ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

INVESTOR ATTENTION AND SOCIAL MEDIA SENTIMENT IN INTERNATIONAL STOCK RETURNS AND TRADING ACTIVITY

Ph.D. THESIS

Selin DÜZ TAN

Department of Management

Management Programme

NOVEMBER 2019

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

INVESTOR ATTENTION AND SOCIAL MEDIA SENTIMENT IN INTERNATIONAL STOCK RETURNS AND TRADING ACTIVITY

Ph.D. THESIS

Selin DÜZ TAN
(403152007)

Department of Management

Management Programme

Thesis Advisor: Prof. Dr. Oktay TAŞ

NOVEMBER 2019

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ SOSYAL BİLİMLER ENSTİTÜSÜ

ULUSLARARASI HİSSE SENEDİ GETİRİLERİ VE İŞLEM HACİMLERİNDE YATIRIMCI İLGİSİ VE SOSYAL MEDYA DUYARLILIĞI

DOKTORA TEZİ

Selin DÜZ TAN
(403152007)

İşletme Anabilim Dalı

İşletme Programı

Tez Danışmanı: Prof. Dr. Oktay TAŞ

KASIM 2019

Selin DÜZ TAN, a Ph.D. student of ITU Graduate School of Arts and Social Sciences student ID 403152007, successfully defended the thesis/dissertation entitled "INVESTOR ATTENTION AND SOCIAL MEDIA SENTIMENT IN INTERNATIONAL STOCK RETURNS AND TRADING ACTIVITY", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Oktay TAŞ** ....
Istanbul Technical University

Jury Members : **Prof. Dr. Murat KIYILAR** ....
Istanbul University

Assoc. Prof. Dr. Cumhur EKİNCİ ....
Istanbul Technical University

Prof. Dr. Suat TEKER ....
Işık University

Prof. Dr. Fuat BEYAZIT ....
Bilgi University

Date of Submission : 7 October 2019
Date of Defense : 20 November 2019

To my family,

FOREWORD

I gratefully thank to my supervisor Prof. Dr. Oktay TAŞ for his guidance, support and encouragement throughout the development of this thesis. I would like to express my gratitude to the members of the thesis monitoring committee, Assoc. Prof. Dr. Cumhur EKİNCİ and Prof. Dr. Murat KIYILAR for their expertise and recommendations.

I would like to thank my soulmate for his endless love, support, inspiration and encouragement throughout my life. I sincerely would like to thank my mother for supporting me spiritually throughout writing this thesis and my life in general. Special thanks to Borsa Istanbul team for their support and friendship during this process.

November 2019 Selin DÜZ TAN

TABLE OF CONTENTS

Page

ABBREVIATIONS

(o-c)	: Open-to-close returns
(o-o)	: Open-to-open returns
ASVI	: Abnormal Search Volume Index
CAPM	: Capital Asset Pricing Model
DJIA	: Dow Jones Industrial Average
EM	: Emerging markets
EMH	: Efficient Market Hypothesis
FEARS	: Financial and Economic Attitudes Revealed by Search
GNH	: Gross National Happiness
HML	: High minus low
IPO	: Initial public offering
MKT	: Equity market return in excess of the risk free rate
S&P 500	: S&P 500 Index
S&P EM	: S&P Emerging Markets Core Index
S&P Europe	: S&P Europe 350 Index
SMB	: Small minus big
SVI	: Search Volume Index
UMD	: Up minus down
VIX	: Chicago Board Options Exchange Volatility Index

LIST OF TABLES

Page

LIST OF FIGURES

Page

INVESTOR ATTENTION AND SOCIAL MEDIA SENTIMENT IN INTERNATIONAL STOCK RETURNS AND TRADING ACTIVITY

SUMMARY

Investors have limited time, effort and cognitive resources to process information. The traditional capital asset pricing model based on the efficient market hypothesis assumes that information on securities is incorporated into prices instantaneously. This is not always true in real life, and some of the investors may have limited knowledge.

Social media platforms, a key enabler of information, opinion, thought and knowledge sharing through online forums, blogs and platforms, have evolved extensively in recent years and the expansion of the social media platforms have enabled researchers to explore the complex behavior of the investors. Following these developments, research on social media data as a measure of the investor behavior have been increasing in recent years.

Behavioral finance challenges the Efficient Market Hypothesis by highlighting the significant role of human emotion, sentiment and mood in financial decision-making. In the last decade, social media such as Twitter and stock message boards have become an important part of these decisions. Investors started to make trading decisions by following these social media tools and forums. Social media sentiment derived from social media tools and forums can capture investors with bounded rationality and these less rational investors are mostly individual investors.

The main purpose of this thesis is to investigate the impacts of investor attention measured by Google Search Volume Index (SVI) and social media activity measured by the number of tweets and social media sentiment measured by Twitter sentiment on individual stock returns and trading activity. The Fama and MacBeth regression model is used as it is widely used in finance literature in order to investigate the relationship between expected returns and factor coefficients. The method is used in asset pricing as it is useful to work with panel data and multiple assets across time. Investor attention is measured by different methods in the literature: Internet search volume, media coverage, abnormal trading volume, extreme returns, advertising expenditure, option trading volume, institutional ownership, firm size, analyst coverage and activity in terminals. The thesis uses Google search volume as a direct proxy of investor attention based on the facts that search volume is likely to capture attention for two main reasons. First, individuals generally use Internet search engine to gather information, which means Google search volume might represent the aggregate household interest on a topic in general. Second, Google search volume index data provides better indication of investors' behaviors or decisions than other investor attention proxies such as turnover, news and abnormal returns, because Internet search volume is a direct and active measure of investor attention that is more likely to be related to an action or buying where other measures are indirect and passive measures. People may search for a firm name for various reasons, i.e. for product information research, store location searches, or job searches and search queries on firm name is based on how the individuals have searched the firm name that it is affected by subjectivity. Firms'

stock tickers are uniquely assigned and since this thesis focuses on the individuals who are interested in investing, this thesis uses Google Search Volume Index using ticker-based search after excluding generic meaning tickers.

Investor attention is mostly measured by quantitative data such as the search volume index, number of news, trading volume and number of analysts, while social media sentiment examines the content and tone of the texts that investors share. Social media is an interactive environment in which people share ideas, emotions and moods that allow people to share information and respond to shared information. Therefore, the information obtained through social media can be analyzed not only quantitatively but qualitatively and social media plays an important role in understanding the behavior of the society. Several social media platforms are used as a measure for investor sentiment such as Twitter, Stocktwits, Facebook and stock message boards. Social media sentiment is important to analyze the complex behavior of the investors with positive and negative texts of comments on stocks as a direct measurement. In recent years, Twitter has been one of the leading social networks around the world considering active users. The tremendous amount of information on tweets that measure the interest and sentiment of the society have been attracting the attention of many academics and professionals. Therefore, in this thesis, Twitter and StockTwits are used as a measure of the social media sentiment.

The thesis is divided into two main parts providing substantial evidence for three hypotheses based on investigating the effect of aggregate investor behavior on individual stock indicators. In the first part, the sole hypothesis examines how stock returns change when attention levels of investors measured by abnormal Google search volume index increase in a sample of stocks from Borsa Istanbul all shares index in Turkey. In accordance with the literature stating that individual investors are net buyers of attention-grabbing stocks, this thesis shows evidence that an increase in abnormal SVI (ASVI) is related to higher future returns. The main finding is that firms attracting abnormally high attention earn higher returns and the price pressure effect of SVI is stronger among small stocks. The reversals for predictability of searches in stock returns shows buying pressure from uninformed investors. Trading strategy shows that forming a portfolio sorting by attention levels creates a significant return premium per week, but only for the short-term. The results suggest that ASVI is likely to grab the attention of individual investors resulting in short-term buying pressure. These findings reveal that stock prices tend to be driven by the behavioral factors due to the investor attention in Turkey.

The second part of the thesis investigates the impact of daily Twitter activity and the impact of daily Twitter sentiment on stock returns and trading activity in multi-country level under two hypotheses. This part focuses on S&P 500, S&P 350 Europe and S&P Emerging Markets Core index constituents with an international investor perspective because investors are mostly active on Twitter for larger firms and sentiment information could be mostly available for these firms. Using a large sample of stocks in international stock markets, the results provide an evidence that Twitter activity and sentiment are associated with trading volume and predict next-day trading volume. The results show that the number of tweets and Twitter sentiment is associated with higher abnormal (raw) stock returns. Daily firm-specific Twitter sentiment contains information for predicting future stock returns, but no such relation exists in the number of tweets or Twitter activity. This predictive power remains significant after controlling the news sentiment. The positive tone of Twitter sentiment has more predictive power in small and emerging market firms. These results are consistent with the literature stating that small firms are hard-to-value and emerging market firms contain high information asymmetry. Overall, these results suggest that social media activity and sentiment provide new information about firms and show that social media present different impacts than traditional news media on firms' information environments. From a practical perspective, investors could potentially use social media sentiment in their trading strategies. The predictive power of Twitter sentiment for stock returns may influence market participants' trading decisions. Trading strategy with long-short portfolio using deciles of sentiment in Twitter sentiment generates significant positive returns per 5 days holding period even after considering trading costs. Due to return reversals, these findings suggest that the predictive power is short sighted, and strategies might be formed only for short term investments.

ULUSLARARASI HİSSE SENEDİ GETİRİLERİ VE İŞLEM HACİMLERİNDE YATIRIMCI İLGİSİ VE SOSYAL MEDYA DUYARLILIĞI

ÖZET

Yatırımcıların bilgi işleme konusunda zaman, çaba ve bilişsel kaynakları kısıtlıdır. Etkin piyasa hipotezine dayanan geleneksel sermaye varlık fiyatlandırma modeli, menkul kıymetler hakkındaki bilgilerin anında fiyatlara yansıdığını varsaymaktadır. Bu durum gerçek hayatta her zaman doğru değildir ve yatırımcılar sınırlı bilgiye sahip olabilir.

Son yıllarda gelişme gösteren sosyal medya platformları araştırmacıların yatırımcıların karmaşık davranışlarını keşfetmelerini sağlamaktadır. Çevrimiçi forumlar, bloglar ve platformlar, bireylerin düşünce, fikir ve bilgileri paylaşmalarına izin vermektedir. Yatırımcı davranışının bir ölçütü olarak sosyal medya verilerini kullanan araştırmalar son yıllarda artış göstermektedir. Davranışsal finans, finansal karar vermede duygu ve ruh halinin önemli rolünü vurgulayarak Etkin Piyasalar Hipotezi'nin geçerliliğini test etmektedir. Son yıllarda, Twitter ve yatırım forumları gibi sosyal medya platformları karar mekanizmalarının önemli bir parçası haline gelmiştir. Yatırımcılar bu sosyal medya araçlarını ve forumlarını takip ederek alım satım kararları almaya başlamıştır. Sosyal medya duyarlılığı, sınırlı rasyonelliğe sahip yatırımcıları yansıtabilmektedir. Daha az rasyonel olan bu yatırımcılar çoğunlukla bireysel yatırımcılardır. Sosyal medya duyarlılığının rolü, teknolojik gelişme ve gelişen işlem platformlarının bireysel yatırımcıları etkilemesi üzerine artış göstermektedir.

Bu tezin asıl amacı, Google arama hacmi endeksi ile ölçülen yatırımcı ilgisinin, tweet sayısı ile ölçülen sosyal medya faaliyeti ve Twitter yatırımcı duyarlılığı ile ölçülen sosyal medya duyarlılığının hisse senedi getirileri ve işlem hacmi üzerindeki etkilerini araştırmaktır. Tezde beklenen getiri ile faktör katsayıları arasındaki ilişkiyi araştırmak için finans literatüründe yaygın olarak kullanılan Fama ve MacBeth regresyon modeli kullanılmaktadır. Bu yöntem literatürde varlık fiyatlandırması alanında zaman içerisinde değişen birden çok varlık fiyatını incelemek amacıyla çoğunlukla kullanılmaktadır. Yatırımcı ilgisi literatürde farklı birçok yöntemle ölçülmektedir: İnternet arama hacmi, haber kapsamı, normal üstü işlem hacmi, aşırı getiri, reklam harcamaları, opsiyon işlem hacmi, kurumsal sahiplik, firma büyüklüğü, analist takibi ve veri terminallerdeki faaliyet. Tezde Google arama hacmi endeksi iki temel nedene dayanarak yatırımcı ilgisinin ölçütü olarak kullanılmaktadır. Birincisi, bireyler genellikle bilgi toplamak için internet arama motorunu kullanmakta olup bu durum Google arama hacminin toplam hane halkının ilgisini temsil edebileceğini göstermektedir. İkincisi, Google arama hacmi endeksi verileri, yatırımcıların davranış veya kararlarını, işlem hacmi, haber sayısı ve normal üstü getiriler gibi diğer yatırımcı ilgisi ölçütlerine göre daha iyi yansıtmaktadır; çünkü internet arama hacmi, yatırımcı ilgisinin diğer değişkenler gibi pasif olmayan alım faaliyetleri ile ilgili doğrudan bir

ölçütüdür. İnsanlar çeşitli nedenlerle, örneğin ürün bilgileri araştırması, mağaza konum aramaları veya iş aramaları için firma adını arama siteleri üzerinden arayabilirler. Firma adına yapılan arama sorguları, bireylerin firma adını nasıl aradıklarına ve yatırım amacından farklı nedenlerle arama sonuçlarına dayanabilmektedir. Firmaların borsa hisse senedi kodları firmalara özel olarak atanmaktadır. Tezin yatırımla ilgilenen kişilere odaklanması sebebiyle, tezde birden fazla anlama gelebilecek hisse senedi kodların elenmesinden sonra hisse senedi kodları üzerinden Google arama verileri kullanılmaktadır.

Yatırımcıların ilgisi çoğunlukla internet arama hacmi endeksi, haber sayısı, işlem hacmi ve analist sayısı gibi nicel verilerle ölçülürken, sosyal medya duyarlılığı yatırımcıların paylaştığı metinlerin içeriğini ve tonunu incelemektedir. Sosyal medya, insanların bilgi paylaşmalarını ve paylaşılan bilgilere cevap vermelerini sağlayan fikirleri, duyguları ve ruh hallerini paylaştığı etkileşimli bir ortamdır. Bu nedenle, sosyal medya aracılığıyla elde edilen bilgiler sadece niceliksel olarak değil niteliksel olarak da analiz edilmekte ve sosyal medya toplumun davranışını anlamada önemli bir rol oynamaktadır. Sosyal medya platformlarından Twitter, Stocktwits, Facebook, hisse senedi mesaj panoları, Yahoo! Finance ve yatırım forumları yatırımcı duyarlılığının ölçüt araçları olarak kullanılmaktadır. Sosyal medya duyarlılığı, yatırımcıların karmaşık davranışlarını doğrudan ölçerek olumlu ve olumsuz yorum metinleriyle analiz etmesi sebebiyle önem taşımaktadır. Son yıllarda, aktif kullanıcı sayıları göz önünde bulundurulduğunda Twitter dünyanın önde gelen sosyal ağlarından biri olarak öne çıkmaktadır. Tweet bilgilerinin toplumun ilgisini ve duyarlılığını ölçmesi, birçok araştırmacının ve profesyonelin dikkatini çekmektedir. Bu nedenle tezde Twitter ve StockTwits sosyal medya duyarlılığının bir ölçütü olarak kullanılmıştır.

Tez, yatırımcı davranışının hisse senedi göstergeleri üzerindeki etkisinin araştırılmasına dayanan iki ana bölüme ayrılmıştır. Birinci bölümde, Borsa İstanbul'da işlem gören hisse senetlerinde normal üstü Google arama hacmi endeksiyle ölçülen yatırımcı ilgisi seviyesi artış gösterdiğinde hisse senedi getirilerinin nasıl değiştiği incelenmektedir. Bireysel yatırımcıların ilgi çekici hisse senetlerinin net alıcıları olduğunu belirten literatüre dayanarak, bu tez normal üstü Google arama hacmi endeksindeki artışın gelecekteki getirilerle ilgili olduğuna dair kanıtlar sunmaktadır. Asıl bulgu, normal üstü derecede yüksek ilgi çeken firmaların daha yüksek getiri elde etmeleri ve arama hacmi endeksinin fiyat baskısı etkisinin piyasa değeri açısından daha küçük hisse senetleri için daha güçlü olmasıdır. Hisse senedi getiri tahminlerindeki geri dönüşler bilgisiz yatırımcıların alım baskısını göstermektedir. Alım satım stratejisi, yatırımcı ilgisi seviyelerine göre sıralama yaparak portföy oluşturmanın haftalık bazda kısa vadede önemli bir getiri primi yarattığını göstermektedir. Sonuçlar, normal üstü Google arama hacmi endeksinin bireysel yatırımcıların ilgisini yansıttığını ve kısa vadeli alım baskısı yarattığını göstermektedir. Bu bulgular, hisse senedi fiyatlarının Türkiye'deki hisse senedi getirilerinin yatırımcının ilgisine bağlı davranışsal faktörlerden etkilenme eğiliminde olduğunu göstermektedir.

Tezin ikinci bölümünde, günlük Twitter faaliyeti ve yatırımcı duyarlılığının birden fazla ülke kapsamında hisse senedi getirileri ve alım satım faaliyeti üzerindeki etkisi incelemektedir. Bu bölüm, S&P 500, S&P 350 Avrupa ve S&P Gelişmekte Olan Piyasalar Çekirdek endeksine dahil hisse senetleri uluslararası yatırımcı perspektifiyle incelemektedir, çünkü yatırımcılar çoğunlukla daha büyük firmalar için Twitter'da aktif durumdadır ve bu firmalar için kolaylıkla duyarlılık bilgisi elde edilebilmektedir. Uluslararası hisse senedi piyasalarında geniş bir hisse senedi örneklemini kullanan çalışma, Twitter faaliyeti ve duyarlılığının işlem hacmi ile ilişkili olduğunu ve ertesi gün işlem hacmini tahmin ettiğini göstermektedir. Sonuçlar, tweet sayısının ve Twitter duyarlılığının, daha yüksek hisse senedi getirileriyle ilişkili olduğunu ve günlük firmaya özgü Twitter duyarlılığının gelecekteki hisse senedi getirilerini tahmin etmek için bilgi içerdiğini, ancak tweet sayısının tahmin gücü bulunmadığını göstermektedir. Bu tahmin gücü, haber duyarlılığı kontrol edildikten sonra da devam etmektedir. Twitter duyarlılığının pozitif tonu, piyasa değeri açısından daha küçük ve gelişmekte olan piyasa firmalarında daha fazla tahmin gücüne sahiptir. Bu sonuçlar, piyasa değeri açısından daha küçük firmaların daha zor değerlendiği ve gelişmekte olan piyasa firmalarının yüksek bilgi asimetrisi içerdiğini belirten literatürle tutarlıdır. Genel olarak, bu sonuçlar sosyal medya aktivitesi ve duyarlılığının firmalar hakkında yeni bilgiler sağladığını ve sosyal medyanın geleneksel haber medyasından farklı olarak firmaların bilgi ortamları üzerinde etkisi olduğunu göstermektedir.

Uygulama açısından bakıldığında, yatırımcılar sosyal medya duyarlılıklarını alım satım stratejilerinde kullanabilirler. Twitter duyarlılığının hisse senedi getirileri için tahmin gücü, piyasa katılımcılarının işlem kararlarını etkileyebilir. Twitter kamuya açık bir platformdur ve yatırım amacıyla alım satım stratejileri için kullanılabilir. Twitter gibi sosyal medya kaynaklarından elde edilen yatırımcı duyarlılığı bilgileri, finansal piyasa katılımcıları için yatırımcı bilgi ve inançlarını yansıtması açısından önemli bir ölçüttür.

Kurumsal yatırımcılar, işlem gören hisse senetlerini profesyonel platformlar ve araçlar yardımıyla takip edebilmekte olup sosyal medya platformları bireysel yatırımcıların bilgiye kolayca erişmesine yardımcı olmaktadır. Bireysel yatırımcılar literatürde psikolojik önyargıları olan daha az bilgiye sahip yatırımcılar olarak tanımlanmaktadır. Kurumsal veya bilinçli yatırımcılar, sosyal medya platformlarını kullanan irrasyonel yatırımcıların davranışlarını ve yatırımcı duyarlılık bilgilerini kar elde etmek amacıyla kullanabilirler. Tezdeki bulgular, Twitter duyarlılığını kullanan uzun ve kısa pozisyonlu portföylere dayalı alım satım stratejisinin, alım satım maliyetleri düşünüldükten sonra bile beş günlük elde tutma süresinde önemli bir getiri sağladığını göstermektedir. Getiri geri dönüşlerinin yaşanması nedeniyle, bu bulgular tahmin gücünün kısa vadeli olduğunu ve stratejilerin yalnızca kısa vadeli yatırımlar için oluşturulabileceğini göstermektedir.

Tezde yatırımcı ilgisi ve sosyal medya faaliyetleri ve duyarlılığının hisse senedi piyasalarındaki etkileri üzerine Türkiye ve uluslararası piyasalar için kanıtlar sunulmaktadır. Yatırımcı ilgisinin Google arama hacim endeksi kullanılarak Türkiye'de işlem gören hisse senetleri üzerine etkisini incelemesi açısından bir ilk olan tez kapsamında arama hacmi endeksi gibi yatırımcı ilgisini sayısal olarak ölçen yatırımcı davranış ölçütünün yanısıra yatırımcıların paylaşımlarının pozitif ve negatif tonlamalarını da ölçen bir duyarlılık ölçütü de kullanılmaktadır. Sosyal medya yatırımcı duyarlılığına ilişkin büyük verilerin işlenmesinde yaşanan zorluklar sebebiyle bu alanda yapılan çalışmalar kısıtlı olmakla birlikte sosyal medyaya dayalı yatırımcı duyarlılığının çok ülkeli ve bölgesel farklılıkları içerecek şekilde finansal piyasalar üzerindeki etkilerine yönelik hiçbir çalışmaya rastlanmamıştır. Bu doğrultuda yatırımcı duygu ve davranışlarına ilişkin bilgilerin çok ülkeli kapsamda incelenmesi açısından bir ilk olma niteliği taşıyan bu tez aynı zamanda yatırımcı ilgisi ve sosyal medya yatırımcı duyarlılığının etkilerine dayanan alım satım stratejileri sunarak mevcut literatürdeki uygulama alanını genişletmektedir. Bu kapsamda yatırımcı davranışlarının etkilerinin anlaşılması, bu etkilerin yatırımcılar ve firmalar tarafından takip edilmesi ve stratejilerde kullanılması açısından literatüre katkı sağlamaktadır.

Tezin genel sonucu olarak Google arama hacim endeksi kullanılarak ölçülen yatırımcı ilgisi ve Twitter ile ölçülen sosyal medya duyarlılığının hisse senedi getirileri ve hacimleri üzerinde etkisi olduğu ve bu yatırımcı davranışları konusunda bilgi içeren değişkenlerin hisse senedi getirilerinin tahmininde önemli bir rol oynadığı gösterilmektedir. Yatırımcılar Google ve Twitter gibi kamuya açık Internet arama motoru ve platformları kullanarak alım satım stratejilerini şekillendirebilir. Firmalar performanslarını etkileyebilecek firmaya özgü arama verileri ve yatırımcı duyarlılık bilgisini takip etmek için bu arama motorları ve platformları izleyebilir. Arama sonuçları ve sosyal medya duyarlılığından elde edilen bilgiler, çeşitli sektörlere uygulanabilmekte olup bu bilgiler yatırımcı ilgi ve duyarlılık seviyelerinin izlenmesini ve karşılaştırılmasını sağlamaktadır.

1. INTRODUCTION

In the last decade, studies on theoretical and empirical models that focus on information flow, supply and demand have increased due to the increase in access to information. The traditional capital asset pricing model based on the efficient market hypothesis assumes that information on securities is incorporated into prices instantaneously (Fama, 1970). This is not always true in real life, and some of the investors may have limited knowledge. Attention is a scarce cognitive resource and fluctuates in time and it is not easy for individual investors to follow all developments in markets closely (Kahneman, 1973; Grossman and Stiglitz, 1980; Barber et al., 2009). Recent studies in the area of behavioral finance show that investor attention affects the asset prices. Recent growing literature emphasizes the power of Google Search Volume Index (SVI) in various fields of financial research. Google SVI is used as a measure of investor attention in many prevalent studies. (Da et al., 2011; Joseph et al. 2011; Mondria and Wu, 2011; Aouadi et al. 2013; Preis et al., 2013; Vozlyublennaia, 2014). Researchers have focused more on the recent evidences on behavioral finance and suggest that investor attention impacts asset pricing (Da et al., 2011; Li and Yu, 2012, Mondria and Wu, 2011). Previous studies mostly focus on investor sentiment using various indexes and variables, Michigan Confidence Consumer Index (Lemmon and Portniaguina, 2006), a sentiment index using trading volume, the dividend premium, the closed-end fund discount, the number and first-day returns on initial public offerings (IPOs), and the equity share in new issues (Baker and Wurgler, 2006), residuals as investor sentiment proxy after regressing weekly trading volumes of benchmark stock indexes on macroeconomic variables (Uygur and Tas, 2014) while there is an increasing interest on firm specific investor attention using firm determinants and Internet.

The main motivation of the first part of the thesis to examine investor attention and its impacts on stock returns in the Turkish stock market that are currently missing in the literature. Information is incorporated into asset prices in longer time in emerging markets such as Turkey where information collecting and processing are more costly for investors (Guner et al., 2004). Individual investors have an importance in emerging stock markets. In 2017, the trading volume of domestic individual investors in Borsa Istanbul all shares index was 75% (Url-1). The high percentage of domestic individual investors in the total volume supports the idea that it is important to examine the individual investor attention in Turkish stock market. These characteristics of emerging markets is our motivation for examining the effects of investor attention on stock returns in Turkey.

Using an attention measure based on Google SVI for individual stocks' ticker symbols, we contribute to the existing literature in various ways. Investor attention is measured by different methods in the literature as advertising expenditure (Grullon et al., 2004), extreme returns (Barber and Odean, 2008), media coverage (Fang and Peress, 2009). We use Google SVI as a measure of investor attention because search volume is likely to capture attention. Individuals generally use Internet search engine to gather information, which indicates Google search volume might show people's interest on a topic in general. Google has 91.79% share of web search volume worldwide by the end of 2017 (Url-2). Search queries are direct proxy for attention and more powerful than other proxies used in the literature. Our findings support the results of Da et al., (2011) stating that Google SVI is likely to capture investor attention and leads other proxies of investor attention, turnover, abnormal return and number of news.

Social media platforms have evolved in recent years and the expansion of the social media platforms enable researchers to explore the complex behavior of the investors. Online forums, blogs and platforms allow individuals to share thoughts, opinions and information. Research on social media data as a measure of the investor behavior have been increasing in recent years. Behavioral finance challenges the Efficient Market Hypothesis by highlighting the significant role of human emotion, sentiment and mood in financial decision-making. In recent years, social media such as Twitter and message boards have become an important part of decisions. Investors started to make trading decisions following these social media tools and forums. The effect of social media on financial markets is still not examined thoroughly due to the difficulties in analyzing big data in social media. In the literature, studies mainly focus on behavioral finance proposing the human behavior factors like animal spirits (Shiller, 1984), social mood (Nofsinger, 2005), investor sentiment (Baker and Wurgler, 2007) or psychological factors (Fenzl and Pelzmann, 2012) as a source of market volatility and anomalies. Many studies use sentiment proxy (Baker and Wurgler, 2007) to grab such human behaviors. Social media provides the opportunity to collect direct data about these human factors at the aggregate level. Investor sentiment can be used as a direction signal for trading purposes. Intuitively, if there is positive information about a certain company, we expect the company's stock price to rise, and if there is any negative information, the stock price would decrease.

Twitter is an online social media service that allows users to send short 280-character messages called tweets. In recent years, Twitter has been one of the leading social networks around the world considering active users. As of the end of 2017, Twitter had 330 million monthly active users (Url-3). The tremendous amount of information on tweets that measures the interest and sentiment of the society have been attracting the attention of many academics and professionals. Research on social media data as a measure of the complex behavior of the investors have been increasing in recent years. Social media sentiment or Twitter sentiment is important for analyzing the positive and negative texts of comments on stocks as a direct measurement.

Considering individual stocks, this thesis has two main objectives: (i) assessing the impact of investor attention measured by Google SVI on stock returns and (ii) Twitter activity and sentiment on return and trading volume. Similar to the methodologies of Da et al. (2011) and Tetlock (2011), we use Fama Macbeth regressions in an investigation of whether the investor attention and social media sentiment predicts returns. The linking mechanisms between aggregated investor behavior directly measured by Internet search, social media and financial markets have many practical implications in investment decisions. Investor attention is mostly measured by quantitative data such as the search volume index, number of news, trading volume and number of analysts, while social media sentiment examines the content and tone of the texts that investors share. Social media is an interactive environment in which people share ideas, emotions and moods that allow people to share information and respond to shared information. Therefore, the information obtained through social media can be analyzed not only quantitatively but also qualitatively and social media plays an important role in understanding the behavior of the society. In this thesis, both the effects of the number of attention and tone of the sentiment are investigated for the practical implications in investment decisions.

In the first part of the thesis, to understand the effect of investor attention on stock returns, we examine how stock returns change when attention levels of investors increase in sample of Borsa Istanbul all shares index stocks over the period April 2013 to September 2017. The hypothesis tests the claim that when investor attention, namely abnormal Search Volume Index (ASVI), is higher, stock returns of the subsequent week will be higher. This hypothesis supports the attention-induced price pressure hypothesis developed by Barber and Odean (2008) and supports that investors are not rational and changes in investor sentiment are an important determinant of prices as stated by Shleifer and Summers (1990). We also test whether price pressure hypothesis induced by investor attention is more pronounced among small stocks since small stocks are more prone to larger price impact. We use following firm characteristics as control variables similar to Da et al. (2011) and Ying et al. (2015): stock turnover, book to market ratio, size or the ratio of stock's market capitalization to all shares index market capitalization, volatility as the standard deviation of the daily stock returns for the week and news, the number of stories published on the recent week. Our findings provide new evidence for attention theory of Barber and Odean (2008) from an emerging country perspective. We find that firms attracting abnormally high attention earn higher returns and the price pressure effect of SVI is stronger among small stocks. Da et al. (2011) propose to use Google SVI for stock ticker as a direct proxy of investor attention and state that SVI captures attention more properly than indirect attention proxies, and mainly measures the individual investor attention. Building on the work of Da et al. (2011), we present SVIs on stock tickers as a likely proxy for investor attention. Our results support the results of Da et al. (2011) which show that an increase in Google SVI predicts higher stock prices in the subsequent weeks. The reversal dynamics of predictability of returns is different from the developed market analysis in Da et al. (2011). We find that predictability effect on return is longer than the effect in developed markets. This result supports the idea that information efficiency is lower in emerging markets and information is incorporated into asset prices in longer time in emerging markets such as Turkey since access to information is more difficult. We provide further evidence on trading strategy that shows portfolios with long in highest decile and short in lowest decile in abnormal investor attention tend to obtain significant alphas for stocks. After using market, size, book to market and momentum factors as controlling factors, there exists high-attention return premium in the short run.

In the second part of the thesis, we focus on S&P 500, S&P 350 Europe and S&P Emerging Markets Core index constituents with the international investor perspective because investors are mostly active on Twitter for larger firms and sentiment information could be easily obtained for these firms. Using a large sample of stocks in international stock markets, we find that Twitter activity and sentiment are associated with trading volume and predicts next-day trading volume. We show that the number of tweets and Twitter sentiment is associated with higher abnormal (raw) stock returns. We find that daily firm-specific Twitter sentiment contains information for predicting future stock returns, but no such relation exists in the number of tweets or Twitter activity. This predictive power remains significant after controlling the news sentiment. The positive tone of Twitter sentiment has more predictive power in small and emerging market firms. These results are consistent with the literature stating that small firms hard-to-value and emerging market firms contain high information asymmetry. Overall, these results suggest that social media activity and sentiment provides new information about firms and show that social media present different impacts than traditional news media on firms' information environments. The results show the role of social media in diffusing sentiment to investors who unintentionally make prices less efficient in the short run.

Institutional investors can follow traded stocks by the help of professional tools, however social media helps individual investors to access the information easily (Chen et al., 2014; Behrendt and Schmidt, 2018). Individual investors are defined as noise traders who have psychological biases (Kyle, 1985; Black, 1986). Institutional investors can exploit the behavior of irrational investors as sentiment driven noise traders who use social media platforms. From a practical perspective, investors could potentially use social media sentiment in their trading strategies. The predictive power of Twitter sentiment for stock returns may influence market participants' trading decisions. We show that a trading strategy based on Twitter sentiment generates significant positive returns even after considering trading costs. Due to return reversals, these findings suggest that the predictive power is short sighted, and strategies might be formed only in short run.

The main contributions on investor attention part of this thesis are summarized as follows. First, we propose a novel stock specific direct investor attention proxy based on Google SVI that have not been studied in prior literature for Turkish stock market.

We support our results with other attention variables as other proxies of investor attention, abnormal turnover, abnormal return and number of news. With the increasing use of the Internet, it has become more popular for investors to use the Internet as a mean of accessing information. It is hard to measure investor attention level with tools such as newspapers and television due to difficulties in measuring time spent and effort of people on these tools. However, investors' attention can be directly measured by Google search tool using user's exact search query. Second, we find evidence of significant and positive relation between abnormal stock returns and ASVI and we show that the price pressure effect of ASVI is stronger among small stocks. Individual investors are important in emerging stock markets. Information is incorporated into asset prices in longer time in emerging markets such as Turkey where information collecting and processing are more costly for investors (Guner et al., 2004). Third, this study contributes to the trading strategy based on attention and shows that a portfolio with long position in high attention stocks and short position in low attention stocks has a significant alpha. While preceding studies mostly focus on US market, we examine Turkey, an emerging market, that received increasing attention from investors due to higher returns. The high percentage of domestic individual investors in the total volume supports the idea that it is important to examine the individual investor attention in Turkish stock market.

The main contributions on social media sentiment of this thesis are summarized as follows. First, we use novel firm specific Twitter sentiment unlike previous related research, and we investigate the impact of Twitter sentiment on individual stock returns in multi-country level. To the best of our knowledge, our study is the first to comprehensively explore the information content of company specific Twitter sentiment rather than stock market indices by comparing regional differences. Emerging and developed markets differ in terms of information environment. Griffin et al. (2011) suggest that emerging market stock returns give slow reaction to news. There has been an increasing interest on how social media sentiment influences the emerging markets. Social media sentiment can capture investors with bounded rationality (De Long et al., 1990; Shleifer and Vishny, 1997; Barberis et al., 1998). These less rational investors are mostly individual investors. Thus, the role of social media sentiment in financial markets has been increasing as a result of an increase in the number of individual investors due to technological developments. Second, we focus on international index constituents with the international investor perspective because investors are mostly active on Twitter for larger firms and sentiment information is more accessible for these firms. The results give an evidence that social media activity and sentiment provides new information about firms and social media present different impacts than traditional news media. Using a large sample of stocks in international stock markets, we find that Twitter activity and sentiment are associated with trading volume. Daily firm-specific Twitter sentiment contains information for predicting future stock returns, but no such relation exists in the number of tweets or Twitter activity. The positive tone of Twitter sentiment has more predictive power in small and emerging market firms. These results are consistent with the literature stating that small firms are hard-to-value and emerging market firms contain high information asymmetry. Third, from a practical perspective, investors could potentially use social media sentiment in their trading strategies and have significant positive returns even after considering trading costs.

The thesis is organized as follows. Section 2 provides a brief literature review and background on the investor attention and sentiment where theoretical limited attention and sentiment models and behavioral finance connections are provided, different measures of investor attention and social media sentiment are discussed, and their empirical results are provided. Section 2 also presents the main hypotheses based on investor attention and social media sentiment and presents the brief outline of the research questions. Section 3 describes the investor attention, sentiment and financial data used in this study, and the methodology of thesis is discussed regarding the effects of investor attention and sentiment on individual stock returns and trading activity. Also, the detailed data obtaining process, main and control variables are provided in this section. Section 4 provides empirical results and discussions on the main hypotheses and implications of trading strategies on investor attention and sentiment. Section 5 gives conclusions regarding the main hypotheses in the thesis. The Appendix provides additional information for the models in the thesis.

2. LITERATURE REVIEW AND HYPOTHESES

This section reviews the literature on investor attention and social media sentiment. First, traditional finance and Efficient Market Hypothesis (EMH) is discussed and criticisms to EMH are given within behavioral finance. Then, the theoretical and empirical basis of the links between behavioral finance and investor attention is given. Common measures used in the literature as investor attention proxies are reviewed by classifying and discussing their empirical results. After that, the theoretical and empirical background of the relation between behavioral finance and social sentiment is given. Lastly, common proxies used in the literature as social media sentiment are reviewed by classifying and discussing their empirical results.

2.1 Traditional Finance and Efficient Market Hypothesis

The basis of traditional finance theories was laid foundation by Markowitz (1952) with studies on portfolio selection and the interaction of risks and expected returns. The Capital Asset Pricing Model (CAPM), developed by Sharpe (1964) and Lintner (1975), combines the cornerstones of traditional finance theory and provides a framework for the relationship between risk, capital structure and expected stock returns. CAPM is used as a starting point for extensive research on asset pricing and cross-sectional return models.

EMH proposed by Fama (1970) states financial markets are efficient where all public and private information are reflected to prices and individual investors cannot constantly beat the market. EMH is built on the assumption that investors are fully rational. Fama (1970) divides the empirical work into three forms on the nature of the informational efficiency: the weak form, the semi-strong form, and the strong form. In strong-form tests, monopolistic access to any information by individual investors or groups is important for price formation. In the semi-strong form, the information subset contains all publicly available information. In the weak form, the information subset is limited to historical price or return sequences. No profit can be gained through technical analysis when the weak form is held. No profit can be made through publicly available information in the semi-strong form. In strong form, in addition to publicly available information, insider information is futile to obtain superior returns. Therefore, within the framework of the EMH, it is not possible for investors to beat the market and make superior returns analyzing the past price movements and fundamentals because available information is already reflected into the prices. EMH has faced many empirical challenges over the years. Fama (1991) reviews the weak form of efficiency and proposes that stock returns are predictable using past returns. With reference to the empirical analysis on prices adjust efficiently to firm-specific information, Fama (1991) proposes event studies and private information analysis for semi-strong and strong form tests. Emphasizing that the anomalies are chance results, Fama (1998) states that return patterns depend on the empirical methodology and long-term return anomalies are tend to disappear with changes in the measurement and techniques.

The traditional finance theory is based on the efficiency of markets. In this theory, there are no arguments on irrational explanations of stock movements that are related to investor sentiment or noise traders. Critics to EMH on the efficient market perspective have increased with the increasing market anomalies that cannot be explained in rational manner. De Bondt and Thaler (1985) state that people tend to overreact to unexpected news affecting stock prices where there are non-rational return reversals after immediate price reactions. De Bondt and Thaler (1985) test the effectiveness of the EMH weak form while attempting to analyze overreaction in financial markets. The weak form argues that investors cannot make superior profits using past price information. However, with their studies, De Bondt and Thaler (1985) show that investors can make profits in the market using past price information. This study leads the way for further studies in behavioral finance.

In real life, these extreme assumptions on human behavior are not held due to limited attention of human beings. Limited attention is one of the psychological biases that comes from the limit on information processing capacity of the people (Kahneman, 1973). Many finance researchers have been increasingly analyzing the role of investor limited attention on asset pricing.

2.2 Behavioral Finance and Investor Attention

Efficient market hypothesis infers that information should be incorporated into stock prices reflecting all relevant information (Fama, 1970). The quantity of financial research on anomalies have been increasing and researchers have focused more on the behaviors that couldn't be explained by this hypothesis and the factors that violate market efficiency. Barberis et al. (1998) and Daniel et al. (1998) propose theories of securities to explain psychological biases and anomalies and they examine reaction types to new information and good-bad news. These studies offer an insight into the investor attention and its role in asset pricing.

Limited attention is one of the psychological biases resulting from the process of information. Investors' allocation of limited attention as a reaction to information and the effect of investors' attention on investor trading is important within this context. As Shleifer and Summers (1990) stated investors are not rational and changes in investor sentiment are an important determinant of prices. If investors were fully rational, they would know that noise trading damages them. Arbitrage is risky and limited. Investors may be classified into two categories as arbitrageurs and noise traders. Arbitrageurs do not answer fully to movements in investor sentiments affecting prices.

Recent evidences in behavioral finance literature show that investor attention affects asset pricing. Many studies focus on how investor sentiment affects security prices. Baker and Wurgler (2006) take the origin of investor sentiment as exogenous affecting patterns in security pricing and they state that investor sentiment has more impact on stocks that are hard to arbitrage or to value. There are two main views of how investor sentiment affects security prices. Merton (1987) states that firms which attract less investor attention have to give higher returns to compensate imperfect diversification. Barber and Odean (2008), with "attention theory", suggest that attention creates buying pressure of uninformed individual investors emphasizing that individual investors are net buyers of attention-grabbing stocks in the short term.

2.3 Measures of Investor Attention

Investor attention, is measured by different methods in the literature: Internet search volume such as Google, Yahoo, Baidu (Da et al., 2011; Lawrence et al., 2016; Zhang et al., 2013), media coverage (Fang and Peress, 2009), abnormal trading volume (Barber and Odean, 2008), extreme returns (Barber and Odean, 2008), advertising expenditure (Grullon et al., 2004), option trading volume (Wang et al., 2018), firm size, analyst coverage (Lee et al., 2019), stock spam messages (Nelson et al., 2013) and activity in Bloomberg terminals (Rephael et al., 2017). Internet search volume is a direct and active measure of investor attention where other measures are indirect and passive measures.

2.3.1 Internet search volume

For investors, investment forums and Google search engine are easily and publicly accessible source of information because investors have limited sources and limited access to professional databases (Bukovina, 2016). Internet search volume is a direct proxy of investor attention. In recent decade, there have been several studies that focus on Google SVI as a proxy for investor attention. The studies in the literature on investor attention can be classified into three main groups. First, these studies can be classified by transmission mechanism as information demand and aggregating moods of the society where various sentiment indexes are formed based on search words of the society such as fear and crisis. Second classification can be made by attention proxy or Internet search engines as Google, Yahoo and Baidu. Third classification method is to focus on main dependent variables, individual stock-level or index.

First, the review of the studies is classified by investors' information demand mechanisms. The main preceding study on Google SVI that suggests SVI is a likely measure of the attention of individual investors is the study of Da et al. (2011). Da et al. (2011) propose the use of Google SVI for stock ticker as a direct proxy of investor attention and introduce that SVI captures attention more properly than indirect attention proxies and SVI mainly measures the individual investor attention. The authors support the idea of price pressure due to attention as proposed by Barber and Odean (2008). Da et al. (2011) shows that an increase in SVI predicts higher stock prices in the next 2 weeks with subsequent price reversals and SVI has impact on the large first-day return and long-run underperformance of IPO stocks. Vlastakis and Markellos (2012) examine firm and market level information demand and supply for Dow 30 stocks. They define information demand with weekly Google SVI and find that information demand and supply have contemporaneous and dynamic relation where market information demand is positively related to historical volatility, implied volatility and trading volume. They also show that information demand increases during the periods of higher returns and investors demand more information with the increase in their level of risk aversion. Joseph et al. (2011) examine S&P 500 stocks and show that online searches predict trading volumes and abnormal stock returns and the sensitivity of returns to search volume is positively related to the arbitrage difficulty of a stock. Mondria and Wu (2011) propose a novel measure of attention as asymmetric attention to examine the attention difference between local and non-local investors in S&P 500 stocks using search queries. They find that firms attracting abnormal high asymmetric attention from local to non-local investors have higher returns and long portfolio that consists of high asymmetric attention stocks has higher alpha per month. Bijl et al. (2016) presents contradicting results compared to the studies listed above. They examine S&P 500 stocks and find that high Google search volumes lead to negative returns. They also suggest that trading strategy on SVI is not profitable when the transaction cost is considered. Different from previous studies, Ding and Hou (2015) use active attention measure as Google SVI on stock tickers and passive attention measures as Google News coverage and advertising expenditure. They analyze S&P 500 stocks and show that passive attention measures cannot explain most of the variation in SVI where SVI increases the shareholder base improving stock liquidity. Chai et al. (2019) use abnormal SVI for the stocks in Australian market and conclude that higher ASVI leads to higher turnover or trading activity, a greater OIB between buy and sell orders, and high liquidity. This study supports the literature that states there is a positive relation between ASVI and stock returns over a short investment horizon and the effect is stronger in stocks with high arbitrage costs.

There are few studies that examine stock markets and investor attention using Google SVI beyond the US. Takeda and Wakao (2014) examine the relation between search intensity with company names on Google and stock-trading behavior in Japan. They find that an increase in Google search activity is associated with an increase in trading, but the relation of search intensity is weakly positive for stock returns. Aouadi et al. (2013) examine the effect of investor attention on trading activity and volatility in France. The authors show that the correlation between investor attention and trading volume is high and attention has significant effects on stock market illiquidity and volatility after controlling crisis effect. Bank et al. (2011) show that an increase in search volume is related to an increase in trading activity and stock liquidity. The authors also state that search queries measure attention from uninformed investors and search volume increase is associated with higher future returns in short term. There are few studies on investor attention in emerging markets in individual stock level. Swamy et al. (2019) examine the impact of Google SVI in forecasting stock returns using the quantile regression approach. The results of this study suggest that a higher SVI predicts positive and significant returns in the next two weeks. Their model with SVI has better predictability performance on excess returns than the model without SVI.

Some of the studies focus on analyzing the impact of search volume on stock indexes instead of individual stocks. Dimpfl and Jank (2016) show the relation between stock market volatility and individual investor attention. The results depict a strong co-movement of the realized volatility and search volumes or investor attention. Greater number of search volumes causes increase in volatility in the next day. Vozlyublennaia (2014) explores the relation between the performance of indexes and investor attention as measured by Google search queries. The study demonstrates that attention affects the performance of indexes in short term. On the contrary, changes in returns significantly influence attention in long-term. The results show the significant interaction effects between lagged returns and investor attention suggesting that attention can affect predictability of index returns. An increase in investor attention decreases return predictability that leads to improved market efficiency. Tantaopas et al. (2016) investigate the relation between investor attention return, volatility and trading volume for Asia-Pacific equity market indexes using Google search volumes. The authors find that one-way causality and change in market variables leads to change in attention. The authors also state that existence of attention is important for predicting returns in the market because investor attention leads to more efficient market. They also show that there is an asymmetric relation between positive and negative market trends and attention. The studies of Vozlyublennaia (2014) and Tantaopas et al. (2016) are different from other presented papers that come to agreement on the behaviors of less rational investors who have the information demand and show this attention effect on their trading decisions. However, these studies suggest that information is demanded by more informed investors leading more efficient markets where previous studies suggest that information is demanded by less sophisticated and noise information traders. Peltomaki et al. (2018) uses two investor attention proxies, SVI and abnormal trading volume and investigate the impact of investor attention on stock market and FX market volatility in emerging economies. The results show that new practical proxies formed by taking the first principal component of the SVI and the abnormal trading value are more likely to capture the complex nature of investor attention. The results of this study also show that investor attention explains stock market volatility and shocks, but do not explain FX market volatility and state that emerging markets are prone to changes to investor attention.

Some studies use Internet search volume to measure society sentiment such as fear, crisis sentiment. Mao et al. (2011) compare sentiment tracking methods (Twitter sentiment, negative news sentiment and tweet & Google search volumes of financial terms) for financial prediction of market indices such as the Dow Jones Industrial Average, trading volumes, and market volatility (VIX), and gold prices. The results depict that weekly Google Search Volume Index on financial search queries carry a predictive value. An indicator of Twitter sentiment and the frequency of occurrence of financial terms on Twitter in the previous 1-2 days are also statistically significant predictors of daily market log return. Da et al. (2015) use the search results related to investor concerns (e.g., "recession," "unemployment," and "bankruptcy") and construct a Financial and Economic Attitudes Revealed by Search (FEARS) index as a proxy for investor sentiment. The results give an evidence that FEARS index predicts short-term return reversals, temporary increases in volatility and mutual fund flows out of equity funds and into bond funds. Preis et al. (2013) work on understanding of collective human behavior and analyze changes in Google SVI for search terms related to finance. The results show the existence of "early warning signs" of stock market moves and suggest that these warning signs in search volume data could be used in profitable trading strategies.

There are various Internet search engines used in the studies such as Google, Yahoo and Baidu. Most studies in the literature use Google which is the most widely used search engine in the world (Url-2). In countries that do not use Google as search engine, for example China, studies use Baidu search engine in their analysis. Ying et al. (2015) use searches obtained from Baidu.com to analyze investor attention in China and find that investor attention is related to stock returns positively. The results show that the effect of attention is reversed after second week, but the transient effect cannot be completely reversed within a year. Institutional ownership makes this transitory effect smaller and causes stronger return reversals after a month. Zang and Wang (2015) investigate the association between the investor attention from Baidu and stock performance based on the ChiNext stock market in China. Their findings show that limited attention of investors leads to positive price pressure with the reversal in the short term. The results further show that investor attention on non-trading days has a significant effect on the open-price differences. Zhang et al. (2013) use stock names in Baidu Index and show that investor attention is an important variable to predict stock abnormal return where granger causality test reveals the bi-directional pattern. Their results suggest that open source information can increase the speed of information diffusion making the market efficient. Yahoo! Finance is another popular web site for individual investors in the US. Due to large number of users, some studies use Yahoo! Finance to analyze the effect of investor attention on financial markets. Lawrence et al. (2016) use Yahoo! Finance search to investigate the impact of investor attention at earnings announcements. The results show that attention from Yahoo! Finance is associated with earning responses and predicts subsequent returns. This study compares abnormal Google search, EDGAR search and volume in explaining earnings responses and subsequent returns and shows that these alternative measures of attention are not as informative as Yahoo! Finance.

There are few studies that focus on sentiment in Turkish stock market. Tan and Tas (2019) show that firms attracting abnormally high attention earn higher returns and the price pressure effect of ASVI is more pronounced among small stocks in Turkey. Their results show that stock returns tend to be driven by the behavioral factors due to the investor attention in Turkey. Ekinci and Bulut (2018) examine the association between Google search and stock returns in BIST 100 stocks. The results of this study show that there is a positive and significant contemporaneous relation between Google SVI and stock returns but direction of this relationship is vague. Analyzing the effects of Google search queries on BIST 100 index, Korkmaz et al. (2017) find there is a weak causal link from investors' attention to stock returns and stock volume is Granger cause of investor attention. Sayım and Rahman (2015) examine the effect of Turkish individual investor sentiment, i.e. monthly Turkish Consumer Confidence Index, on the Istanbul Stock Exchange 100 index returns and volatility. Uygur and Tas (2014) use residuals as investor sentiment proxy regressing trading volumes of benchmark stock indexes on group of macroeconomic variables and show that earning shocks are more influential on the conditional volatility in high sentiment periods. In the literature, studies on information efficiency show that emerging markets have lower information efficiency (Bekaert and Harvey, 2002). Lower information efficiency leads to the result that information is incorporated into asset prices in longer time in emerging markets such as Turkey where information collecting, and processing are more costly for investors. As Brzeszczyński et al. (2015) state, investor sentiment may have more effects on this environment that is dominated by individual investors in the shortage of high-quality information.

2.3.2 Media coverage

Media and news coverage in newspapers is a commonly used passive attention measure in the literature. Media coverage could influence market valuation by influencing investors' behaviors. Media can also affect market activity by directing investor attention. Merton (1987) proposes a model of incomplete information in which investors are uninformed of a subset of stocks and do not use them in forming their portfolios. The results show that visibility in media increases a firm's investor base leading to increase in its market value and decrease in its expected return. Fang and Peress (2009) examine the hypothesis that claim mass media can lessen informational frictions that affects security prices. They use daily newspapers with nationwide circulation in the US and find that stocks with no media coverage earn higher returns than stocks with high media coverage. This effect is stronger among small stocks and stocks with high individual ownership, low analyst following, and high idiosyncratic volatility. These results are broadly consistent with Merton's (1987) theory suggesting stocks with lower investor recognition offer higher expected returns to compensate being imperfectly diversified. Kaniel et al. (2007) examine sample of mutual funds using daily newspapers in the US in the times before investors use less Internet for the information to investigate the role of media coverage in investment decisions of mutual fund investors. The findings show that media coverage has an impact on net investor flows into the fund and fund characteristics affect the probability of a news story. Solomon et al. (2014) use widely circulated national newspapers: The Wall Street Journal, the NewYork Times, the Washington Post, and USA Today and present evidence that media coverage tends to contribute to investors' chasing of past returns. The findings show that if the stocks were recently presented in the media, fund holdings with high past returns attract extra flows. Griffin et al. (2011) examine the information content of news announcements in 56 markets and show how the financial media affects international markets. The findings depict that in most of the developed markets a firm's stock price volatility is higher on days when the number of public news about the firm is higher whereas in many emerging markets number of news do not affect volatility. The results suggest that stock price reactions are different in cross-country level and this difference is caused by insider trading and differences in the quality of the news dissemination. Engelberg and Parsons (2011) separate the causal impact of media motivated by the observation that investors have local demand and tilt their portfolios towards geographically local stocks. The findings depict that local media coverage has predictability on local trading, after controlling for earnings, investor, and newspaper characteristics. Drake et al. (2014) use comprehensive dataset of business press and define business press as an information intermediary in the market. The study provides evidence that press coverage of the annual earnings announcement decrease cash flow mispricing.

2.3.3 Abnormal trading volume

Trading volume is commonly used passive investor attention measure in the literature. Barber and Odean (2008) find that unsophisticated investors are more likely to buy salient stocks due to limits on attention and short sales. This study shows that investors are net buyers of attention and buy stocks that are in the news or experiencing high abnormal trading volume or the ones with extreme one-day returns. Cheng et al. (2015) use prior turnover as the measure for investor attention to investigate the difference in stock price performance and show that firms with low investor attention have greater underreaction to repurchase announcements than firms with high attention. Hur and Singh (2017) find evidence for investor attention using two measures, abnormal trading volume and Google Search Volume Index. The findings show that stocks that reach maximum daily returns at the end of the month and have investor attention are mispriced and show greater reversals. Lin et al. (2014) focus on listed US firms with trading turnover as a proxy for investor attention and investigate whether investor attention and analyst coverage affect the diffusion of information. This study finds evidence that the effect of analyst coverage on stock synchronicity is higher when investor attention is high. When firms have less analyst coverage, they become more relied on investor attention to adjust information. Wang et al. (2018) use option trading volume as investor attention proxy and give an evidence for suggesting higher pre-earnings option trading helps to reduce stock market under-reaction. The findings also show that when pre-earnings option trading is high the initial stock market's response increases and the post-earnings announcement drift decreases. Using trading volume as a proxy for investor attention, Hou et al. (2009) show that price momentum profits are higher among high volume stocks and in up markets while post earnings momentum profits are higher among low volume stocks and in down markets.

2.3.4 Other measures of investor attention

This subsection reviews the studies that investigate the impact of other measures of investor attention on financial markets. These measures are advertising expenditure, analyst coverage, activity in Bloomberg terminals, firm size and stock spam messages.

Previous research proposes that advertising expenditure is a measure of passive investor attention. Advertisement activity facilitates the awareness of the product of the firm leading to an increase in the awareness of the same firm (Grullon et al., 2004; Ding and Hou, 2015). Using product advertisement of firms, Grullon et al. (2004) find that firms with higher advertisement expenditures have more investors and higher liquidity. Their results show that awareness and familiarity of consumers or investors with a firm affect its cost of capital and value. Managers also use advertising tool to attract consumers' or investors' attentions that influence stock returns and value of the company. Lou (2014) examines the implications of firm advertising and show that an increase in advertising expenditure is related to a contemporaneous increase in retail buying and abnormal stock returns and lower future returns. The findings report that inverted V-shaped pattern in advertising spending around insider sales is most consistent with managers' adjustments for the profitable temporary return effect. The findings of Bali et al. (2013) show that there are two possible reasons for stock market underreaction to liquidity shocks; limited investor attention and illiquidity. Stock size and analyst coverage are defined as proxies for investor attention. The study also compares the relation between liquidity shocks and subsequent returns in different investor attention subparts. The results show that the inattention-based part is more powerful for the longer-term return predictability even if investor inattention and illiquidity contribute to the underreaction in short term. Hirshleifer and Teoh (2003) use analyst following, firm size, and the fraction of shares owned by financial institutions as investor attention proxies in their theoretical models. These models display that limited investor attention can lead to underreaction to information and slow changes in prices.

Previous studies mostly focus on measuring individual investor attention, however the study of Rephael et al. (2017) uses news reading activity on Bloomberg terminals to measure institutional attention and suggests using this activity as a direct measure of abnormal institutional investor attention (AIA) because users of Bloomberg terminals are generally institutional investors. Their findings show that institutional attention responds more quickly to major news events than Google search volume and has a leading effect on retail attention. The study provides evidence for price drifts that follow both earnings announcements and analyst recommendation changes are driven by insufficient investor attention.

Analyst coverage and institutional ownership are commonly used proxies for investor attention in the literature. Several studies use different measures of investor attention to explain stock performance variables. Qian et al. (2017) use number of shareholders, analyst following, institutional ownership and number of employees as investor attention proxies to examine turnover with three different components as liquidity, firm specific uncertainty, and investor attention. The findings of the study provide evidence that turnover is positively related to uncertainty and investor attention and show that a positive relationship between turnover and price delay. Lee et al. (2019) measure investor attention using firm size, analyst coverage, institutional ownership, and media coverage. The results show that the returns of technology-linked firms predict focal firm returns. This effect is more pronounced if the firm receives lower investor attention.

Several studies use email endorsements of stocks, i.e. stock spam, as an investor attention proxy and investigate the association between stock spam mails and stock market. Stock spam include unsolicited emails that recommends stocks and spam messages are ubiquitous because spammers do not disclose sales figures (Bohme and Holz, 2006). The study of Bohme and Holz (2006) provide evidences for significant reactions of traded value and market valuation to spam campaigns in the short run. Nelson et al. (2013) argue that stock spam provides a natural quasi-experimental environment to investigate the effect of investor attention. This study includes a sample of firms that investors are not aware prior to becoming the target of a stock spam campaign. The findings of the study show that content of the spam message influences the predictability of market reaction to spam. Returns and volume at the spam date are higher for stocks targeted by spam emails that have optimistic target price projections with the information in previously issued company press release.

A detailed list of studies showing commonly used measures of investor attention proxies is given in Table 2.1. The table depicts that most of the direct Internet search as a measure of investor attention replaces the indirect investor attention proxies such as trading volume, media coverage and advertising expenditures over time. Table 2.1 also indicates that Google SVI is the most commonly used Internet search tool analyzed in previous literature. Indirect investor attention proxies such as trading volume, media coverage and advertising expenditures can be considered within information supply mechanism and these proxies are typically used to investigate the limited investor attention effect where Internet search volume proxy is used to measure information demand of investors.

Table 2.1 : Investor attention proxies in the literature.

SOURCE	MECHANISM	ATTENTION PROXY	COUNTRY	SCOPE
Dimpfl and Jank (2016)	Information demand	Internet Search (Google)	USA	DJIA
Mao et al. (2011)	Aggregating moods of the society	Internet Search (Google)	USA	DJIA, VIX, volume, gold
Vozlyublennaia (2014)	Information demand	Internet Search (Google)	USA	DJIA, S&P 500, Nasdaq composite index, 10-year treasury note yield index, gold index, commodities index
Da et al. (2015)	Aggregating moods of the society	Internet Search (Google)	USA	S&P 500 index
Preis et al. (2013)	Aggregating moods of the society	Internet Search (Google)	USA	DJIA index
Tantaopas et al. (2016)	Information demand	Internet Search (Google)	Asia Pasific	Asia Pasific indexes
Peltomaki et al. (2018)	Information demand	Google SVI, Abnormal Trading Volume	Emerging markets	MSCI emerging market index, S&P 500, MSCI emerging market currency index
Da et al. (2011)	Information demand	Internet Search (Google)	USA	Russell 3000 companies
Joseph et al. (2011)	Information demand	Internet Search (Google)	USA	S&P 500 companies
Mondria and Wu (2011)	Information demand	Internet Search (Google)	USA	S&P 500 companies
Vlastakis and Markellos (2012)	Information demand	Internet Search (Google)	USA	DJIA companies
Bijl et al. (2016)	Information demand	Internet Search (Google)	USA	S&P 500 companies
Ding and Hou (2015)	Information demand	Internet Search (Google), news, advertising expenditure	USA	S&P 500 companies
Takeda and Wakao (2014)	Information demand	Internet Search (Google)	Japan	Nikkei 225 companies
Aouadi et al. (2013)	Information demand	Internet Search (Google)	France	CAC 40 companies
Bank et al. (2011)	Information demand	Internet Search (Google)	Germany	Xetra-listed companies
Swamy et al. (2019)	Information demand	Internet search (Google)	India	S&P BSE 500 companies
Ying et al. (2015)	Information demand	Internet Search (Baidu)	China	A-share listed companies
Zhang and Wang (2015)	Information demand	Internet Search (Baidu)	China	ChiNext market companies
Zhang et al. (2013)	Information demand	Internet Search (Baidu)	China	30 companies from ChiNext, the SME Board and the Main Board
Lawrence et al. (2016)	Information demand	Internet Search (Yahoo! Finance)	USA	US publicly - listed companies

Table 2.1 (continued) : Investor attention proxies in the literature.

SOURCE	MECHANISM	ATTENTION PROXY	COUNTRY	SCOPE
Fang and Peress (2009)	Media Coverage	Daily newspapers	USA	All companies listed on the NYSE and 500 randomly selected from Nasdaq
Kaniel et al. (2007)	Media Coverage	Daily newspapers	USA	Sample of mutual funds
Solomon et al. (2014)	Media Coverage	Daily newspapers	USA	Sample of mutual funds (open-end domestic equity funds)
Griffin et al. (2011)	Media Coverage	News archives	Global	Common equities around the world
Engelberg and Parsons (2011)	Media Coverage	Daily newspapers	USA	S&P 500 companies
Drake et al. (2014)	Media Coverage	News archives	USA	US publicly listed companies
Wang et al. (2018)	Limited investor attention	Option trading volume prior to earnings announcement	USA	US publicly listed companies
Lee et al. (2019)	Limited investor attention	Size, analyst coverage	USA	Common stocks excluding financial firms
Cheng et al. (2015)	Limited investor attention	Trading turnover	Taiwan	Stock repurchase programs in the Taiwan Securities Exchange (TWSE)
Lin et al. (2014)	Limited investor attention	Trading turnover	USA	NYSE, AMEX and Nasdaq listed securities
Hou et al. (2009)	Limited investor attention	Trading volume	USA	NYSE and AMEX listed securities
Barber and Odean (2008)	Limited investor attention	Abnormal trading volume, extreme one-day returns, media coverage	USA	Trading and position records for the investments of households
Hur and Singh (2017)	Limited investor attention	Abnormal trading volume, Google	USA	Common stocks on the NYSE, AMEX and Nasdaq
Rephael et al. (2017)	Institutional investor attention	Activity in Bloomberg terminals	USA	Accounts for individual investors
Grullon et al. (2004)	Limited investor attention	Advertising expenditure	USA	US publicly listed companies
Lou (2014)	Limited investor attention	Advertising expenditure	USA	Reduced sample using all US stocks
Bali et al. (2013)	Limited investor attention	Stock size, analyst coverage	USA	All common stocks traded on the NYSE, AMEX, and Nasdaq
Qian et al. (2017)	Limited investor attention	Analyst following, institutional ownership, number of shareholders and employees	China	China’s A-shares
Nelson et al. (2013)	Unsophisticated investors affected by spam campaigns	Stock spam message	USA	Targeted companies
Bohme and Holz (2006)	Unsophisticated investors affected by spam campaigns	Stock spam message	USA	Targeted companies

2.4 Behavioral Finance and Social Media Sentiment

Behavioral finance is a constantly evolving area that examines how psychology, cognition and irrational manner of investors affect their decision-making. Behavioral finance challenges the EMH (Fama, 1970) by highlighting the significant role of human emotion, sentiment and mood in financial decision-making. In EMH, financial markets are efficient with public and private information that is fully incorporated into prices stating that individual investors cannot consistently beat the market. As the main assumption of EMH, investors are fully rational, and their decisions are based on maximizing wealth. In real life, research studies on behavioral finance and economics claim that irrational investors affect asset prices (Lee et al., 1991; Lee et al., 2002; Baker and Wurgler, 2007; Ho and Hung, 2009).

The literature on behavioral finance mainly focuses on studies that propose human behavior factors like social dynamics (Shiller, 1984), social mood (Nofsinger, 2005), investor sentiment (Baker and Wurgler, 2006) or psychological factors (Fenzl and Pelzmann, 2012) as the source of market volatility and anomalies.

Shiller (1984) suggests that social dynamics arising from observations of human nature and participants in the market is likely to influence speculative asset price movements. Shiller (1984) provides evidence that social dynamics, fashion or fads are the important cause of speculative asset price movements. Fashions are unpredictable in nature, caused by the overreaction of ordinary investors to earnings news or dividends news. Decision makers are affected by human behaviors such as social mood (Nofsinger, 2005). According to Nofsinger (2005) social mood determines the forms of decisions made by consumers, investors and corporate managers where optimism and pessimism in society is reflected by the emotions. Stock market reflects the social mood since stock transactions have emotional nature. The stock market is influenced by the social mood positively. Therefore, stock market is identified as a measure of social mood. Increases in stock market valuation is a measurement for a rising (optimistic) mood where declining stock valuation indicates a declining social mood. There is a time lag between the rising stock market and the economic activity, but the time lag is asymmetric between increases and decreases in mood. The author suggests that investors do not have negative social mood because the stock market has fallen where the market has fallen because people have negative social mood. Stock market moves faster to reflect changes in social mood where other financial actions such as M&A activity take longer to reveal. Another important statement of Nofsinger (2005) is that extremely positive or negative social moods are associated with extreme behaviors which can cause stock market bubbles. Fenzl and Pelzmann (2012) state that nonrational herding impulses (mainly mass psychological dynamics of human aggregate behavior) of financial market actors in complex and uncertain conditions cause non mean reverting dynamism in financial markets. The study emphasizes that collective behavior and social interactions between market participants and social environment leads to unintentional aggregate outcomes such as financial booms.

Behavioral finance literature examines two types of investors: irrational traders who are prone to exogenous sentiment and rational arbitrageurs. In noise trader approach, all investors are not rational and their demand for risky assets is affected by their beliefs or sentiments. Noise traders are examined by Kyle (1985) and Black (1986) assuming investors are classified into two groups as rational informed traders and uninformed noise traders with an irrational behavior. In this context, one of the significant contributions of behavioral finance research is the existence of investors with bounded rationality. De Long et al. (1990) present a model which shows that risk created by the unpredictability of unsophisticated investors' beliefs reduces the attractiveness of arbitrage. Rational arbitrageurs have short horizons and limited risk-bearing capacity that leads to large difference between market prices and fundamental values where noise traders who bears a disproportionate amount of risk enables them to earn a higher expected return. The actions of irrational investors lead to a change in investor sentiment.

Behavioral patterns of individual investors influence financial markets. The literature discusses these patterns analyzing the changes of investor sentiment in financial markets and asset pricing. Barberis et al. (1998) develop a model of investor sentiment to show the impact of investor overreaction and underreaction to public information for parameters such as post-earnings announcement drift and momentum. This study describes market inefficiencies focusing on how investors form their beliefs and defines the links between conservatism and representativeness heuristic to explain under-reaction and overreaction.

Institutional investors can follow traded stocks using professional tools, however social media helps individual investors to access the information easily (Chen et al., 2014, Behrendt and Schmidt, 2018). Individual investors are defined as noise traders who have psychological biases (Kyle, 1985; Black, 1986). Easley and O'hara (1987) and Hirshleifer and Teoh (2003) define individual investors as uninformed traders. Institutional investors can exploit the behavior of irrational investors as sentiment driven noise traders who use social media platforms.

Investor sentiment have been measured in various approaches in the literature. In traditional models, sentiment is measured by observing analyst estimates, survey data, news stories, put/call ratios and relative strength indicators. These approaches consist of financial market-based measures such as volume, VIX index, surveys such as consumer confidence index, non-economic factors such as news and weather conditions and textual sentiment data from social media such as Twitter and Facebook. Investor sentiment has been identified as a fundamental factor in determining asset prices. Many studies in the literature examine how changes in investors' sentiment affect stock prices. Baker and Wurgler (2006) construct a sentiment index that combines common variation in six proxies: the closed-end fund discount, NYSE share turnover, the number of average first day returns on IPOs, the equity share in new issues, and the dividend premium. The study of Baker and Wurgler (2006) challenges the view of classical finance theory that states investor sentiment does not play any role in the cross-section of stock prices and returns. Baker and Wurgler (2006) state that investor sentiment affects securities more that are highly subjective valued and difficult to arbitrage stocks. They find that when proxies for sentiment are initially low, subsequent returns are relatively high for small, young, high volatile, unprofitable, non-dividend-paying, extreme growth and distressed stocks that earn relatively low subsequent returns in high sentiment environment. Huang et al. (2015) propose a new investor sentiment index to predict the aggregate stock market return. Their findings support that investor sentiment is more predictive for the aggregate stock market than previous commonly used proxies. They find that the return predictability of investor sentiment is originated from investors' biased belief about future cash flows instead of discount rates. The new aligned investor sentiment measure can forecast stock returns either at the aggregate level or portfolio level.

In line with the noise trader models and a sentiment-based theory of return comovement, Barberis et al. (2005) and Fisher and Statman (2000) analyze the sentiment of three groups as Wall Street strategists; individual investors, newsletter writers and large investors, and they show that groups' sentiments are not alike. They suggest that sentiment can be used for tactical asset allocation. The impact of retail trading patterns on stock return comovement is examined by Kumar and Lee (2006) using a large data set of retail trades in the US. The study shows that the trading activities of retail investors contain a common directional component, and this result suggests that changes in portfolio-level retail sentiment may lead to comovement in stock returns. With the direct measurement of investor sentiment using retail investor trading activities, the results report that the stocks in the portfolios have higher excess returns. Investor concentration on firms that are smaller, low priced and have higher book to market, lower institutionally ownership ratio and high arbitrage costs are more sensitive to changes in retail investor sentiment.

Market sentiment or investor attention represent main attitude of investors. Sentiment is defined as optimism or pessimism, bullish versus bear behaviors in the literature. Using a media content as a measure of the interaction between the media and the stock market, Tetlock (2007) shows that high media pessimism predicts low stock prices and high market trading volume is predicted by an unusual high or low pessimism in line with theoretical models of noise and liquidity traders. The study does not support the idea claiming that media content as a proxy for new information about fundamental asset values where pessimism measure grabs temporary decreases in returns.

Research on social media data as a measure of the complex behavior of the investors have been increasing in recent years. In the literature, commonly the impacts of traditional media or news tone is investigated as the frequency of negative words used in an article (Tetlock et al., 2008; Tetlock, 2011). On the contrary to traditional media sources, social media is an interactive platform. Social media sentiment or Twitter sentiment is essential for analyzing the positive and negative texts of comments on stocks as a direct measurement. Social media sentiment can capture investors with bounded rationality (De Long et al., 1990; Shleifer and Vishny, 1997; Barberis et al., 1998). These less rational investors are typically individual investors. Thus, the role of social media sentiment will increase as a result of an increase in the number of individual investors caused by technological developments and growing number of trading platforms. Social media provides the opportunity to collect direct data about these human factors at the aggregate level.

2.5 Measures of Social Media Sentiment

This subsection reviews the main measures of social media sentiment and their empirical findings documented in the literature. The linking mechanism between financial markets and aggregated investor behavior directly measured by social media has many practical implications in investment decisions. Investor attention is typically measured by quantitative data such as the search volume index, number of news, trading volume and number of analysts, while social media sentiment examines the content and tone of the texts that investors share. Social media is an interactive environment in which people share ideas, emotions and moods that allow people to share information and respond to shared information. Therefore, the information obtained through social media can be analyzed not only quantitatively but also qualitatively and social media plays an important role in understanding the behavior of the society and information dissemination.

Social media applications and websites such as Twitter and Facebook allow people to interact each other and share their ideas. This subsection reviews most commonly used social media platforms with the aim of having and sharing an information on securities and finance: Twitter and Stocktwits, Facebook and stock message.

2.5.1 Twitter sentiment

There are few studies focusing on firm-specific Twitter sentiment methodology due to the difficulties in analyzing big-data in social media. However, with the advances in data analytics tools, recent empirical studies in the literature increasingly focus on the analysis of the relation between asset prices and investor sentiment obtained from social media. Bukovina (2016) surveys the literature on the link between social media and capital markets and emphasizes the role of social media big data as a tool to track the aggregate behavior of people.

Twitter sentiment is one of the direct ways to measure sentiment in the stock market. Recent studies mainly focus on index level sentiment analysis in the US market. Zhang et al. (2011) analyze Dow Jones, NASDAQ and S&P 500 Indexes, and show that the positive and negative moods on Twitter has negative correlation with indexes where it has a significant positive correlation with VIX. Bollen et al. (2011) construct measurement of collective mood states derived from large scale Twitter feeds and they find that public mood states, measured by the OpinionFinder and GPOMS mood time series, has predictive power of changes in Dow Jones Industrial Average (DJIA) closing values. Mao and Bollen (2011) compare sentiment tracking methods (Twitter sentiment, negative news sentiment and Tweet & Google Search volumes of financial terms) for financial prediction of market indices such as the DJIA, trading volumes, and market volatility (VIX), and gold prices. Their results depict that weekly Google SVI on financial search queries carry a predictive value. An indicator of Twitter sentiment and the frequency of occurrence of financial terms on Twitter in the previous 1-2 days are also statistically significant predictors of daily market log return. In recent studies, investor sentiment is investigated by taking advantage of natural language processing techniques to analyze sentiment of the society. Zhang et al. (2016) calculate the daily happiness sentiment using Twitter and investigate the effect of this sentiment in eleven international stock markets. The findings of the paper show that correlation coefficient between happiness sentiment and index return, the coefficient between index return, and the range-based volatility is higher in high happiness group, and happiness sentiment explains index return better in these groups. Granger-cause results depict that daily happiness causes index return.

A large set of empirical studies investigate the impact of Twitter sentiment on individual stock returns and trading activity. Liew and Wang (2016) investigate the cross-sectional relationship between the IPO's first day returns and Twitter sentiment using iSENTIUM LLC sentiment data. The findings of the study indicate that IPO sentiment the day before can signal and predict IPO's first-day returns. Sprenger et al. (2014) analyze S&P 100 companies using computational linguistics on stock-related daily messages and find associations between tweet sentiment and stock returns, message volume and trading volume, as well as disagreement and volatility. The findings indicate that increase in bullishness is a proxy for positive investor sentiment indicated by rising stock prices. Ranco et al. (2015) examine DJIA index companies and they find a significant dependence between the Twitter sentiment and abnormal returns at the peaks of Twitter volume. Bartov et al. (2017) focus on Russell 3000 firms and hypothesize whether individual tweets about a company's prospects can predict its earnings and the stock price reactions. Their findings reveal a positive relation between the aggregate opinion and the immediate abnormal stock price reaction to the quarterly earnings announcement. Focusing to the number of followers mechanism in Twitter, Sul et al. (2017) analyze S&P 500 stocks and show that Twitter sentiment about a specific firm from users with less than the median of the sample, have a significant effect on the stock's returns on the following day, 10 days and 20 days. Twitter sentiment from users with fewer than median followers that were not retweeted have the highest impact on future stock returns. Leitch and Sherif (2017) investigate the impact of Twitter sentiment about the announcement of Chief Executive Officer (CEO) succession on stock returns for a sample of firms that are listed on the indexes of S&P 100 and FTSE 100. The results provide evidence supporting the idea that Twitter sentiment is negatively contemporaneous related to stock returns and CEO succession announcements and CEO age is at announcement positively related to stock returns. Using social media metrics, Liu et al. (2015) suggest to group firms based on their Twitter accounts and predict stock comovement for US stocks. The results depict that returns of firms with official Twitter accounts have much higher comovement than those without Twitter accounts. Social media groupings also increase the accuracy of comovement prediction better than industry categories. The study of Yu et al. (2013) is different from earlier studies because the study compares the impacts of social media and conventional media on short term stock market performances for US companies. Blogs, forums, and Twitter are selected as social media platforms whereas major newspapers, television broadcasting companies, and business magazines are selected as means of conventional media. The results show that both social and conventional media have effects on stock performance while the effect of social media is higher on the daily basis. Using local Twitter activity, Baik et al. (2016) find that local Twitter users' tweets about the firms that have high information asymmetry and Twitter activity is positively related to trading volume for local stocks. Future stock returns and subsequent earnings announcement returns are predicted by the negative tone of local tweets that are also positively related to higher bid-ask spreads and lower market depths. Focusing on S&P 1500 firms, Crowley et al. (2018) investigate the dynamic information dissemination role of Twitter and find that firms are inclined to disclose corporate events on Twitter and select Twitter to post financial disclosures more frequently around financial firm events such as earnings announcements. The results provide evidences supporting the ideas that firms are more likely to disseminate significant good or bad news on Twitter and firms with limited attention are more inclined to exercise discretion facilitating future financial tweets and use of media and links.

Several recent studies provide evidences for the intraday effects of Twitter sentiment. Behrendt and Schmidt (2018) examine the relation between individual-level stock return volatility measured by absolute 5-minute returns and Twitter sentiment for DJIA constituents. In the study, intraday Twitter sentiment and Twitter publication count data for all DJIA constituents are obtained from Bloomberg. Their findings show that there are significant feedback effects of return volatility to Twitter sentiment as well as Twitter count and vice versa in a bivariate VAR framework. However, they emphasize that estimated coefficients are small, and the effects do not have a significant economic impact. Renault (2017) proposes an intraday, half-hour interval, lexicon of words on the bullishness or the bearishness of the stock market using StockTwits. The study shows that the sentiment is driven by the change in the sentiment of novice traders. The study provides evidence for the idea stating that investor sentiment forecasts intraday stock index returns, and the first half-hour change in sentiment predicts the last half-hour S&P 500 index ETF return. Using intraday sentiment from Thomson Reuters MarketPsych Indices based on a textual analysis of sources from news wires, internet news sources, and social media, Sun et al. (2016b) support that return predictability is most likely driven by noise traders. The authors show that lagged half-hour investor sentiment predicts intraday S&P 500 index returns and this effect persists in at least the last two hours of a trading day. Li et al. (2018) examine S&P 100 companies in daily, Apple stock in 15-min basis intraday analysis using stock related tweets and computational linguistics. The results of the study show that sentiment of messages is positively associated with contemporaneous daily abnormal stock returns and message volume predicts 15-min subsequent returns, trading volume, and volatility. Disagreement in tweets has a positive effect on stock features. The trading strategy in the paper indicates that it is possible to have profitable strategy even after transaction costs are included.

Wisdom of crowds represents the collective information of a group of individuals that results in better predictions than those of an individual member or single expert. Recent studies on the Wisdom of crowds examine social media sentiment and highlight the importance of the aggregate opinion from individual tweets in predicting asset returns. Azar and Lo (2016) show that tweets containing information about stock prices and tweets on the Federal Open Market Committee around these meetings is informative to predict future returns. After gathering tweets between 2007 and 2014 that mention the terms “FOMC” or “Federal Reserve”, “Bernanke” or “Yellen” they associate the outcome of each tweet with a polarity score and report that this score can be used to predict the returns of the Center for Research in Security Prices alue-Weighted Index. Karagozoglu and Fabozzi (2017) use sentiment data provided by PsychSignal from Twitter and StockTwits and Hive-Mind market volatility detection system. With investor sentiment and market volatility data on S&P 500 Index, the authors show that information in the volatility sentiment from social media can be used to create profitable trading strategies for stock market volatility.

StockTwits is a social media platform where investors, traders and market participants share ideas. The platform developed in 2008 currently has 2 million registered members, market professionals and public companies (Url-4). StockTwits is an investor platform where users share short messages about a particular stock or index using a $ symbol before the ticker symbol. Liew and Budavari (2017) identify the Social Media Factor and show that security characteristics derived from social media information significantly explains the daily returns for the sample of 15 stocks. Their social media factor which uses daily tweet sentiments provide significant characteristics. Employing the Fama–French five factors model, the residuals are examined as two separate components: Social Media Factor and the original residual. Their results suggest that the Fama–French five-factor model should be followed as a six-factor model, with the sixth factor of the Social Media Factor. Sun et al. (2016a) investigate the importance of textual information in StockTwits to predict the stock market. The distinction of this study is based on the model which leverages market information included in high-volume social media data rather than news articles without the need to evaluate the sentiment in each message.

2.5.2 Facebook sentiment

Facebook is one of the most commonly used social media platform in the world (Url-5). In 2009, Facebook constructed a “Gross National Happiness” index that consists of a multidimensional model by using thirty three indicators based on nine sections: psychological wellbeing, health, education, time use, cultural diversity and resilience, good governance, community vitality, ecological diversity and resilience, and living standards (Siganos et al., 2014). Several studies use this happiness index to investigate the effect of investor sentiment on stock market indicators. Siganos et al. (2014)

investigate the impact of daily Facebook sentiment on trading behavior in twenty international markets. The results indicate that sentiment is positively related to stock returns but this effect reverses in the subsequent weeks and causality exists from sentiment to stock returns. Karabulut (2013) proposes to use Facebook's Gross National Happiness (GNH) as a direct measure of investor sentiment. The findings of the study depict that GNH predicts changes in daily returns and trading volume, but the effect is temporary and reversed in next weeks, consistent with noise trader models. Siganos et al. (2017) proposes to use the distance between people with positive and negative sentiment using Facebook status updates for twenty international markets. Based on a divergence of sentiment, the results indicate that divergence of sentiment is positively associated with trading volume, volatility. These relations are more pronounced when investors are more likely to trade, and country-specific effects differentiates with market integration levels.

Few studies in the literature focus on the relation between individual level trading activity and Facebook activity. Siikanen et al. (2018) collect daily numbers of posts and related comments, likes, and shares from Facebook wall for the stock Nokia and investigates the relation between Facebook data and investors' decision making. The paper shows that less sophisticated investors, passive households and nonprofit organizations are more related to Facebook activity and inclined to decide to buy versus sell.

2.5.3 Stock message boards

Internet message boards are tools that investors spend considerable amount of time and effort posting and reading the messages. There are mixed results on the prediction of subsequent stock returns using public information on the Internet message boards. Antweiler and Frank (2004) report an evidence that rejects all message board talk is just noise and there exists financial relevant information. Their study focuses on more than 1.5 million messages posted on Yahoo! Finance and Raging Bull for forty-five companies in the Dow Jones Industrial Average and the Dow Jones Internet Index. Their findings show that stock messages help the prediction of volatility where the positive shock effects to message board posting predicts negative returns that are statistically significant but economically small. The results also indicate that disagreement among the posted messages is related to subsequent trading volume.

Using Yahoo! message boards, Wysocki (1998) finds that changes in daily posting volume is positively related to changes in stock trading volume and returns and overnight message volume has a predicting power in subsequent day trading volume and returns for US stocks. Das and Chen (2007) develop a methodology for small investor sentiment on stock message boards. The empirical results of their study show that tech sector message postings are associated with stock index levels, volumes and volatility. The study presents the algorithms that may be used to assess the impact of investor opinion and used to analyze the herding mechanism and market monitoring. Using Yahoo! Finance message board by a machine learning classification, Kim and Kim (2014) examine the association between stock message boards and stock market variables with causality tests. The results show that investor sentiment is positively affected by prior stock price performance, but investor sentiment does not forecast future stock returns. Using Yahoo! Finance message board, Jiang et al. (2014) suggest a stakeholder-based event analysis framework that uses online stylometric analysis to group the forum participants in stakeholder basis. The findings of this study indicate that some stakeholders grouped by the system has stronger market performances than the groups formed by other web forum users. Using messages posted on TheLion.com, Sabherwal et al. (2008) find that thinly traded micro-cap stocks with low institutional holdings and low analyst coverage are typically discussed stocks. Focusing on micro-cap stocks, the results of the study show that the number of messages posted predicts the abnormal returns on the subsequent day. Chen et al. (2014) use textual analysis of user-generated opinions and articles from Seeking Alpha, one of the most popular investor social media platforms in the United States. They find that the opinions on this website significantly predict future stock returns and earnings surprises by controlling other traditional advice sources, such as financial analysts and news media.

Twitter is considered as one of the most widely used microblogging social media platforms. There are differences between stock message boards (or blogging sites) and Twitter because of microblogging features. First, in microblogging sites people can update their thoughts more frequently. Thus, microblogging platforms are more active than blogging sites even if there may be outdated information on stock message boards. A blogging site allows people to write unlimited words on a topic while a microblogging site allow people to post a content of limited words. Second, message boards classify postings for firms and can archive all postings related to specific stocks.

However, in microblogging sites like Twitter, postings or tweets have conversational characteristics and firms can be followed by stock tickers within these conversational postings. Third, mentioning, retweeting and following mechanisms are important for microblogging sites where tracking information diffusion is possible (Sprenger et al., 2014).

Detailed list of studies that shows commonly used measures of social media sentiment proxies is given in Table 2.2. This table depicts that most of the studies focus on US markets and the publication years of the studies based on stock message boards are quite older than publication years of the studies based on Twitter sentiment because microblogging platforms replace message boards as an updated version of posting platforms. Table 2.2 indicates that firm-specific sentiment or calculating bullishness or bearishness of the market using text mining from Twitter is the most commonly used mechanism to analyze complex behavior of the investors and the society.

2.6 Hypotheses

This subsection presents the hypotheses on the impacts of investor attention and social media sentiment on stock returns and trading activity.

To understand the effect of investor attention on stock returns, we investigate the following hypothesis. We examine how stock returns change with the changes in attention levels of investors measured by ASVI in stocks listed in Turkey. We also focus on the interaction effect of the firm size and investor attention in individual stock returns.

There are extensive evidences suggesting that individual investors have limited attention. Limited attention executes a constraint on the amount of information that investors can process and respond. Barber and Odean (2008) find that unsophisticated investors are likely to buy salient stocks due to limited cognitive capacities of investors. They show that investors are net buyers of attention and buy stocks, in the news, stocks experiencing high abnormal trading volume, and stocks with extreme one-day returns. Da et al. (2011) propose to use the Google SVI for stock ticker as a direct proxy of investor attention and introduce that search volume captures attention more properly than indirect attention proxies and mainly measures the individual investor attention. Building on the work of Da et al. (2011), this hypothesis use SVI on stock tickers as a direct proxy for investor attention. One might expect to find that stock returns of the subsequent week will be higher when investor attention, namely ASVI, is higher. This hypothesis supports the attention-induced price pressure hypothesis developed by Barber and Odean (2008) within an emerging country perspective. The high percentage of domestic individual investors in the total volume supports the idea claiming that it is important to examine the individual investor attention in Turkish stock market. Taken together, this hypothesis tests the idea of Da et al. (2011) claiming that the searches for ticker symbols serves as a valid proxy for investor attention, is useful for predicting stock returns in the short term. The price pressure hypothesis states excess demand of uninformed participants cause that prices temporarily diverge from their information-efficient values to be compensated and prices return to their efficient values over a short horizon (Scholes, 1972). We also test whether price pressure hypothesis due to individual buying activity induced by ASVI effect is more pronounced among small stocks since small stocks are more prone to larger price impact (Da et al., 2011). Our sample consists of all stocks in Borsa Istanbul all shares index instead of large cap stocks. We would expect a larger price increase with an increase in ASVI among smaller Turkish stocks. Therefore, first hypothesis based on investor attention is formed as follows.

Hypothesis 1: Investor attention measured by abnormal Google search volume index is associated with stock returns.

As the second main hypothesis, we test whether social media environment contains valuable information that is not fully incorporated in stock market performance indicators. To analyze the effect of Twitter activity and Twitter sentiment on stock returns and trading activity, we investigate two hypotheses. In the first hypothesis, we examine how the number of tweets or Twitter activity and sentiment affects trading activity or volume measured by abnormal turnover for stocks that are constituents of international indexes. The trading volume measure is abnormal turnover as in Tetlock (2011). Van Bommel (2003) argues that investors are motivated to tell their friends and environment about their investments to make them follow their actions and the reason behind this inclination is trying to gain reputation. The study states that spreading rumors increases the demand and price of a security. Therefore, people tend to post tweets about their trades. Wysocki (1998), Sprenger et al. (2014) and Li et al. (2018) find that the number of tweets or message volume predicts the following day trading volume. We would expect that an increase in the number of tweets or Twitter activity would be associated with higher trading volume and high Twitter activity would predict trading volume in the next day. Therefore, first sub hypothesis is formed as follows.

Hypothesis 2.a.: Increases in Twitter sentiment and activity is associated with higher trading volume.

In the second sub hypothesis, we investigate whether the Twitter activity and sentiment have impacts on stock returns. Social media sentiment can capture investors with bounded rationality (De Long et al., 1990; Shleifer and Vishny, 1997; Barberis et al., 1998). These less rational investors are typically individual investors. Thus, the role of social media sentiment would increase as a result of an increase in the number of individual investors caused by technological developments. Social media provides the opportunity to collect direct data about these human factors at the aggregate level. Investor sentiment and attention can be used as a direction signal for trading purposes. Intuitively, if there is positive information about a certain company, one might expect the company's stock price to rise, and if there is any negative information, the stock price would decrease.

Institutional investors can follow traded stocks with the help of professional tools, however social media helps individual investors to access the information easily (Chen et al., 2014, Behrendt and Schmidt, 2018). Individual investors are defined as noise traders who have psychological biases (Kyle,1985; Black, 1986). Easley and O'hara (1987) and Hirshleifer and Teoh (2003) define individual or individual investors as uninformed traders. Institutional or informed investors can exploit the behavior of irrational investors as sentiment driven noise traders who use social media platforms.

The increase in the number of tweets is an indication that new information has arrived on the market (Sprenger et al., 2014). Most of the tweets or messages denotes buy signals and an increase in the number of tweets would be associated with higher stock returns (Bartov et al., 2017; Sprenger et al., 2014; Chen et al., 2014; Li et al., 2018). DeMarzo et al. (2003) suggest a bounded rationality model in which individuals have persuasion bias and they fail to account for possible repetition in the received information. Their model proposes that social influence and well-connecting in the social network determines communication. Social media platforms such as Twitter can be given as an example of this model based on its follower mechanism. We would expect that an increase in the number of tweets or Twitter activity and the tone of tweets or Twitter sentiment would be associated with higher stock returns and high Twitter sentiment would predict stock returns in the next day. Therefore, the hypothesis is formed as follows.

Hypothesis 2.b.: Increases in Twitter sentiment and activity is associated with higher stock returns.

These hypotheses are constructed to investigate the research questions that examine whether investor attention measured by Google SVI has an impact on stock returns in Turkey, and whether social media sentiment and activity measured by Twitter are influential on stock returns and trading volume in multi-country context. The literature provides information showing that Google SVI has been used as a proxy for investor attention, but the literature on the impacts of search index on asset pricing in emerging markets is limited and no study has investigated the impact of direct investor attention in Turkey in emerging markets perspective. The impacts of social media proxies on stock markets, mainly stock indexes, are also investigated in the literature. However, the literature provides no evidence on the impacts of social media sentiment on stock markets in multi-country context using large number of stocks by comparing regional differences.

Table 2.2 : Social media sentiment proxies in the literature.

SOURCE	MECHANISM	SOCIAL MEDIA PROXY	COUNTRY	SCOPE
Renault (2017)	Bullishness (bearishness) of the market	StockTwits	USA	S&P 500 index ETF
Liew and Budavari (2017)	Firm-specific sentiment	StockTwits	USA	Sample of 15 companies
Sun et al. (2016)	Firm-specific sentiment	StockTwits	USA	S&P 500 companies
Karagozoglu and Fabozzi (2017)	Firm-specific sentiment	StockTwits and Twitter	USA	S&P 500 companies
Zhang et al. (2011)	Aggregating moods of the society	Twitter	USA	DJIA, Nasdaq and S&P 500 indexes
Bollen et al. (2011)	Aggregating moods of the society	Twitter	USA	DJIA index
Sun et al. (2016)	Bullishness (bearishness) of the market	Twitter	USA	S&P 500 index ETF
Azar and Lo (2016)	Tweets mentioning FOMC meetings	Twitter	USA	CRSP value-weighted market index
Zhang et al. (2016)	Aggregating moods of the society	Twitter	11 countries	11 international stock market benchmark indexes
Behrendt and Schmidt (2018)	Firm-specific sentiment	Twitter	USA	DJIA companies
Li et al. (2018)	Firm-specific sentiment	Twitter	USA	S&P 100 companies-daily, only Apple Inc. intraday
Liew and Wang (2016)	Firm-specific sentiment	Twitter	USA	325 IPOs going public on the NYSE or Nasdaq
Baik et al. (2016)	Firm-specific sentiment	Twitter	USA	Randomly selected 1044 companies
Ranco et al. (2015)	Firm-specific sentiment	Twitter	USA	DJIA index companies
Bartov et al. (2017)	Firm-specific sentiment	Twitter	USA	Russell 3000 companies
Sprenger et al. (2014)	Firm-specific sentiment	Twitter	USA	S&P 100 companies
Sul et al. (2017)	Firm-specific sentiment	Twitter	USA	S&P 500 companies
Liu et al. (2015)	Grouping Twitter accounts	Twitter	USA	Sample of companies listed on the NYSE and Nasdaq
Leitch and Sherif (2017)	Firm-specific sentiment	Twitter	USA, UK	Sample of companies in S&P 100 and FTSE 100 indexes
Crowley et al. (2018)	Firm-specific tweets	Twitter	USA	S&P 1500 companies
Karabulut (2013)	Aggregating moods of the society	Facebook	USA	Dow Jones, NYSE Composite, S&P 500 ETFs
Siganos et al. (2014)	Aggregating moods of the society	Facebook	20 countries	Country MSCI return indexes
Siganos et al. (2017)	Aggregating moods of the society	Facebook	20 countries	Country-level return indexes and trading volume
Siikanen et al. (2018)	Firm-specific Facebook activity	Facebook	Finland	Nokia
Chen et al. (2014)	Firm-specific opinions	Stock message board (Seeking Alpha)	USA	Sample US companies
Sabherwal et al. (2008)	Firm-specific posting volume	Stock message board (TheLion.com)	USA	Sample of 135 companies
Antweiler and Frank (2004)	Firm-specific opinions, posting volume	Stock message board (Yahoo! Finance, Raging Bull)	USA	45 stocks from DJIA and DJ Internet Commerce Index
Kim and Kim (2014)	Firm-specific sentiment	Stock message board (Yahoo! Finance)	USA	Sample 91 US companies
Das and Chen (2007)	Bullishness (bearishness) of the market	Stock message board (Yahoo! Finance)	USA	Sample of 24 tech-sector companies in the Morgan Stanley High-Tech Index
Wysocki (1998)	Firm-specific posting volume	Stock message board (Yahoo! Finance)	USA	Sample of 50 companies

3. DATA AND METHODOLOGY

This section explains the methodology of the Fama and MacBeth (1973) regression approach used in this thesis. In asset pricing theories, risk factors such as size and ratio are widely used to explain asset returns. The Fama-MacBeth regression is a practical way of testing how these factors affect portfolio or asset returns and is relevant as it is commonly used in asset pricing models in analyzing the mechanism between stock return and risk. The Fama and MacBeth (1973) model, which was developed by Fama and MacBeth (1973), is widely used in finance literature to investigate the relationship between expected returns and factor coefficients. The model is used in asset pricing because it is compatible to work with panel data and multiple assets across time. The model allows the coefficients of explanatory variables to change over time.

The Fama-MacBeth regression is a two-stage procedure used to test the CAPM using time series of cross-sections. This procedure begins with the estimation of cross-sectional regressions and each portfolio's return is regressed on one or more factor time series. In the first step, the cross-section of returns is regressed against the factor exposures for each time and it gives a time series of risk premia coefficients for each factor. In the second step, the time-series averages of the coefficients in the cross-sectional regressions are calculated. The aim is to find the premium from exposure to the factors.

Fama-MacBeth procedure as defined in Url-6 is provided below.

(i) Run time series regressions to get betas,

$$Rt_t^{\mathrm{ei}} = a_i + \beta_{\mathrm{i}}' f_t + \varepsilon_t^i \text{ , } t = 1, 2, \ldots T \text{ for each } i \tag{3.1}$$

(ii) Run cross sectional regression at each time period,

$$Rt_t^{\mathrm{ei}} = \beta_{\mathrm{i}}' \lambda_t + \alpha_{it},\ i = 1, 2, \ldots N \text{ for each } t \tag{3.2}$$

(iii) Then, estimates of λ and α are the averages across time,

$$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_t, \quad \hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\alpha}_{it} \tag{3.3}$$

In this model, the standard errors are adjusted for cross-sectional dependence. This is generally not challenging when the number of cross-sectional units is large and a time series for cross-sectional units is smaller. In this thesis, we find time-series averages of the coefficients and their standard errors which can be corrected for time-series dependence using Newey and West (1987) standard errors.

3.1 Effects of Investor Attention on Stock Returns: Evidence from Borsa Istanbul

In the first hypothesis of the thesis, we investigate the impacts of investor attention measured by Google SVI on individual stock returns in listed stocks in Turkey. We start with 481 stocks ever involved as a constituent in the Borsa Istanbul all shares index in Turkey in sample period, from April 2013 to September 2017, to remove survivorship bias and the effect of adding and removing stocks to the index.

Table 3.1 : Definitions of variables in investor attention models.

Variable	Definition	
Ret	Raw stock return	Weekly stock returns
AbRet	Abnormal return	DGTW adjusted abnormal return (Daniel et al.,1997)
SVI	Google SVI	Search frequency from Google Trends based on stock ticker
Name_SVI	Google SVI on firm name	Search frequency from Google Trends based on firm name
ASVI	Abnormal attention	The log of SVI during the week minus the log of median SVI for the previous 8 weeks
Size	Market value	The log of stock’s market capitalization in week t-1
BM	Book to market ratio	The book value divided by market capitalization of the stock in week t-1
Abnturnover	Abnormal turnover	The log of turnover relative to mean of last 52 weeks turnover
Lturnover	Turnover	The log of stock turnover in week t-1
Volatility	Stock return volatility	Standard deviation of daily stock returns for the week t-1
News	Number of news stories	The log of one plus number of stories published on the most recent week from Bloomberg

Google SVI returns zero for tickers that are rarely searched. Zero abnormal SVI values are eliminated to have valid SVI results. The sample contains firms for which more than 15 weeks of SVI are provided to eliminate the stocks with few observations. After these eliminations, the sample consist of weekly observations of 313 Borsa Istanbul all shares index stocks, 42,902 weekly firm observations during 232 weeks in Turkey between April 2013 and September 2017. All variables used in this study are defined in Table 3.1.

3.1.1 Proxies of investor attention

In this subsection, other proxies of investor attention are examined and compared SVI to other most common measures for attention in the literature. Investor attention is measured by different methods such as trading volume (Hou et al., 2009), extreme returns (Barber and Odean, 2008), media coverage (Fang and Peress, 2009). In this context, we select abnormal returns (Barber and Odean, 2008; Da et al., 2011; Ying et al. 2015), abnormal turnover (Da et al., 2011; Hou et al., 2009; Lin et al., 2014) and number of news (Fang and Peress, 2009; Engelberg and Parsons, 2011). The relation between direct investor attention measure, Google SVI and indirect investor attention measures (abnormal turnover, absolute abnormal return and the number of news) as in Da et al. (2011) is investigated to observe the capturing and likely effects of direct investor attention (ASVI). These variables are based on the assumption that investors increase their attention when there is extreme return or volume and large number of news in the media about the firm. However, extreme returns or volume may be factors that do not attract investors' attention, and newspaper articles or news do not necessarily increase investor interest unless investors read it (Da et al., 2011).

3.1.2 Google search volume index (SVI)

Google Trends (Url-7) is a website of Google that analyzes the popularity of top search queries and provides search volume index data from 2004 to present. Google Trends gives search volume index that is a standardized score between 0 to 100 where the maximum search volume is scaled to 100. Google Trends provides relative data by giving the highest 100 score to the absolute searches in the interval and determining the scores of the remain part with the relative score of the highest level. Google search data is available on a daily basis for maximum 90-day periods and on a weekly basis for maximum 5-year periods and on a monthly basis for more than 5-year periods. 90-day daily data period may have seasonal effects and may not reflect the investor behavior. In addition, it may be difficult to catch investor attention in monthly data.

Figure 3.1 demonstrates an example of SVI output obtained from Google Trends (Url-7) for a 5-year period for the term "GARAN", the stock ticker of "Türkiye Garanti Bankası A.Ş.". Google Trends define this index as "Numbers show search interest relative to the highest point on the chart for the given region and time. 100 is the peak point for the term where 50 means that the search query is half as popular. 0 means there was not enough data for this search query".

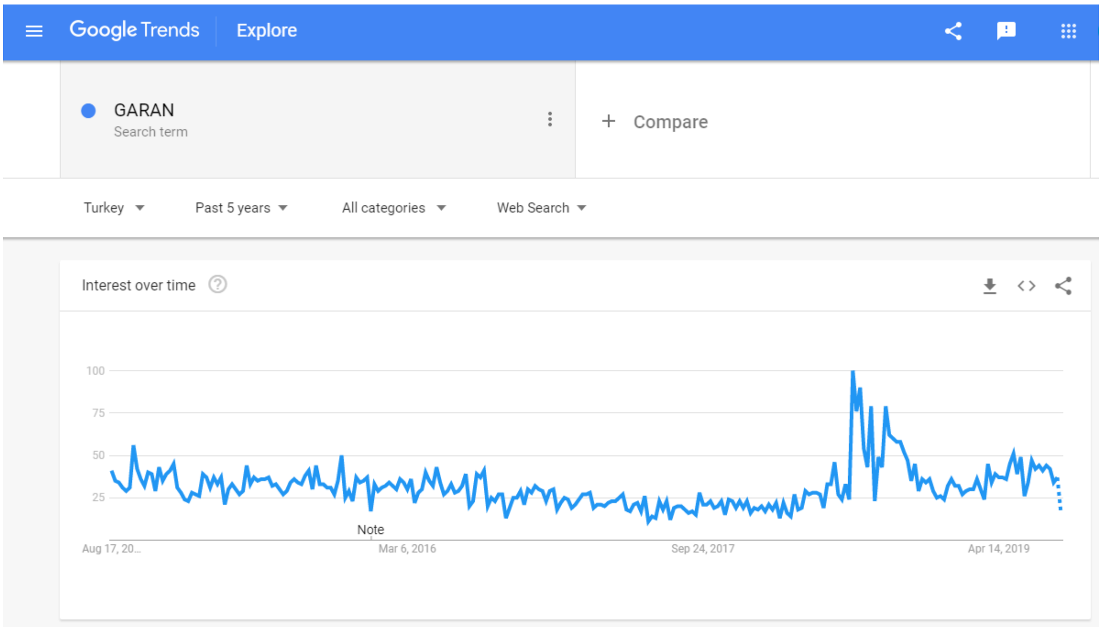

Figure 3.1 : Google search volume index for the term "GARAN".

Weekly investor attention of SVI data are collected from the Google Trends website. We use Google SVI as a proxy of investor attention and focus on ticker-based search to eliminate generic meaning search terms. People may search for a firm name for various reasons such as getting information on products, store locations, or job openings (Da et al., 2011). Search queries on firm name is based on how the individuals have searched the firm name and it may be subjective. Since we study the effect of investor attention on asset pricing, we want to focus on the individuals who are interested in investing. Thus, we choose to use firms' stock tickers which are uniquely assigned. Google Trends provides search categorization option which includes investing category. However, SVIs on firms' names in investing category is useless for the Turkish stocks because of low search frequencies and several missing values. For these reasons, we use stock tickers as the search term in Google Trends and manually exclude generic meaning tickers in Turkish (e.g. SISE, KONYA). We only exclude stock tickers that have a generic meaning in Turkish since we choose search region as Turkey.

For the listed companies in Turkey, we obtain weekly investor attention data from Google Trends. Other stock specific variables are obtained from Bloomberg database.

We follow Da et al. (2011) methodology where abnormal search volume is defined as,

$$ASVI_{it} = \ln(SVI_{it}) - \ln[Med(SVI_{t-1}, \dots, SVI_{t-8})] \tag{3.4}$$

where $\ln(SVI_{it})$ is the log of SVI_{it} (Google SVI) for firm i in week t, and $\ln[Med(SVI_{t-1}, \dots, SVI_{t-8})$ is the log of the SVI_{it} median for the previous eight weeks. This procedure allows Google SVI to be robust against recent jumps and to remove low-frequency seasonalities, time trends and the effects of macroeconomic changes on attention.

We use abnormal turnover in Barber and Odean (2008) as,

$$Abnturnover_{it} = \ln(Turn_{it}) - \ln[Mean(Turn_{it-1}, \dots, Turn_{it-52})] \tag{3.5}$$

where $\ln(Turn_{it})$ is the log of stock turnover, $Turn_{it}$ for the week t, and $\ln[Mean(Turn_{it-1}, \dots, Turn_{it-52})]$ is the log of the mean of $Turn_{it}$ over the previous 52 weeks.

3.1.3 SVI and stock returns

We use two indicators, calculated with Daniel et al. (1997) characteristic-based benchmarks, for the dependent variable to describe stock returns: raw stock returns and abnormal return. First, we group each stock into quintiles based on market capitalization. Then the stocks in each market capitalization quintile are grouped into quintiles on book to market values. These 25 portfolios grouped by their market capitalization and book to market values are further grouped into quintiles based on their momentum. We define momentum as a stock's cumulative return from *t*-44 to *t*-4 on a weekly basis (Fama and French, 2012). Thus, finally we obtain 125 benchmark portfolios grouped by size, book to market and momentum. We use the weekly average return of these portfolios as a benchmark return. The equations that show the impact of SVI on stock returns are given as,

$$Ret_{it} = \alpha_i + \beta_{1i} ASVI_{it} + \sum_{k=1}^{K} \beta_{2ki} Control_{ki,t-1} + \varepsilon_{it} \quad (3.6)$$

$$Ret_{it} = \alpha_i + \beta_{1i} ASVI_{i,t-1} + \sum_{k=1}^{K} \beta_{2ki} Control_{ki,t-1} + \varepsilon_{it} \quad (3.7)$$

$$Ret_{it} = \alpha_i + \beta_{1i} ASVI_{i,t-1} + \beta_{2i} SizeASVI_{i,t-1} + \sum_{k=1}^{K} \beta_{3ki} Control_{ki,t-1} + \varepsilon_{it} \quad (3.8)$$

where Ret_{it} is raw return in week *t*, $ASVI_{i,t-1}$ is abnormal Google Search Volume Index for firm *i* in week *t*-1, *K* denotes the number of control variables. Control variables, $Control_{ki,t-1}$, include $Ret_{i,t-1}$ that denotes return in week t-1, $Volatility_{i,t-1}$ that denote standard deviation of last 7 trading days' returns, $Size_{i,t-1}$ that denotes the log of stock's market capitalization, $BM_{i,t-1}$ that denotes book value of equity divided by market capitalization, $News_{i,t-1}$ that denotes log of 1 plus the number of stories published on the most recent week, $Lturnover_{i,t-1}$ that denotes the change in log of stock turnover at *t*-1. As in Da et al. (2011), Ying et al. (2015), Mondria and Wu (2011), we use size, book to market ratio, stock turnover and volatility to control company-specific size, value, turnover and volatility effects and we expect that these effects may be positive on stock returns. As earlier studies in the literature suggest, news coverage has an impact on stock returns and the number of news variable is used as control variable to test whether SVI has significant effects beyond news (Fang and Peress, 2009; Da et. al., 2011; Ying et al. 2015). We expect the news variable to have significant and positive effect on stock returns.

3.2 Effects of Social Media Sentiment on International Stock Returns and Trading Activity

The main purpose of the second hypothesis is to investigate the effects of social media activity and sentiment on individual stock returns and trading activity in international stock markets.

3.2.1 Twitter sentiment

Twitter is an online social media service that allows users to send short 280-character messages called tweets. In Twitter, hashtags (#), at sign (@), and cashtags ($) are used as text modifiers to create structured tagging for a term. Cashtags are particularly used for stocks. Using these hashtags and cashtags in front of the term (e.g. #AAPL,

$AAPL), platform users can easily find related tweets. Users can reach the users' profiles and previous tweets by clicking on a username tagged with at sign. In recent years, Twitter has been one of the leading social networks around the world based on the number of active users. By the end of 2017, Twitter had 330 million monthly active users (Url-8). StockTwits is another social media platform where investors, traders and market participants share ideas. StockTwits is an investor platform where users share short messages about a particular stock or index using a $ symbol before the ticker symbol. The platform developed in 2008 has 2 million registered members, market professionals and public companies (Url-4). Bloomberg integrated Twitter feeds into its platform in April 2013. Bloomberg also started to generate Bloomberg Social Velocity (BSV) alerts to track a company where BSV scans tweets and StockTwits for so-called cashtags, and any mentions of the company's name. With this service, professionals or clients can see the overall volume of tweets and the mix of positive, negative and neutral comments, and details of individual Twitter postings (Url-11).

Twitter sentiment data used in this study is obtained from Bloomberg. Bloomberg uses the raw message feeds from both StockTwits and Twitter as inputs and apply a proprietary natural language processing algorithm to classify each tweet. This classification methodology is similar to the polarity score constructed by Azar and Lo (2016). The sentiment calculation process is defined as follows (Url-9). First, a human expert manually assigns a positive, negative or neutral score to each news story or tweet. Second, the annotated data is fed into machine-learning models, such as a support vector machine. When a new message arrives, the model automatically assigns a positive, negative or neutral score to each news story or tweet. Third, story-level sentiment is calculated where real time score is a categorical value, e.g., 1, -1 and 0 and confidence is a numerical value ranging from 0 to 100. Company-level sentiment is defined as the confidence-weighted average of story-level sentiment. Finally, company-level daily sentiment scores are the confidence-weighted average of the past 24 hours' story-level sentiments for both news and Twitter and are published every morning about 10 minutes before market open. Market open time is determined based on the composite exchange of the equity being traded for the company.

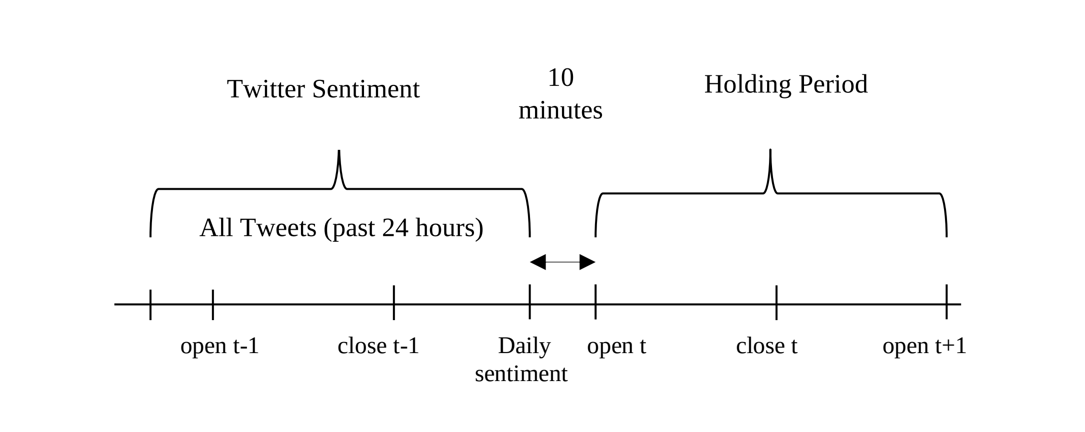

Figure 3.2 : Timeline of Twitter sentiment strategy.

Figure 3.2 shows the timeline for the Twitter sentiment trading strategy for open-to-open returns. The daily Twitter sentiment is a lagging indicator as it is an aggregation of the previous 24 hours' story-level sentiment. At the open (market open) of each day t, we sort stocks into decile portfolios based on their twitter sentiments on day t and compute the equal and value weighted return of a long short portfolio that buys stocks in the top decile with high Twitter sentiment and sells stocks in the bottom decile with low Twitter sentiment. For open-to-close return-based strategy, holding period is from open to close time at day t. This trading strategy is possible, but it is needed to act quickly to be able to decide and trade in a few minutes after observing the score. However, thanks to technological developments, trading in a few minutes after observing the sentiment score is not demanding in automated systems.

3.2.2 Twitter sentiment and trading activity

This study analyzes the effect of Twitter sentiment on stock returns for S&P 500, S&P 350 Europe and S&P Emerging Markets Core Index stocks. The sample period includes daily observations from January 2015 to the end of December 2017. The sample period begins in January 2015 because Twitter sentiment data was made available after this date for companies. Same companies listed on different exchanges are eliminated. The final sample includes 1,063 stocks consisting of 552 stocks from S&P 500, 372 stocks from S&P 350 Europe and 139 stocks from S&P Emerging Markets Core Index. All variables except Twitter and news sentiments and activity are used in terms of US dollars. All variables except Twitter and news sentiment are winsorized at the 1% level to minimize outlier effects. Detailed country breakdown information is given in Table A.1 in Appendix. All variables used in this study are defined in Table 3.2.

This subsection investigates the relation between Twitter activity, sentiment and individual stock trading volume. We follow the methodology of Sprenger et al. (2014) and Tetlock (2011) to investigate the association between Twitter activity, sentiment and trading volume.

Table 3.2 : Definitions of variables in social media sentiment models.

Variable	Definition	
Ret	Open to open return	Open-to-open (open-to-close) daily stock return.
AbRet	Abnormal return	Abnormal return, raw daily return minus S&P, S&P 350 Europe and S&P EM Core index returns as in Tetlock (2011).
Twitter	Twitter sentiment	Stock-specific Twitter sentiment from Bloomberg. Sentiment based on Twitter varies from -1 to 1, with -1 representing the most negative sentiment and +1 representing the most positive sentiment over a 24-hour period.
Ntweet	Number of tweets	The log of 1 plus total number of tweets for the company over a 24-hour period.
News	News sentiment	Stock-specific news sentiment from Bloomberg. Sentiment based on stories varies from -1 to 1, with -1 representing the most negative sentiment and +1 representing the most positive sentiment over a 24-hour period.
Size	Market capitalization	Firm size, defined as the log of stock market capitalization on day t-1.
Abturn	Abnormal turnover	Firm's log turnover on day t minus its average log turnover on days t −5 to t −1 as in Tetlock (2011).
Vola	Volatility	Park volatility (Parkinson 1980) measure on day t.
Ret[−5,−1]	Cumulative return	Cumulative raw returns on days t – 5 to t – 1.
AbRet[−5,−1]	Cumulative abnormal return	Cumulative abnormal returns on days t – 5 to t – 1.
Vola[−5,−1]	Volatility	Park volatility (Parkinson 1980) measure averaged over days t – 5 to t – 1.
Illiq[−5,−1]	Illiquidity	Amihud's (2002) illiquidity measure averaged over days t − 5 to t – 1.

The trading volume is abnormal turnover ($AbTurn_{it}$), which is measured by firm i's log turnover on day t minus its average log turnover on days t-5 to t-1 (Tetlock, 2011). The equations that show the impact of Twitter activity and sentiment on trading volume, abnormal turnover are given as,

$$AbTurn_{it} = \alpha_i + \beta_{1i}Ntweet_{it} + \sum_{k=1}^{K}\beta_{2ki}Control_{kit} + \varepsilon_{it} \tag{3.9}$$

$$AbTurn_{it} = \alpha_i + \beta_{1i}Twitter_{it} + \sum_{k=1}^{K}\beta_{2ki}Control_{kit} + \varepsilon_{it} \tag{3.10}$$

$$AbTurn_{it} = \alpha_i + \beta_{1i}Ntweet_{i,t-1} + \sum_{k=1}^{K}\beta_{2ki}Control_{kit} + \varepsilon_{it} \tag{3.11}$$

$$AbTurn_{it} = \alpha_i + \beta_{1i}Twitter_{i,t-1} + \sum_{k=1}^{K}\beta_{2ki}Control_{kit} + \varepsilon_{it} \tag{3.12}$$

where $AbTurn_{it}$ is the trading volume, abnormal turnover (*AbTurn*), which is measured by firm *i*'s log turnover on day *t* minus its average log turnover on days *t*-5 to *t*-1. $Twitter_{i,t-1}$ is stock *i*'s specific Twitter sentiment for day *t*-1, $Ntweet_{i,t-1}$ is stock i's number of tweets for day *t*-1, $Control_{it}$ is firm *i*'s $Size_{i,t-1}$, log of market capitalization on day *t*-1 , Ret_{it} [−5,−1], cumulative returns in the previous week from day *t*-1 to *t*-5, $Vola_{it}$[−5,−1], Park volatility (Parkinson, 1980) measure on days *t*-5 to *t*-1, $Illiq_{it}$[−5,−1], Amihud's (2002) illiquidity measure averaged over days *t*-5 to *t*-1.

Stock price volatility is measured based on intraday high, P_{it}^{h} and low price, P_{it}^{l} for firm *i* for day *t* with Park volatility (Parkinson, 1980) defined as,

$$Vola_{it} = \sqrt{\frac{\left(\ln\left(P_{it}^{h}\right) - \ln\left(P_{it}^{l}\right)\right)^2}{4\ln(2)}} \tag{3.13}$$

Amihud's (2002) daily illiquidity measure for day *t* is defined as

$$Illiq_{it} = \frac{|Ret_{it}|}{Volume_{it}} \tag{3.14}$$

where $Volume_{it}$ is the stock's dollar volume. The daily illiquidity measure is multiplied by 10^6 consistent with the studies in the literature (Tetlock, 2011).

3.2.3 Twitter sentiment and stock returns

This subsection investigates the relation between Twitter activity, sentiment and individual stock returns. Tetlock (2011) examines the impact of firm-specific news on stock returns. We follow the methodology of Tetlock (2011) to investigate the impact of Twitter sentiment on stock returns and use daily cross-sectional regressions given in Fama and MacBeth (1973) as,

$$Ret_{it} = \alpha_i + \beta_{1i} Ntweet_{it} + \sum_{k=1}^{K} \beta_{2ki} Control_{kit} + \varepsilon_{it} \quad (3.15)$$

$$Ret_{it} = \alpha_i + \beta_{1i} Twitter_{it} + \sum_{k=1}^{K} \beta_{2ki} Control_{kit} + \varepsilon_{it} \quad (3.16)$$

$$Ret_{it} = \alpha_i + \beta_{1i} Ntweet_{i,t-1} + \sum_{k=1}^{K} \beta_{2ki} Control_{kit} + \varepsilon_{it} \quad (3.17)$$

$$Ret_{it} = \alpha_i + \beta_{1i} Twitter_{i,t-1} + \sum_{k=1}^{K} \beta_{2ki} Control_{kit} + \varepsilon_{it} \quad (3.18)$$

$$Ret_{it} = \alpha_i + \beta_{1i} Twitter_{i,t-1} + \beta_{2i} News_{i,t-1} + \sum_{k=1}^{K} \beta_{2ki} Control_{kit} + \varepsilon_{it} \quad (3.19)$$

where Ret_{it} is open-to-open (open-to-close) return for stock *i* from day *t* to *t*+1 (return for stock *i* from open to close at day *t*), $AbRet_{it}$ is open-to-open (open-to-close) abnormal return, raw return minus the return of S&P 500, S&P 350 Europe and S&P EM Core index as in Tetlock (2011). $Twitter_{i,t-1}$ is stock i's specific Twitter sentiment for day t-1, $Ntweet_{i,t-1}$ is stock *i*'s number of tweets for day t-1, $News_{i,t-1}$ is stock i's specific News sentiment for day t-1 , $Control_{it}$ consists of $Size_{i,t-1}$, firm i's log of market capitalization on day t-1, Ret_{it} [−5,−1], cumulative returns in the previous week from day *t*-1 to *t*-5, $AbRet_{it}$ [−5,−1], cumulative abnormal returns in the previous week from day *t*-1 to *t*-5, abnormal turnover $Abnturn_{i,t-1}$, firm *i*'s log of turnover on day t minus its average log of turnover on days *t*-1 to *t*-5, $Vola_{it}$[−5,−1], Park volatility (Parkinson, 1980) measure on days *t*-1 to *t*-5, $Illiq_{it}$[−5,−1], Amihud's (2002) illiquidity measure averaged over days *t*-1 to *t*-5.

4. RESULTS

This section gives empirical results of the main hypotheses and indicates their contribution to the investor attention and social media sentiment research area. First subsection gives evidences of the effects of investor attention measured by Google SVI on individual stock returns in Turkey. Second subsection reports the effects of social media activity and sentiment, mainly Twitter sentiment, on stock returns and trading volume in multi-country context.

4.1 Effects of Investor Attention on Stock Returns: Evidence from Borsa Istanbul

This subsection discusses the relation between SVI and other indirect measures of investor attention, and then investigates the empirical results on the effects of the abnormal SVI on individual stock returns. Lastly, the practical implications of the trading strategy on ASVI are discussed.

Table 4.1 presents descriptive statistics for the variables used in the regressions. *Ret* is weekly stock returns, *AbRet* is weekly is DGTW adjusted abnormal return, *SVI* is search frequency from Google Trends based on stock ticker, *ASVI* is the natural log of SVI during the week minus the natural log of median SVI for the previous 8 weeks, *Size* is the log of ratio of stock's market capitalization in week *t*-1, *BM* is the ratio of book value to market capitalization or value of the stock in week *t*-1, *Abnturnover* is the log of turnover relative to median of last 52 weeks turnover, *Lturnover* is the change in log of turnover in week *t*-1, *Volatility* is the standard deviation of the daily stock returns for the week *t*-1, *News* is the log of 1 plus the number of stories published on the most recent week. The mean value of *SVI* per week is 26.056, indicating that the average popularity search volume index of the sample firm is given as 26.056 per week on Google searches. The mean value of *ASVI* per week is 0.164, showing that log value difference from last eight weeks *ASVI* is positive and increasing popularity in searches. 10th and 90th percentile values for SVI is 7 and 53, respectively indicating that 90th percentile is two times of mean value.

Table 4.1 : Descriptive statistics of the variables in investor attention models.

Variable	Mean	Std. Dev.	Min	Max	Median	10^{th} Percentile	90^{th} Percentile
Ret	0.002	0.066	-1.083	0.882	0.000	-0.057	0.064
AbRet	-0.001	0.064	-1.015	0.900	-0.001	-0.052	0.051
SVI	26.056	19.042	1	100	21	7	53
ASVI	0.164	0.717	-2.811	5.298	0.134	-0.714	1.076
ASVI_{it-1}	0.132	0.711	-2.811	5.298	0.105	-0.734	1.033
Size	5.650	1.985	0.842	10.745	5.464	3.273	8.379
BM	1.006	1.091	-6.410	34.483	0.781	0.218	2.064
Lturnover	0.003	0.071	-0.826	4.796	-0.003	-0.060	0.071
Volatility	0.022	0.019	0.000	0.219	0.017	0.007	0.044
News	2.367	0.825	0.000	7.043	1.946	1.792	3.584
Abnturnover	0.063	0.897	-8.428	6.749	-0.024	-0.908	1.169

Table 4.2 presents correlations among the variables. The correlation between *Ret* and *ASVI* is 0.104 indicating that ASVI is positively related to the returns. The positive correlation level between *Ret, AbRet* and *ASVI* and $ASVI_{it-1}$ shows that both contemporenaous and lagged value of abnormal search volume index is positively related.

Table 4.2 : Correlation matrix of the variables in investor attention models.

	Ret	AbRet	SVI	ASVI	ASVI_{it-1}	Size	BM	Lturn over	Volatility	News	Abn turnover
Ret	1										
AbRet	0.687	1									
SVI	0.077	0.042	1								
ASVI	0.104	0.074	0.525	1							
ASVI_{it-1}	0.099	0.063	0.103	0.347	1						
Size	-0.015	-0.061	0.109	-0.028	-0.038	1					
BM	0.025	0.06	-0.02	0.004	-0.007	-0.207	1				
Lturn over	-0.042	-0.014	0.039	0.082	0.107	-0.021	-0.01	1			
Volatility	-0.036	-0.037	0.053	0.049	0.138	-0.167	0.023	0.262	1		
News	-0.005	-0.025	0.078	-0.035	-0.034	0.586	-0.14	0.028	-0.034	1	
Abn turnover	0.281	0.206	0.16	0.263	0.392	-0.028	-0.002	0.213	0.196	-0.028	1

4.1.1 SVI and indirect measures of investor attention

We investigate the relation between direct investor attention measure, Google SVI and indirect investor attention measures (abnormal turnover, absolute abnormal return and the number of news) as in Da et al. (2011) to reveal the likely effects of direct ASVI.

Table 4.3 : Correlations between proxies of investor attention.

	Ln (SVI)	News	Absolute AbRet	Abnturnover	Ln (Name_SVI)
Ln (SVI)	1				
News	0.092	1			
Absolute AbRet	0.094	-0.016	1		
Abnturnover	0.134	0.055	0.240	1	
Ln (Name_SVI)	0.136	0.376	-0.020	-0.003	1

Table 4.3 depicts the correlations between proxies of investor attention. The correlations between search volume and other proxies are low. The correlation between *Ln (SVI)* and *Ln (SVI_Name)* is 13.6%. This shows that searches on firm name may be done by individuals for many other reasons unrelated to financial information about the stock. Search queries on firm name is based on how the individual have searched the firm name. Thus, search volume on firm name is affected by subjectivity. Abnormal returns and turnover are other measures of investor attention. There is positive but weak correlation between SVI and other proxies. *Absolute AbRet* and *Ln (SVI)* correlation is 9.4%, and *Abnturnover* and *Ln (SVI)* correlation is 13.4%. This low correlation shows that the relation may be affected by many economic factors other than investor attention. Number of news is another commonly used proxy of investor attention. Table 4.3 shows that there is positive correlation, 9.2%, between *Ln (SVI)* and News. This result supports that newspaper articles or news do not necessarily increase investor interest unless investors read it (Da et al., 2011).

Table 4.4 shows regression results where the dependent variable in each model is ASVI. In Column 1, *Absolute AbRet* is a variable with positive and significant effects on *ASVI*. The results in Column 2 and 4 show that *Absolute AbRet* and *Abnturnover* have significant effects on *ASVI*. These results show that abnormal turnover and abnormal return are the variables that create attention. Column 4 reports that News does not have significant effect on ASVI. R^2 values in all regressions is around only 15.6% or below, showing that alternative measures explain a small portion of the change in abnormal search volume, ASVI.

In our models, we use time lags to examine the leading effect of measures of attention. Table 4.5 reports the relation between ASVI and indirect measures of investor attention with time lags. The dependent variables are *ASVI, Abnturnover, Absolute AbRet* and *News*. The coefficient of first lag *ASVI* is statistically significant on all other indirect investor attention variables.

Table 4.4 : ASVI and indirect measures of investor attention.

	(1)	(2)	(3)	(4)
Absolute AbRet	1.522***		0.789***	0.788***
	(7.74)		(6.03)	(6.02)
Abnturnover		0.184***	0.173***	0.173***
		(20.89)	(20.07)	(19.96)
News				0.00200
				(0.25)
Intercept	0.109***	0.153***	0.125***	0.120***
	(14.99)	(279.12)	(26.10)	(6.52)
R^2	0.1183	0.1539	0.1567	0.1567
N	42,897	42,897	42,897	42,897

The t-statistics are calculated using two-way clustered standard errors and reported in parentheses. $*$, $**$, and $***$ are significance levels at the 10%, 5%, and 1% level, respectively.

According to the results in Table 4.5, ASVI has positive coefficients on other attention proxies and explain other investor attention proxies better since R square values are 25.26%, 11.46% and 61.02% in the models where $ASVI_{t-1}$ is independent variable and these R square values are greater than 9.69% where other lagged value of investor attention proxies are explanatory variables of ASVI. Thus, ASVI captures investor attention better than abnormal turnover, abnormal returns and news in accordance with the results of Da et al. (2011).

Table 4.5 : ASVI and indirect measures of investor attention with time lags.

	ASVI	Abnturnover	Absolute AbRet	News
	(1)	(2)	(3)	(4)
Abnturnover_{t-1}	0.107***		0.003***	0.006
	(14.11)		(9.25)	(1.56)
$\text{Absolute Abnret}_{t-1}$	0.436***	2.043***		0.194**
	(4.33)	(7.45)		(2.90)
News_{t-1}	-0.009	-0.0005	0.0008*	
	(-1.18)	(-0.06)	(1.84)	
ASVI_{t-1}		0.450***	0.009***	0.036***
		(28.68)	(13.94)	(6.14)
Intercept	0.088***	-0.046**	0.032***	2.361***
	(4.58)	(-1.98)	(28.97)	(942.58)
R^2	0.0969	0.2526	0.1146	0.6102
N	35,347	35,347	42,897	35,347

The t-statistics are calculated using two-way clustered standard errors and reported in parentheses. $*$, $**$, and $***$ are significance levels at the 10%, 5%, and 1% level, respectively.

Table 4.5 also shows that the first lag of Absolute AbRet has a significant impact on ASVI. This relation supports that high abnormal return creates attention in the following week. The results in Column 1 to 3 in Table 4.5 supports that investors initially pay attention to stocks then decide to buy related stocks, and search attention could increase before important firm specific announcements. This decision process leads to the fact that ASVI has major impact on abnormal turnover, abnormal returns and news.

4.1.2 SVI and stock returns

We examine whether investor attention has a significant impact on stock returns at individual stock level to hypothesize that high investor attention by SVI explains increasing stock returns. In our study, we will use the Fama and MacBeth (1973) cross sectional regression model to perform our tests. The Fama and MacBeth cross sectional regression allows us to empirically examine the relation between abnormal SVI and stock returns, while at the same time controlling time-related effects. The two-step Fama and MacBeth regression first applies cross section regressions, then calculates the average of the coefficients obtained from the first step regressions. Using the Fama and MacBeth regression method, all regression variables were cross sectionalized, and all independent variables were standardized. The regression coefficients show the effect of a standard deviation variation on independent variables on the dependent variable. Skoulakis (2008) shows that the Fama and MacBeth cross sectional regression is more effective than the ordinary least squares regression for panel data with both a larger cross section and a longer data time period. The standard errors are calculated with the Newey and West (1987) formula with four lags to deal with autocorrelation and heteroskedasticity in the error terms. Fama and MacBeth cross sectional regression results are reported in Tables 4.6, 4.7 and 4.8.

In Table 4.6, contemporaneous regression results are reported to show the association between ASVI and returns. The positive coefficient of ASVI indicates that ASVI is positively related to returns in week *t*. The main finding is that ASVI is associated with higher contemporaneous raw and abnormal returns since all coefficients are highly significant.

In Table 4.7, the raw weekly stock return is the dependent variable, Ret is weekly raw return for the following weeks and cumulative stock return between weeks 5-52. In Table 4.8, *AbRet* is weekly DGTW adjusted abnormal return for the following weeks and cumulative abnormal return between weeks 5-52.

Table 4.6 : ASVI and stock returns (contemporaneous regressions).

	Ret	Ret	AbRet	AbRet
	(1)	(2)	(3)	(4)
ASVI	0.007***	0.009***	0.005***	0.006***
	(10.49)	(11.66)	(9.33)	(9.95)
Ret_{t-1}		0.031**		
		(2.34)		
Size		-0.0007***		-0.002***
		(-2.78)		(-8.46)
dLnturnover		-0.014*		-0.031***
		(-1.70)		(-3.40)
BM		0.001***		0.003***
		(2.99)		(7.61)
Volatility		-0.218***		-0.196***
		(-4.64)		(-5.00)
News		0.001**		0.002***
		(2.20)		(4.73)
$AbRet_{t-1}$				0.106***
				(4.75)
Intercept	0.0001	0.005**	-0.002***	0.005***
	(0.07)	(2.01)	(-5.41)	(3.86)
R^2	0.0205	0.1253	0.0139	0.1306
N	42,897	35,325	42,897	35,325
Time periods	232	229	232	229

The t statistics are given in parentheses. *, **, and *** are significance levels at the 10%, 5%, and 1% level, respectively.

In Table 4.7 and 4.8, the main and controlling firm characteristic variables for all regressions are as follows. *Ret* is weekly stock returns, *AbRet* is weekly is DGTW adjusted abnormal return, *SVI* is search frequency from Google Trends based on stock ticker, *ASVI* is the natural log of SVI during the week minus the natural log of median SVI for the previous 8 weeks, *Size* is the log of ratio of stock's market capitalization in week t-1, *BM* is the ratio of book value to market capitalization or value of the stock in week t-1, *Abnturnover* is the log of turnover relative to median of last 52 weeks turnover, *dLturnover* is the change in log of turnover in week t-1, *Volatility* is the standard deviation of the daily stock returns for the week t-1, *News* is the log of 1 plus the number of stories published on the most recent week. All independent variables are controlled by variance inflation factors to account for multicollinearity problem. Maximum level of variance inflation factor is typically 10 in the literature (Hair et al., 1995). All independent variables' variance inflation factors are at acceptable levels.

We find a significant positive impact of an increase in investor attention measured by ASVI on returns as Barber and Odean (2008) stated that individual investors are net buyers of attention-grabbing stocks. Table 4.7 and 4.8 indicate that significant predictability of returns endures for three weeks. The effect is completely reversed within a year in the Turkish stock market, different from Da et al. (2011)'s study which shows predictability continue two weeks and reversals occur after three weeks for US stock market. This difference shows that the significant effect of predictability is longer in emerging markets and the time for buying pressure from uninformed investors in emerging markets is greater than the time for buying pressure in developed markets. The effect between weeks 5 to week 52 is significant and negative, demonstrating that with the return reversals, the positive ASVI impact on stock returns in the first week is balanced in a year. The first price increases due to the temporary price pressure and the reversal effect in the long-term support the price pressure hypothesis.

Table 4.7 : ASVI and stock returns (raw returns).

	Week 1	Week 1	Week 2	Week 3	Week 4	Week 5-52
	(1)	(2)	(3)	(4)	(5)	(6)
ASVI	0.018***	0.007***	0.006***	0.006***	-0.0003	0.002
	(8.57)	(9.93)	(6.82)	(5.62)	(-0.78)	(0.55)
Ret	0.023**	0.025**	0.041***	0.057***	0.006	0.058
	(2.14)	(2.32)	(2.96)	(3.44)	(0.82)	(0.97)
Size	-0.0002	-0.0005**	-0.0006	-0.0008	-0.00009	-0.0007
	(-1.09)	(-2.21)	(-1.40)	(-1.33)	(-0.39)	(-0.26)
dLturnover	-0.011*	-0.010	-0.001	-0.004	0.009	0.164***
	(-1.73)	(-1.55)	(-0.19)	(-0.39)	(1.57)	(3.31)
BM	0.001***	0.001***	0.002***	0.003***	0.0009**	0.015***
	(3.48)	(3.53)	(3.65)	(3.22)	(2.20)	(3.55)
Volatility	-0.253***	-0.249***	-0.408***	-0.479***	-0.063**	-2.103***
	(-6.18)	(-6.16)	(-7.33)	(-7.03)	(-2.57)	(-6.60)
News	0.0006	0.0009*	0.001*	0.002**	0.0005	0.013***
	(1.27)	(1.89)	(1.96)	(2.10)	(1.22)	(4.17)
Size*ASVI	-0.002***					
	(-6.12)					
Intercept	0.003	0.004*	0.005	0.006	-0.0001	0.045**
	(1.42)	(1.79)	(1.41)	(1.17)	(-0.07)	(2.06)
R^2	0.1201	0.1088	0.1009	0.0954	0.0659	0.0705
N	42,873	42,873	42,864	42,852	42,835	30,580
Time periods	232	232	232	232	232	181

The t statistics are given in parentheses. *, **, and *** are significance levels at the 10%, 5%, and 1% level, respectively.

.

In Column 2 of Table 4.7 and 4.8, the positive association between ASVI and following week returns supports the price pressure hypothesis. We use an alternative measure of investor attention as control variables in regressions.

Table 4.8 reports the regression results where the dependent variable is the following week's abnormal returns computed with risk characteristic benchmark returns. We use the method of Daniel et al. (1997) where adjusted abnormal return is the dependent variable to test the robustness of these regression results. The abnormal return is calculated using size, book-to-market ratio and momentum factors.

Table 4.8 : ASVI and stock returns (abnormal returns).

	Week 1	Week 1	Week 2	Week 3	Week 4	Week 5-52
	(1)	(2)	(3)	(4)	(5)	(6)
ASVI	0.012***	0.004***	0.003***	0.003**	-0.0005	-0.012
	(6.66)	(6.64)	(3.83)	(2.34)	(-1.15)	(-1.53)
AbRet	0.096***	0.098***	0.192***	0.275***	0.070***	1.841***
	(4.52)	(4.58)	(5.34)	(5.46)	(4.52)	(4.00)
Size	-0.001***	-0.002***	-0.003***	-0.005***	-0.001***	-0.098***
	(-7.88)	(-8.54)	(-8.69)	(-8.25)	(-5.58)	(-8.15)
dLturnover	-0.031***	-0.030***	-0.033***	-0.056***	-0.008	-0.244**
	(-3.89)	(-3.70)	(-2.87)	(-3.59)	(-1.13)	(-2.29)
BM	0.003***	0.003***	0.006***	0.009***	0.002***	0.122***
	(8.37)	(8.37)	(8.83)	(9.32)	(7.30)	(10.85)
Volatility	-0.224***	-0.223***	-0.367***	-0.437***	-0.0637***	-1.722***
	(-6.23)	(-6.29)	(-7.36)	(-6.78)	(-2.71)	(-4.43)
News	0.001***	0.002***	0.003***	0.004***	0.001***	0.086***
	(3.89)	(4.32)	(4.87)	(5.06)	(2.72)	(6.61)
Size*ASVI	-0.001***					
	(-4.61)					
Intercept	0.005***	0.006***	0.009***	0.012***	0.002*	0.156***
	(3.88)	(4.26)	(3.91)	(3.68)	(1.94)	(5.34)
R^2	0.1273	0.1178	0.1210	0.1221	0.0749	0.1485
N	42,873	42,873	42,659	42,415	41,928	30,580
Time periods	232	232	231	230	228	181

The t statistics are given in parentheses. *, **, and *** are significance levels at the 10%, 5%, and 1% level, respectively.

The results in Table 4.8 show and support the conclusion that there is a positive and significant impact of ASVI on the following week's stock returns and this predictability endures for three weeks. We find strong evidence of positive return changes for abnormal returns with an increase in investor attention. Column 1 in Table 4.8 shows that one increase in ASVI standard deviation results in a significant positive return change of 0.012 for abnormal returns among stocks. Table 4.7 and 4.8 show that the price pressure effect of SVI is stronger among small stocks since price pressure is related to individual buying activity and small stocks are prone to a larger price effect. As individual buying pressure leads to price increase and the smaller stocks are prone to the larger price impact, the interaction term between size and *ASVI*, *Size*ASVI*, is statistically significant and negative.

In Table 4.8, Column 1 shows that number of news in the previous week, News, has positive and significant coefficients on abnormal return. These results support the idea of Barber and Odean (2008) which demonstrates that individual investors show attention leaded buying behavior and they are net buyers when stocks are in the news. However, if companies make important announcements, individual investors immediately begin to search for stocks and SVI increases. The increasing SVI immediately predicts prices or returns for the next period, while the news about the company is slowly incorporated into stock prices as also shown in Da et al. (2011). We find a significant effect of ASVI after controlling for company characteristics. Table 4.7 and 4.8 show the negative effect of size and positive effects of book to market ratio on following week returns as in Mondria and Wu (2011) and Ying et al. (2015). We find a negative effect of volatility and positive effect of news variable on following week returns in parallel with earlier studies of Da et al. (2011) and Ying et al. (2015).

4.1.3 Trading strategy on abnormal investor attention

The previous results show that positive impact of ASVI on abnormal returns endures for three weeks. To test whether ASVI contains valuable information, we create a trading strategy based on search volume to examine a portfolio that goes long in high attention stocks and short in low attention stocks has an economic sense. We construct three different portfolios for each week: (i) the high-attention portfolio including stocks with investor attention above the 90th percentile; (ii) the low-attention portfolio including stocks with investor attention below the 10th percentile; (iii) the long-short portfolio longs in high-attention stocks and shorts in low-attention stocks as zero-investment portfolio. We show the excess return for each portfolio and alphas factor models, Capital Asset Pricing Model (CAPM), Fama French Three Factor Model (Fama and French, 1993) and Carhart Four Factor Model (Carhart, 1997).

Table 4.9 : Portfolios sorted by abnormal attention.

	Low attention	High attention	High-Low
Equal weighted	-0.005***	0.011***	0.017***
	(-3.000)	-4.35	-8.55
Value weighted	-0.001	0.004	0.005**
	-0.655	-1.45	-2.1

Excess returns for portfolios sorted by abnormal attention. The t statistics for mean tests are given in parentheses. ∗, ∗∗, and ∗∗∗ are significance levels at the 10%, 5%, and 1% level, respectively.

Table 4.9 shows univariate analysis of the trading strategy based on portfolios sorted by abnormal investor attention. We split the sample into low, medium and high-attention parts for the weeks.

Table 4.9 depicts that the mean of equal weighted portfolio weekly returns with high attention is 0.011 and that of low attention is -0.005. The difference is 0.017 and 0.005 per week and the effect is significant between high and low abnormal investor attentions in equal and value weighted portfolios. This result shows that equal weighted portfolios sorted by abnormal attention generates significant return premium using high investor attention stocks.

Table 4.10 : Portfolios sorted by abnormal attention (attention-based trading strategy).

	Alpha	MKT	SMB	HML	UMD	R^2
Panel A. Equal Weighted Portfolio						
CAPM Model	0.0172***	-0.089				0.0086
	(7.57)	(-1.25)				
Fama French Three Factor Model	0.0166***	0.033	0.495***	-0.059		0.0946
	(7.35)	(0.53)	(4.73)	(-0.45)		
Carhart Four Factor Model	0.0168***	0.031	0.497***	-0.024	0.082	0.097
	(7.33)	(0.49)	(4.67)	(-0.18)	(0.79)	
Panel B. Value Weighted Portfolio						
	Alpha	MKT	SMB	HML	UMD	R^2
CAPM Model	0.0049**	-0.054				0.0025
	(2.02)	(-0.60)				
Fama French Three Factor Model	0.0049**	0.014	0.312**	0.050		0.0261
	(2.06)	(0.18)	(2.32)	(0.39)		
Carhart Four Factor Model	0.0058**	0.006	0.321**	0.180	0.309**	0.0512
	(2.57)	(0.08)	(2.38)	(1.21)	(2.15)	

The t statistics are given in parentheses. ∗, ∗∗, and ∗∗∗ are significance levels at the 10%, 5%, and 1% level, respectively.

Table 4.10 reports the multivariate analysis of the attention related investment strategy with Capital Asset Pricing Model (CAPM), Fama French Three Factor and Carhart Four Factor Models. The dependent variable is the returns of the long-short portfolio. The return of the market factor is represented as *MKT*; the return of size factor is represented as *SMB*, the return of book-to-market ratio is represented as *HML*; and momentum factor is represented as *UMD*. Table 4.10 indicates that the trading strategy with long in high attention stocks and short in low attention stocks has positive and significant alphas in all models for both equal and value weighted portfolios. The alphas of equal weighted portfolio and value weighted portfolio are 168 and 58.7 basis points per week, respectively, and these results indicate that smaller firm effect on predictability is higher. Univariate and multivariate analysis indicate that there is a significant return premium for high-investor attention after using the factors as control variables. This evidence supports that high and low attention-based stock portfolios have significant return differences that cannot be related to traditional CAPM, Fama French Three Factor and Carhart Four Factor Models. This result shows that trading strategy is profitable only in the short run since return reversals are observed after three weeks. However, if the strategy executes trades every week, the returns would be depending on round-trip and trading costs. This result may support that return premiums in short run exist in emerging markets where information efficiency is lower, and information is incorporated into asset prices in longer time. The positive coefficients on the size factor (*SMB*) show that the zero-trading strategy with long in high-attention stocks and short in low-attention stocks generates positive pressure for small stocks that shows individual investor attention may have a greater role in asset pricing.

4.2 Effects of Social Media Sentiment on International Stock Returns and Trading Activity

This subsection firstly gives results on the effects of the number of tweets and Twitter sentiment on individual stock returns and trading activity indicator which is abnormal turnover as a measure of trading volume. Then, the controlling effect of traditional media measurement, news sentiment is investigated. Finally, the effectiveness of trading strategy on Twitter sentiment is discussed.

Table 4.11 depicts the descriptive statistics of the returns for Twitter activity, sentiment, main dependent and control variables. The table also shows that the 10th, 50th (median), and 90th percentiles of Twitter sentiment are -0.1068, 0 and 0.2642, respectively, and the number of tweet counts are 1, 11 and 91, respectively. The mean number of tweets per day is 53.43, indicating that the firm is tweeted 53.43 times per day on average on Twitter and Stocktwits. Twitter sentiment (*Twitter*) represents the average value of twitter sentiment for the parent company over a 24-hour period. Firm specific sentiment based on Twitter varies from -1 to 1, with -1 representing the most negative sentiment and +1 representing the most positive sentiment over a 24-hour period. The average Twitter sentiment of firms is 0.0378 indicating that the effect of positive sentiment is higher on average. News sentiment (*News*) represents the average value of news sentiment for the parent company over a 24-hour period. Firm specific news sentiment is obtained from Bloomberg. Sentiment based on stories varies from -1 to 1, with -1 representing the most negative sentiment and +1 representing the most positive sentiment over a 24-hour period. The average news sentiment of firms is 0.1527 indicating that the effect of positive sentiment is higher on average.

Table 4.11 : Descriptive statistics of the variables in social media sentiment models.

	Mean	Std. Dev.	10th Percentile	Median	90th Percentile	N
Twitter	0.0378	0.2063	-0.1068	0.0000	0.2642	776,642
News	0.1527	0.2871	-0.1484	0.1172	0.5142	775,454
Ntweet	2.5888	1.4635	0.6931	2.4849	4.5218	713,001
Number of tweets	53.43	176.48	1.00	11.00	91.00	713,001
Ret(o-c)	0.0000	0.0143	-0.0162	0.0000	0.0159	776,642
Ret(o-o)	0.0002	0.0173	-0.0187	0.0000	0.0188	776,642
AbRet(o-c)	-0.0003	0.0144	-0.0161	-0.0003	0.0153	776,642
AbRet(o-o)	-0.0001	0.0173	-0.0186	0.0000	0.0185	776,642
AbRet[−5,−1](o-c)	-0.0013	0.0323	-0.0385	-0.0013	0.0356	772,784
AbRet[−5,−1](o-o)	-0.0004	0.0347	-0.0394	-0.0001	0.0386	772,784
Ret[−5,−1](o-c)	-0.0001	0.0335	-0.0391	0.0006	0.0379	776,461
Ret[−5,−1](o-o)	0.0009	0.0384	-0.0428	0.0018	0.0433	776,461
Size	6.8687	2.7973	3.9314	5.2354	10.5277	775,727
Vola[−5,−1]	0.0002	0.0003	0.0000	0.0001	0.0005	776,461
Illiq[−5,−1]	0.0006	0.0012	0.0001	0.0003	0.0011	776,418
AbTurn_{t-1}	-0.0009	0.4662	-0.4782	-0.0293	0.5274	775,934
AbTurn	-0.0009	0.4665	-0.4784	-0.0294	0.5273	776,455
Vola	0.0002	0.0004	0.0000	0.0001	0.0005	776,642

In Table 4.12, the descriptive statistics in different regions (index basis) are given. As observed from the number of observations, S&P 500 stocks constitute half of the sample. The highest average market capitalization level is observed in S&P 500 stocks whereas the lowest levels are observed in S&P 350 Europe stocks. The table depicts higher average number of tweets and Twitter sentiment for S&P 500 stocks, 93.24 and 0.0463, respectively. Emerging market stocks have relatively lower number of tweets and Twitter sentiment, 21.04 and 0.0227, respectively.

Table 4.12 : Descriptive statistics in different regions.

	N	Mean	Std. Dev.	10th Percentile	Median	90th Percentile
Panel A. S&P 500						
Twitter	404,753	0.0463	0.2108	-0.1068	0.0080	0.2921
News	404,248	0.1385	0.2825	-0.1333	0.0915	0.5072
Number of tweets	380,056	93.24	369.64	4.00	19.00	129.00
Market cap (US dollars)	404,643	37,502.52	57,544.56	6,142.88	17,474.09	88,465.76
Panel B. S&P 350 Europe						
	N	Mean	Std. Dev.	10th Percentile	Median	90th Percentile
Twitter	275,198	0.0305	0.2048	-0.1030	0.0000	0.2360
News	274,879	0.1848	0.2919	-0.1525	0.1907	0.5401
Number of tweets	236,889	28.97	88.35	1.00	5.00	57.00
Market cap (US dollars)	274,430	27,004.89	33,824.79	4,788.43	15,053.86	62,437.69
Panel C. S&P EM Core						
	N	Mean	Std. Dev.	10th Percentile	Median	90th Percentile
Twitter	96,691	0.0227	0.1892	-0.1068	0.0000	0.2039
News	96,327	0.1209	0.2833	-0.2039	0.0825	0.4865
Number of tweets	96,056	21.04	184.74	1.00	2.00	31.00
Market cap (US dollars)	96,680	31,979.46	50,638.75	4,013.53	15,689.89	69,371.51

Table 4.13 : Correlation matrix of variables in social media sentiment models.

	Twitter	News	Ntweet	Ret (o-c)	Ret (o-o)	AbRet (o-c)	AbRet (o-o)	AbRet [−5,−1] (o-c)	AbRet [−5,−1] (o-o)	Ret [−5,−1] (o-c)	Ret [−5,−1] (o-o)	Size	Vola [−5,−1]	Illiq [−5,−1]	$AbTurn_{t-1}$	AbTurn	Vola
Twitter	1																
News	0.144	1															
Ntweet	-0.0084	0.0293	1														
Ret(o-c)	0.0482	0.0418	0.0128	1													
Ret(o-o)	0.0743	0.0686	0.0105	0.8017	1												
AbRet(o-c)	0.0442	0.033	0.007	0.8058	0.6387	1											
AbRet(o-o)	0.0713	0.0617	0.0057	0.6397	0.8669	0.7975	1										
AbRet[−5,−1] (o-c)	0.0311	0.0281	0.011	0.0092	0.0113	-0.0088	-0.0034	1									
AbRet[−5,−1] (o-o)	0.0738	0.0754	0.0007	-0.0178	-0.0262	-0.0823	-0.0794	0.7461	1								
Ret[−5,−1] (o-c)	0.0298	0.0333	0.0126	0.0143	-0.0102	0.024	-0.0022	0.8204	0.6693	1							
Ret[−5,−1] (o-o)	0.0667	0.0736	0.0032	-0.0116	-0.0422	-0.0465	-0.0711	0.5508	0.8599	0.7867	1						
Size	-0.0453	0.0604	-0.2699	-0.0212	-0.0019	-0.0151	0.0033	-0.0252	0.0206	-0.0361	0.0067	1					
Vola[−5,−1]	-0.0391	-0.0549	0.0211	-0.0073	0.0083	-0.0297	-0.0101	0.0027	-0.0359	-0.0424	-0.0699	0.01	1				
Illiq[−5,−1]	-0.0222	-0.0509	-0.1764	-0.0034	0.0004	-0.0066	-0.0021	-0.0167	-0.0202	-0.0269	-0.0277	0.0339	0.2611	1			
$AbTurn_{t-1}$	0.0118	0.0278	0.0396	0.0062	0.0081	-0.009	-0.0043	0.0466	0.0304	0.0072	-0.0053	-0.0006	0.0255	0.0107	1		
AbTurn	0.0106	0.0367	0.0578	0.0223	0.0253	0.055	0.0526	0.0144	-0.0015	-0.0043	-0.0173	-0.0006	-0.0576	0.0211	0.2247	1	
Vola	-0.0319	-0.0249	0.0571	0.0068	0.0198	0.069	0.0718	-0.0196	-0.0942	-0.0599	-0.1201	0.0098	0.5009	0.1792	0.0798	0.2051	1

Table 4.13 presents correlations among the dependent and control variables. Twitter sentiment is positively related to returns and abnormal returns (open-to-open return) and the correlations are 0.0743 and 0.0713, respectively. The correlation between Number of tweets (*Ntweet*) and returns, and the correlation between Number of tweets and abnormal returns (open-to-open return) are 0.0105 and 0.005, respectively. The correlation between abnormal turnover (*AbTurn*) and number of tweets (*Ntweet*), and the correlation between abnormal turnover and Twitter sentiment (*Twitter*) are positive, 0.0578 and 0.0106, respectively. The correlations table shows that control variables do not have high correlations with any of the control variables that may cause multicollinearity problems.

4.2.1 Twitter activity, sentiment and trading activity

This subsection gives estimation results of the contemporaneous and predictability regressions on the Twitter activity and sentiment and stock market trading activity. The trading activity or trading volume is abnormal turnover (*AbTurn*), which is measured by firm *i*'s log turnover on day *t* minus its average log turnover on days *t*-5 to *t*-1 (Tetlock, 2011). Table 4.14 and 4.15, Panel A parts show results from daily Fama-MacBeth (1973) regressions on day *t*. Dependent variable is trading volume measure, $Abturn_{i,t}$ is firm i 's log of turnover on day *t* minus its average log of turnover on days *t*-1 to *t*-5. Columns 2, 3 and 4 are the subsamples for S&P 500, S&P 350 Europe and S&P EM Core companies. Newey-West (1987) standard errors robust to heteroskedasticity and six days of autocorrelation. Panel A displays the results from the contemporaneous cross-sectional regressions on the effect of the number of tweets (*Ntweet*). The coefficients in Table 4.14 Panel A shows that the number of tweets has positive and significant effect on abnormal turnover at day *t* for all region stocks. The regression coefficient on *Ntweet* shows that one-standard-deviation increase in the number of tweets, *Ntweet* is associated with a significant positive abnormal turnover change of 0.0319 at day t.

Contemporaneous regression results in Table 4.14 support the argument that people tend to post tweets about their trades with the motivation of informing their friends and environment to make them follow their actions and to gain reputation (Van Bommel, 2003).

Table 4.14 : Twitter activity and trading volume.

	All	S&P 500	S&P 350E	S&P EM
Panel A. Contemporaneous regressions				
	AbTurn			
	(1)	(2)	(3)	(4)
Ntweet	0.0319***	0.0524***	0.0232***	0.0167***
	(28.34)	(31.66)	(11.95)	(8.33)
Ret[−5,−1]	0.100***	0.0789*	0.0314	0.171***
	(2.81)	(1.95)	(0.38)	(2.65)
Size	-0.0277***	-0.0845***	-0.0184***	-0.0122***
	(-14.54)	(-16.43)	(-7.75)	(-5.36)
Vola[−5,−1]	-193.9***	-288.2***	-229.6***	-174.3***
	(-22.34)	(-23.62)	(-16.72)	(-14.37)
Illiq[−5,−1]	12.63***	66.04***	19.78***	3.234***
	(12.99)	(18.26)	(7.59)	(4.45)
Intercept	0.126***	0.215***	0.157***	0.136***
	(9.65)	(10.86)	(5.92)	(5.86)
Region dummy	Yes			
R^2	0.0992	0.0770	0.0601	0.0850
N	723,263	381,172	237,693	104,398
Time periods	781	781	781	781
Panel B. Predicting trading volume				
	AbTurn			
	All	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)
Ntweet_{t-1}	0.0252***	0.0395***	0.0203***	0.0188***
	(20.95)	(24.03)	(9.36)	(8.97)
Ret[−5,−1]	0.0909**	0.0503	0.0351	0.181***
	(2.53)	(1.24)	(0.43)	(2.77)
Size	-0.0224***	-0.0610***	-0.0174***	-0.0131***
	(-11.57)	(-11.55)	(-7.25)	(-5.78)
Vola[−5,−1]	-191.4***	-271.0***	-229.9***	-177.6***
	(-21.63)	(-22.30)	(-16.58)	(-14.62)
Illiq[−5,−1]	13.18***	66.41***	19.62***	3.182***
	(13.42)	(18.38)	(7.65)	(4.38)
Intercept	0.104***	0.155***	0.153***	0.143***
	(7.65)	(7.67)	(5.76)	(6.23)
Region dummy	Yes			
R^2	0.0947	0.0625	0.0590	0.0860
N	723,890	381,634	237,812	104,444
Time periods	782	782	782	782

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 4.15 : Twitter sentiment and trading volume.

Panel A. Contemporaneous regressions				
	AbTurn			
	All	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)
Twitter	0.0159***	0.0262***	0.00425	0.0118
	(5.40)	(6.87)	(0.74)	(1.14)
Ret[−5,−1]	0.0800**	0.0517	0.0258	0.149**
	(2.19)	(1.27)	(0.32)	(2.18)
Size	-0.0007	0.0150***	-0.0016	-0.0036
	(-0.49)	(4.26)	(-1.01)	(-1.63)
Vola[−5,−1]	-155.8***	-181.7***	-191.2***	-162.8***
	(-19.68)	(-17.58)	(-15.59)	(-13.08)
Illiq[−5,−1]	16.17***	66.41***	22.69***	3.784***
	(16.02)	(19.22)	(8.50)	(5.12)
Intercept	0.0182	-0.0622***	0.0369*	0.0749***
	(1.52)	(-3.62)	(1.91)	(3.04)
Region dummy	Yes			
R^2	0.0850	0.0433	0.0469	0.0841
N	775,584	404,536	274,382	96,666
Time periods	781	781	781	781

Panel B. Predicting trading volume				
	AbTurn			
	All	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)
Twitter_{t-1}	0.0126***	0.0144***	0.00830	0.0294***
	(4.28)	(3.94)	(1.25)	(2.61)
Ret[−5,−1]	0.0804**	0.0514	0.0269	0.149**
	(2.19)	(1.26)	(0.34)	(2.15)
Size	-0.000694	0.0148***	-0.00146	-0.00359
	(-0.47)	(4.20)	(-0.89)	(-1.62)
Vola[−5,−1]	-156.8***	-183.0***	-190.5***	-164.2***
	(-19.67)	(-17.61)	(-15.48)	(-13.22)
Illiq[−5,−1]	16.20***	66.30***	22.56***	4.121***
	(16.02)	(19.12)	(8.49)	(5.45)
Intercept	0.00937	-0.0600***	0.0353*	0.0744***
	(0.80)	(-3.49)	(1.82)	(3.02)
Region dummy	Yes			
R^2	0.0848	0.0429	0.0473	0.0856
N	776,200	405,009	274,490	96,701
Time periods	782	782	782	782

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Panel B in Table 4.14 reports the regression results for predictability of number of tweets when the dependent variable is the next day abnormal turnover. The coefficients of $Ntweet_{t-1}$ are positive and significant indicating that Twitter activity is positively related to trading volume on the subsequent day for all subsamples. The regression coefficient on $Ntweet_{t-1}$ shows that next-day abnormal turnover is 0.0252 higher after one-standard deviation increase in the number of tweets. These results suggest that Twitter activity predicts next day trading volume consistent with the informational role of social media (Wysocki,1998; Sprenger et al., 2014; Li et al., 2018).

Table 4.15 Panel A depicts that Twitter sentiment published on a given day is positively related to contemporaneous trading volume for all sample and S&P 500 stocks. The coefficients of *Twitter* are positive and significant indicating that Twitter sentiment is positively related to trading volume on the same day. The regression coefficient on *Twitter* shows that one-standard-deviation increase in the sentiment, *Twitter* is associated with a significant positive abnormal turnover change of 0.0159 at day *t*. Panel B in Table 4.15 displays the regression results for predictability of Twitter sentiment when the dependent variable is the next day abnormal turnover. These results suggest that Twitter activity predicts next day trading volume consistent with the informational role of social media (Sprenger et al., 2014). The coefficients of $Twitter_{t-1}$ are positive and significant except S&P 350 Europe stocks which have lowest average market capitalization levels in the sample. The regression coefficient on $Twitter_{t-1}$ shows that next-day abnormal turnover is 0.0126 higher after one-standard deviation increase in Twitter sentiment. This result indicates that when the language of the post is more positive or Twitter sentiment is high, trading volume on the next day will be higher. The main finding is that both the number of tweets and Twitter sentiment are associated with higher contemporaneous abnormal turnover and both are positively related to trading volume on the next day for the all firms in the sample. Coefficients of main variables, number of tweets and Twitter sentiment are statistically and economically significant.

4.2.2 Twitter activity, sentiment and stock returns

This section examines whether Twitter sentiment and tweets have informational role. We expect that positive sentiment and number of tweets to be related to future stock returns. To test the impact of Twitter activity and sentiment on stock returns, we follow the methodology of Tetlock (2011) and we use daily Fama and MacBeth (1973) cross-sectional regressions. In this part, regression test results on that the relation and predictability between number of tweets, Twitter sentiment and stock returns are given.

Tables 4.16 to 4.19 show results from daily Fama-MacBeth (1973) regressions. The standard errors in parentheses are robust to heteroskedasticity and autocorrelation for six days with the Newey and West (1987) method. For simplicity, firm's abnormal return (*AbRet*) is measured as its raw return minus the return on the weighted S&P 500, S&P 350 Europe and S&P EM Core indexes as in Tetlock (2011). The main variables are the number of tweets and lag values ($Ntweet_{i,t-1}$) and Twitter sentiment with lag values ($Twitter_{i,t-1}$). Following studies of Tetlock (2011), Chen et al. (2014) and Sprenger et al. (2014), we include additional firm characteristics that might affect stock returns as control variables to investigate whether the number and sentiment of tweets have incremental power to predict stock. We include five control variables. Firm's past stock performance as the cumulative raw (abnormal) returns on days t-5 to t-1 ($Ret_{it}[-5,-1]$ and $AbRet_{it}[-5,-1]$), Volatility, Park volatility (Parkinson, 1980) measure averaged over days t-5 to t-1 ($Vola_{it}[-5,-1]$), trading volume, abnormal turnover, firm's log turnover on day t minus its average log turnover on days t-5 to t-1 ($AbTurn_{i,t-1}$), firm i's log of market capitalization on day t-1 ($Size_{i,t-1}$), the Amihud's (2002) illiquidity measure averaged over days t-5 to t-1 and multiplied by 10^6 ($Illiq_{it}[-5,-1]$). Columns 2 and 3 in Tables 4.16 to 4.19 show subsamples of below-median and above-median firm size. Columns 4, 5 and 6 are the subsamples for S&P 500, S&P 350 Europe and S&P EM Core companies.

Table 4.16 reports the regression results of Twitter activity and raw stock returns where Table 4.17 reports the regression results of Twitter activity and abnormal stock returns which is calculated by raw return minus return of weighted index. For both tables, in Panel A sections, contemporaneous regression results for day *t* are reported. For all columns except S&P 350 Europe stocks, the coefficient on the number of tweets (*Ntweet*) is positively significant. The regression coefficient on *Ntweet* shows that one-standard-deviation increase in the number of tweets is associated with a significant positive return change of 1.6 basis points at day *t*.

Table 4.16 : Twitter activity and returns (raw open-to-open returns).

	All	Small	Big	S&P 500	S&P 350E	S&P EM
Panel A. Contemporaneous regressions						
Ret (open-to-open)						
	(1)	(2)	(3)	(4)	(5)	(6)
Ntweet	0.00016***	0.00021***	0.00016***	0.00029***	0.00004	0.00016*
	(5.01)	(4.16)	(4.76)	(5.65)	(1.01)	(1.78)
Ret[−5,−1]	-0.0122***	-0.0115***	-0.0131***	-0.0026	-0.0143***	-0.0242***
	(-4.96)	(-4.68)	(-4.74)	(-1.09)	(-4.18)	(-7.33)
Vola[−5,−1]	-1.061***	-1.031***	-0.911**	-1.672***	-1.170**	0.107
	(-2.85)	(-2.75)	(-2.17)	(-3.49)	(-2.28)	(0.24)
Illiq[−5,−1]	-0.0745**	-0.0731*	0.0681*	-0.203*	-0.00145	-0.0247
	(-2.50)	(-1.80)	(1.76)	(-1.72)	(-0.02)	(-1.21)
AbTurn_{t-1}	0.0002***	0.0001*	0.0003***	0.0002**	0.0002**	0.0002
	(3.16)	(1.74)	(3.38)	(2.20)	(2.02)	(1.57)
Size	-0.0001***	0.00004	-0.0003***	-0.0006***	-0.0001	-0.00007
	(-3.46)	(0.38)	(-4.31)	(-4.07)	(-1.46)	(-0.91)
Intercept	0.0015***	-0.0007	0.0009*	0.0024***	0.0013*	0.0005
	(3.71)	(-1.07)	(1.73)	(3.52)	(1.88)	(0.67)
Region dummy	Yes	Yes	Yes			
R^2	0.1327	0.1421	0.1556	0.1051	0.1002	0.1263
N	722,795	360,803	361,781	380,704	237,693	104,398
Time periods	781	781	781	780	781	781
Panel B. Predicting returns						
Ret (open-to-open)						
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Ntweet_{t-1}	-0.00002	-0.00004	0.000004	0.000005	-0.00003	-0.000003
	(-0.77)	(-1.22)	(0.15)	(0.13)	(-0.91)	(-0.04)
Ret[−5,−1]	-0.0121***	-0.0111***	-0.0129***	-0.0021	-0.0141***	-0.0241***
	(-4.90)	(-4.52)	(-4.64)	(-0.88)	(-4.14)	(-7.38)
Vola[−5,−1]	-0.864**	-0.815**	-0.634	-1.109**	-1.086**	0.172
	(-2.27)	(-2.11)	(-1.48)	(-2.20)	(-2.10)	(0.39)
Illiq[−5,−1]	-0.0615**	-0.0794*	0.0766**	-0.219*	0.0081	-0.0207
	(-2.07)	(-1.95)	(1.98)	(-1.85)	(0.12)	(-1.03)
AbTurn_{t-1}	0.0002***	0.0002**	0.0003***	0.0002***	0.0002**	0.0002
	(3.50)	(2.20)	(3.59)	(2.72)	(2.05)	(1.64)
Size	-0.00004	0.0001	-0.00008	-0.0001	-0.00005	0.000003
	(-0.79)	(1.04)	(-1.18)	(-0.99)	(-0.78)	(0.04)
Intercept	0.0008**	-0.0006	0.00008	0.0010	0.0010	-0.000006
	(2.31)	(-0.87)	(0.17)	(1.56)	(1.46)	(-0.01)
Region dummy	Yes	Yes	Yes			
R^2	0.1317	0.1403	0.1543	0.1015	0.0993	0.1250
N	723,419	361,189	362,230	381,163	237,812	104,444
Time periods	782	782	782	781	782	782

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 4.17 : Twitter activity and returns (abnormal open-to-open returns).

Panel A. Contemporaneous regressions						
	AbRet (open-to-open)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Ntweet	0.00016***	0.00020***	0.00016***	0.00029***	0.00004	0.00017*
	(4.98)	(4.09)	(4.83)	(5.60)	(0.90)	(1.93)
AbRet[−5,−1]	-0.0104***	-0.0094***	-0.0123***	-0.0035*	-0.0123***	-0.0222***
	(-4.96)	(-4.55)	(-4.81)	(-1.72)	(-4.11)	(-7.03)
Vola[−5,−1]	-1.027***	-0.982***	-0.902**	-1.644***	-1.141**	0.0669
	(-2.76)	(-2.65)	(-2.10)	(-3.42)	(-2.23)	(0.15)
Illiq[−5,−1]	-0.0790***	-0.0767*	0.0621	-0.218*	-0.0044	-0.0241
	(-2.64)	(-1.86)	(1.60)	(-1.83)	(-0.07)	(-1.17)
AbTurn_{t-1}	0.0002***	0.0001	0.0003***	0.0002**	0.0002*	0.0002*
	(3.07)	(1.62)	(3.44)	(2.19)	(1.78)	(1.65)
Size	-0.0001***	0.00002	-0.0003***	-0.0006***	-0.0001	-0.00007
	(-3.47)	(0.21)	(-4.32)	(-4.07)	(-1.50)	(-0.90)
Intercept	0.0011***	-0.0003	0.0006	0.0021***	0.0011*	0.0005
	(2.83)	(-0.39)	(1.27)	(3.46)	(1.67)	(0.58)
Region dummy	Yes	Yes	Yes			
R^2	0.1393	0.1499	0.1619	0.1035	0.0998	0.1259
N	719,656	359,385	360,056	378,768	236,783	104,105
Time periods	776	776	776	776	776	776
Panel B. Predicting returns						
	AbRet (open-to-open)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Ntweet_{t-1}	-0.00002	-0.00004	0.000006	0.000004	-0.00004	0.00001
	(-0.75)	(-1.25)	(0.20)	(0.11)	(-1.02)	(0.17)
AbRet[−5,−1]	-0.0121***	-0.0110***	-0.0129***	-0.0023	-0.0139***	-0.0241***
	(-4.85)	(-4.50)	(-4.59)	(-0.98)	(-4.06)	(-7.35)
Vola[−5,−1]	-0.842**	-0.790**	-0.645	-1.031**	-1.004**	0.180
	(-2.23)	(-2.07)	(-1.50)	(-2.06)	(-1.99)	(0.40)
Illiq[−5,−1]	-0.0615**	-0.0760*	0.0677*	-0.227*	0.0039	-0.0205
	(-2.05)	(-1.85)	(1.76)	(-1.91)	(0.06)	(-1.02)
AbTurn_{t-1}	0.0002***	0.0001**	0.0003***	0.0002***	0.0002*	0.0002
	(3.41)	(2.00)	(3.55)	(2.64)	(1.96)	(1.58)
Size	-0.00004	0.0001	-0.00008	-0.0001	-0.00005	-0.000003
	(-0.84)	(0.94)	(-1.25)	(-1.05)	(-0.75)	(-0.04)
Intercept	0.0003	-0.0001	0.0002	0.0007	0.0008	0.000008
	(0.96)	(-0.25)	(0.52)	(1.30)	(1.18)	(0.01)
Region dummy	Yes	Yes	Yes			
R^2	0.1405	0.1496	0.1625	0.1014	0.0990	0.1246
N	720,467	359,875	360,592	379,283	237,005	104,179
Time periods	777	777	777	777	777	777

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

The second and third columns in Table 4.16 and 4.17 report the regression results for "small" firms and "big" firms using the median of all sample firm size on day *t*. These columns depict that the association of the number of tweets with stock returns is more pronounced in small firms (0.00021) than in big firms (0.00016). A reason for this difference can be the fact that Twitter activity is more difficult to measure in big firms because their information environments are more complex (Tetlock, 2011). Tetlock (2011) states that measurement errors are larger for big firms. News and sources of information are abundant for big firms and even sophisticated text mining techniques may fail to catch details and similarities in a wide range of news stories. These results show that the increase in the number of tweets is an indication that new information has arrived on the market (Sprenger et al., 2014). These contemporaneous regression results in Tables 4.16 and 4.17 support that most of the tweets or messages denote buy signals and the increase in the number of tweets would be associated with higher stock returns (Bartov et al., 2017; Sprenger et al., 2014; Chen et al., 2014; Li et al., 2018). In Panel B in Tables 4.16 and 4.17, we report the regression results on the next-day raw returns and abnormal returns. We test whether the number of tweets (*Ntweet*), as the main variable, predicts future returns. The results show that the effect of the number of tweets of a firm reverses the next day and does not predict its next-day stock returns.

Table 4.18 reports the regression results of Twitter sentiment and raw stock returns where Table 4.19 displays the results of Twitter sentiment and abnormal stock returns which is calculated by raw return minus return of weighted index. For both tables, in Panel A sections, contemporaneous regression results for day *t* are reported. For all columns, sizes and regions, the coefficient on the Twitter sentiment (*Twitter*) is positively significant at the same and next day. The second and third columns in Tables 4.18 and 4.19 report the regression results for "small" firms and "big" firms using the median of all sample firm size on day *t*. These columns show that the Twitter sentiment is associated with an increase in stock returns is stronger in small firms (0.0065) than in big firms (0.0049). A reason for this difference can be the fact that Twitter sentiment is more difficult to measure in big firms because their information environments are more complex (Tetlock, 2011).

Table 4.18 : Twitter sentiment and returns (raw open-to-open returns).

Panel A. Contemporaneous regressions						
	Ret (open-to-open)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Twitter	0.00583***	0.00653***	0.00490***	0.00764***	0.00346***	0.00504***
	(28.39)	(24.42)	(25.85)	(28.18)	(17.20)	(8.62)
Ret[−5,−1]	-0.0152***	-0.0143***	-0.0161***	-0.00818***	-0.0166***	-0.0248***
	(-6.00)	(-5.61)	(-5.80)	(-3.36)	(-4.80)	(-7.23)
Vola[−5,−1]	-0.801**	-0.746*	-0.613	-1.038**	-0.962*	0.499
	(-2.03)	(-1.90)	(-1.35)	(-2.06)	(-1.79)	(1.06)
Illiq[−5,−1]	-0.0657**	-0.113***	0.0826**	-0.140	-0.0070	-0.0277
	(-2.21)	(-2.76)	(2.16)	(-1.21)	(-0.11)	(-0.96)
$AbTurn_{t-1}$	0.0002***	0.0001	0.0003***	0.0001*	0.0002**	0.0002
	(2.76)	(1.47)	(3.10)	(1.88)	(2.00)	(1.37)
Size	-0.00005	0.00001	-0.00001	-0.00001	-0.00007	-0.00001
	(-1.31)	(0.11)	(-0.23)	(-0.14)	(-1.39)	(-0.23)
Intercept	0.0002	0.000002	0.0006	0.00007	0.0010*	0.00007
	(0.58)	(0.00)	(1.57)	(0.14)	(1.78)	(0.09)
Region dummy	Yes	Yes	Yes			
R^2	0.1392	0.1500	0.1567	0.1164	0.0955	0.1234
N	775,091	376,798	398,067	404,043	274,382	96,666
Time periods	781	781	781	780	781	781
Panel B. Predicting returns						
	Ret (open-to-open)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
$Twitter_{t-1}$	0.00103***	0.00101***	0.000962***	0.000671***	0.00103***	0.00258***
	(9.48)	(6.78)	(7.29)	(4.80)	(6.73)	(5.60)
Ret[−5,−1]	-0.0127***	-0.0112***	-0.0140***	-0.0029	-0.0157***	-0.0241***
	(-4.97)	(-4.41)	(-4.99)	(-1.17)	(-4.54)	(-6.94)
Vola[−5,−1]	-0.947**	-0.944**	-0.731	-1.177**	-1.066**	0.474
	(-2.39)	(-2.33)	(-1.61)	(-2.31)	(-1.99)	(1.00)
Illiq[−5,−1]	-0.0687**	-0.0983**	0.0771**	-0.232**	0.0002	-0.0214
	(-2.25)	(-2.36)	(2.00)	(-1.99)	(0.00)	(-0.69)
$AbTurn_{t-1}$	0.0002***	0.0001**	0.0003***	0.0002***	0.0002**	0.0002
	(3.32)	(2.09)	(3.43)	(2.72)	(2.14)	(1.56)
Size	-0.00007*	0.00007	-0.00005	-0.0001	-0.00008	-0.00002
	(-1.70)	(0.66)	(-0.97)	(-1.25)	(-1.53)	(-0.31)
Intercept	0.0004	0.0001	0.0005	0.0009*	0.0012**	0.0001
	(1.23)	(0.16)	(1.30)	(1.80)	(2.05)	(0.21)
Region dummy	Yes	Yes	Yes			
R^2	0.1323	0.1423	0.1505	0.1016	0.0922	0.1222
N	775,712	377,179	398,533	404,521	274,490	96,701
Time periods	782	782	782	781	782	782

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 4.19 : Twitter sentiment and returns (abnormal open-to-open returns).

	All	Small	Big	S&P 500	S&P 350E	S&P EM
	Panel A. Contemporaneous regressions					
	AbRet (open-to-open)					
	(1)	(2)	(3)	(4)	(5)	(6)
Twitter	0.00577***	0.00649***	0.00487***	0.00761***	0.00342***	0.00463***
	(28.82)	(24.47)	(26.01)	(28.10)	(17.27)	(9.86)
AbRet[−5,−1]	-0.0151***	-0.0140***	-0.0160***	-0.0083***	-0.0164***	-0.0246***
	(-5.90)	(-5.56)	(-5.71)	(-3.41)	(-4.73)	(-7.14)
Vola[−5,−1]	-0.782**	-0.709*	-0.632	-0.986**	-0.911*	0.500
	(-1.99)	(-1.82)	(-1.38)	(-1.96)	(-1.71)	(1.05)
Illiq[−5,−1]	-0.0684**	-0.116***	0.0764**	-0.147	-0.0102	-0.0141
	(-2.28)	(-2.83)	(1.99)	(-1.26)	(-0.15)	(-0.63)
$AbTurn_{t-1}$	0.0002***	0.0001	0.0003***	0.0001*	0.0002*	0.0002
	(2.73)	(1.39)	(3.10)	(1.83)	(1.91)	(1.42)
Size	-0.00006	-0.000003	-0.00002	-0.00002	-0.00007	-0.000003
	(-1.43)	(-0.03)	(-0.34)	(-0.23)	(-1.42)	(-0.04)
Intercept	0.0002	-0.0001	0.0002	-0.0001	0.0008	-0.0001
	(0.78)	(-0.21)	(0.51)	(-0.37)	(1.37)	(-0.14)
Region dummy	Yes	Yes	Yes			
R^2	0.1470	0.1576	0.1644	0.1162	0.0952	0.1222
N	771,908	375,415	396,262	401,989	273,497	96,422
Time periods	776	776	776	776	776	776
	Panel B. Predicting returns					
	AbRet (open-to-open)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
$Twitter_{t-1}$	0.00101***	0.00101***	0.00093***	0.00066***	0.00103***	0.00233***
	(9.30)	(6.76)	(7.12)	(4.76)	(6.66)	(5.74)
AbRet[−5,−1]	-0.0127***	-0.0112***	-0.0141***	-0.0031	-0.0157***	-0.0243***
	(-4.96)	(-4.44)	(-4.96)	(-1.28)	(-4.53)	(-7.05)
Vola[−5,−1]	-0.911**	-0.842**	-0.739	-1.097**	-1.026*	0.477
	(-2.31)	(-2.15)	(-1.62)	(-2.18)	(-1.92)	(1.00)
Illiq[−5,−1]	-0.0723**	-0.105**	0.0694*	-0.239**	-0.0021	-0.0163
	(-2.37)	(-2.57)	(1.80)	(-2.04)	(-0.03)	(-0.74)
$AbTurn_{t-1}$	0.0002***	0.0001*	0.0003***	0.0002***	0.0002**	0.0003
	(3.27)	(1.95)	(3.42)	(2.66)	(2.06)	(1.61)
Size	-0.00008*	0.00005	-0.00006	-0.0001	-0.00008	-0.00001
	(-1.82)	(0.45)	(-1.08)	(-1.33)	(-1.56)	(-0.21)
Intercept	0.0003	-0.0003	0.0004	0.0007	0.0010	0.00003
	(1.09)	(-0.56)	(0.92)	(1.42)	(1.63)	(0.04)
Region dummy	Yes	Yes	Yes			
R^2	0.1414	0.1511	0.1594	0.1016	0.0920	0.1207
N	772,736	375,905	396,831	402,534	273,715	96,487
Time periods	777	777	777	777	777	777

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

In Tables 4.18 and 4.19 Panel B, we report the regression results on the next-day raw returns and abnormal returns. We test whether the Twitter sentiment (*Twitter*), as the main variable, predicts future returns. The results on the coefficient of $Tweet_{t-1}$ suggest that the effect of Twitter sentiment for all sizes and regions are positively significant and predict its next-day stock returns. In Table 4.18, the results depict that one-standard-deviation increase in Twitter sentiment leads to a significant positive return change of 10.3 basis points in the next day for raw returns. Column 2 and 3 show that the impact of Twitter sentiment in predicting returns is more pronounced in small stocks. Regression coefficient on $Tweet_{t-1}$ implies that next-day returns are 10.1 and 9.6 basis points higher after each one-standard deviation increase in Twitter sentiment in small and big stocks, respectively. Baker and Wurgler (2006) state that stocks that are difficult to arbitrage or to value are most affected by sentiment. One might expect the investor sentiment to have a stronger effect on stocks that are difficult to value such as small firms. The results for small and big stocks support the argument stating that predictability of investor sentiment is stronger on small firms.

The results in Table 4.18 Panel B indicate that the coefficient of S&P EM is greater than (0.00258) other region coefficients suggesting that Twitter sentiment tend to have greater impact on predictability of stock returns in emerging market stocks. This result is consistent with the result of Calomiris and Mamaysky (2019) where the authors show that news and content measures tend to have higher predictive power for returns and risks in emerging markets. US stocks are relatively larger and highly followed by analysts. Our results are in accordance with the results of Baik et al. (2016) which find the relation between the negative tone of local tweets and future returns in stocks with high information asymmetry such as less liquid firms, non-S&P 500 firms and firms with followed by lower analysts.

Overall, the results on Twitter activity and sentiment show that the coefficients on the number of tweets (*Ntweet*) is insignificant in predictability analysis while the coefficient on Twitter sentiment *(Twitter)* is significant. These results suggest that the stock returns increase when the tone or the language of the post is more positive in Twitter, but no predictability power of the number of Twitter activity exists whereas Twitter sentiment has incremental information able to predict future stock returns. These results are consistent with the arguments stating that Twitter sentiment would be associated with higher stock returns and high Twitter sentiment would predict stock returns in the next day (Sprenger et al. 2014; Sul et al., 2017; Li et al., 2018).

4.2.3 Twitter sentiment and stock returns with news sentiment

This section gives the results on whether Twitter sentiment has predictive value for stock returns incremental to that of the news sentiment. Tetlock et al. (2008) examine that firm-specific linguistic media content can be used to predict firms’ accounting earnings and stock returns and find that firms’ stock prices underreact to the information in negative words. To test the impact of Twitter sentiment on stock returns, we use daily Fama and MacBeth (1973) regressions using news sentiment as a control variable. The main research question is whether predictability of Twitter sentiment comes from the information diffusion from traditional media to social media. We expect our main variables on Twitter to be significant after controlling traditional media and news. Since we obtain sentiment data, number of news and tweets from Bloomberg, the procedure in all calculations of the variables are consistent.

Table 4.20 shows that firm-specific news sentiment (*News*) predicts future returns. However, the predictive power of Twitter sentiment is not counteracted by news sentiment. After controlling news sentiment, the coefficient of Twitter sentiment is still significant and positive. After controlling news sentiment effect, one-standard-deviation increase in Twitter sentiment leads to a significant positive return change of 8.7 basis points in the next day. With reference to low correlation between news sentiment and Twitter sentiment (0.14) and significant regression results, the results suggest that social media contains information that is incremental to that contained in traditional news media.

Overall, these results show that Twitter sentiment predicts stock returns over the next day without subsequent reversal. We further test whether Twitter sentiment contains fundamental information about stocks. To test the impact of Twitter sentiment on return reversals, we use daily Fama and MacBeth (1973) regressions for the following-day returns.

Table 4.20 : Twitter sentiment and stock returns with news sentiment.

	All	Small	Big	S&P 500	S&P 350E	S&P EM
Panel A. Raw return						
Ret (open-to-open)						
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathrm{Twitter}_{t-1}$	0.00087***	0.00085***	0.00080***	0.00060***	0.00086***	0.00221***
	(8.50)	(5.92)	(6.28)	(4.43)	(5.89)	(4.83)
News_{t-1}	0.00087***	0.00099***	0.00073***	0.00028***	0.00113***	0.00247***
	(9.56)	(8.40)	(7.03)	(3.16)	(7.35)	(9.29)
Ret[−5,−1]	-0.0134***	-0.0121***	-0.0147***	-0.0031	-0.0168***	-0.0257***
	(-5.24)	(-4.72)	(-5.21)	(-1.27)	(-4.78)	(-7.41)
Vola[−5,−1]	-0.932**	-0.952**	-0.703	-1.182**	-1.026*	0.465
	(-2.35)	(-2.34)	(-1.55)	(-2.32)	(-1.92)	(0.98)
Illiq[−5,−1]	-0.0653**	-0.0906**	0.0740*	-0.233**	0.0093	-0.0158
	(-2.13)	(-2.16)	(1.91)	(-1.99)	(0.14)	(-0.53)
AbTurn_{t-1}	0.0002***	0.0001*	0.0003***	0.0002***	0.0002**	0.0002
	(2.99)	(1.79)	(3.10)	(2.63)	(1.98)	(1.55)
Size	-0.00008*	0.00005	-0.00006	-0.0001	-0.00009*	-0.00005
	(-1.95)	(0.53)	(-1.15)	(-1.32)	(-1.71)	(-0.64)
Intercept	0.0001	0.00001	0.0003	0.0009*	0.0011*	0.00008
	(0.47)	(0.02)	(0.68)	(1.79)	(1.83)	(0.11)
Region dummy	Yes	Yes	Yes			
R^2	0.1341	0.1450	0.1536	0.1042	0.0971	0.1343
N	774,538	376,705	397,833	404,027	274,170	96,341
Time periods	782	782	782	781	782	782
Panel A. Abnormal return						
AbRet (open-to-open)						
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathrm{Twitter}_{t-1}$	0.00084***	0.00086***	0.00077***	0.00060***	0.00085***	0.00196***
	(8.30)	(6.01)	(6.09)	(4.40)	(5.79)	(4.88)
News_{t-1}	0.00088***	0.00099***	0.00074***	0.00028***	0.00114***	0.00247***
	(9.60)	(8.35)	(7.09)	(3.12)	(7.40)	(9.23)
AbRet[−5,−1]	-0.0135***	-0.0121***	-0.0147***	-0.0034	-0.0167***	-0.0260***
	(-5.23)	(-4.76)	(-5.18)	(-1.38)	(-4.78)	(-7.55)
Vola[−5,−1]	-0.897**	-0.845**	-0.710	-1.105**	-0.982*	0.474
	(-2.28)	(-2.16)	(-1.56)	(-2.19)	(-1.85)	(1.00)
Illiq[−5,−1]	-0.0689**	-0.0984**	0.0664*	-0.240**	0.00693	-0.0129
	(-2.26)	(-2.39)	(1.71)	(-2.05)	(0.10)	(-0.57)
AbTurn_{t-1}	0.0002***	0.0001*	0.0003***	0.0002**	0.0002*	0.0002
	(2.94)	(1.66)	(3.09)	(2.56)	(1.90)	(1.56)
Size	-0.00009**	0.00003	-0.00007	-0.0001	-0.00009*	-0.00004
	(-2.07)	(0.32)	(-1.27)	(-1.41)	(-1.74)	(-0.57)
Intercept	0.0004	-0.00007	0.0005	0.0007	0.0009	-0.000005
	(1.29)	(-0.10)	(0.99)	(1.42)	(1.42)	(-0.01)
Region dummy	Yes	Yes	Yes			
R^2	0.1431	0.1536	0.1624	0.1041	0.0969	0.1328
N	771,585	375,448	396,137	402,058	273,400	96,127
Time periods	777	777	777	777	777	777

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 4.21 : Return reversals at different time horizons (all sample).

	All						
	Ret (open-to-open)						
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 1-5	Day 6-10
Twitter_{t-1}	0.00103***	0.00005	0.00015	0.00014	0.0001	0.00149***	-0.00019
	(9.48)	(0.49)	(1.56)	(1.41)	(0.97)	(4.39)	(-0.56)
$\text{Ret}[-5,-1]$	-0.0127***	-0.0043**	-0.0015	0.0003	-0.0009	-0.0194**	0.00006
	(-4.97)	(-2.02)	(-0.73)	(0.16)	(-0.48)	(-2.56)	(0.01)
$\text{Vola}[-5,-1]$	-0.947**	-0.840**	-0.842**	-0.703*	-0.639	-3.972**	-4.082**
	(-2.39)	(-2.11)	(-2.05)	(-1.71)	(-1.55)	(-2.37)	(-2.34)
$\text{Illiq}[-5,-1]$	-0.0687**	-0.0617**	-0.0392	-0.0532*	-0.0425	-0.265**	-0.183
	(-2.25)	(-2.03)	(-1.28)	(-1.67)	(-1.38)	(-2.31)	(-1.56)
AbTurn_{t-1}	0.0002***	0.00009	-0.00002	-0.0001*	-0.0001**	0.00006	0.00003
	(3.32)	(1.35)	(-0.29)	(-1.68)	(-2.22)	(0.35)	(0.18)
Size	-0.00007*	-0.00006	-0.00005	-0.00004	-0.00004	-0.0002*	-0.0003*
	(-1.70)	(-1.57)	(-1.31)	(-1.11)	(-1.09)	(-1.67)	(-1.85)
Intercept	0.0004	0.0003	0.0003	0.0002	0.0006**	0.0015	0.0012
	(1.13)	(1.05)	(1.11)	(0.61)	(1.98)	(1.29)	(1.09)
Region dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.1323	0.1293	0.1281	0.1278	0.1265	0.1232	0.1210
N	775,712	774,650	773,588	772,526	771,465	771,465	766,160
Time periods	782	781	780	779	778	778	773

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 4.21 shows the duration of the return reversal for different time horizon of the dependent return variable for all firms and regions. Table 4.21 displays the results for one-day periods (through day 1 to 5), one-week period as cumulative return from day 1 to 5 and two-week period as cumulative return from day 6 to 10. Table 4.21 reports that Twitter sentiment has small-positive but insignificant effect on returns from day 2 to 5. The coefficient on Twitter sentiment for the cumulative return from day 1 to 5 is 0.0014 and positively significant depicting that Twitter sentiment has some fundamental information that has not been incorporated into stock prices. However, the negative coefficient on cumulative return from day 6 to 10 shows that this effect disappears thereafter. Higher Twitter sentiment is associated with higher next day return and lower subsequent returns after 5 trading days. The effect reverses after fundamentals are come out. This result suggests that higher Twitter sentiment is related to temporary overvaluation that reverses after one week. Regression results are robust to open to close returns as in Tables A.2 to A.7 in Appendix.

4.2.4 Trading strategy on Twitter sentiment

The regressions in previous parts indicate that Twitter sentiment is positively associated with subsequent returns after including several control variables. The cumulative positive return in the first week suggest that trading strategy on Twitter sentiment could earn significant premium. To build a trading strategy, we follow the portfolio methodology of Tetlock (2011). For each trading day, we determine the deciles for Twitter sentiment. We sort firms based on top and bottom 10th and 90th percentile Twitter sentiments. We form long-short portfolios that buy the stocks with Twitter sentiment in the top decile and sell the stocks with Twitter sentiment in the bottom decile. The portfolios are long on firms experiencing average positive sentiment with high Twitter sentiment, and short on firms experiencing average negative sentiment with low Twitter sentiment. We form the equal and value weighted Twitter sentiment portfolios when Twitter sentiment is updated 10 minutes before market opens. We hold long and short portfolios during 24 hours for the trading day until the market opens the next day for open-to-open returns and rebalance at the beginning of each trading day. Tables A.8 to A.10 in the Appendix section show long-short portfolio results in open-to-close returns for the trading day until the market closes the same day and it is rebalanced at the beginning of each trading day. We compute the risk-adjusted returns on four return factors: market (*MKT*), size (*SMB*), book-to-market (*HML*) factors proposed in Fama and French (1993), and the *UMD* factor based on the momentum. We show the excess return for each portfolio and alphas factor models, Capital Asset Pricing Model (CAPM), Fama French Three Factor Model (Fama and French, 1993) and Carhart Four Factor Model (Carhart, 1997). We obtain global, US and Europe risk factors daily from Ken French's website (Url-10). We compute daily Fama and French (1993) three factors and momentum for emerging markets using the same methodology with Ken French. The momentum factor is computed as a long-short portfolio sorting on prior returns of two to twelve months. Twitter sentiment-based portfolios' risk-adjusted returns are the alphas or intercepts in the regressions of the portfolio's raw return on the risk factors. We use adjusted Newey and West (1987) standard errors for heteroskedasticity and up to six days of serial correlation.

Table 4.22 : Trading strategy on Twitter sentiment (global portfolio).

	Equal-weighted			Value-weighted		
	(1)	(2)	(3)	(1)	(2)	(3)
MKT	-0.0006**	-0.0005**	-0.0003	-0.0007**	-0.0005*	-0.0004
	(-2.41)	(-2.04)	(-1.11)	(-2.51)	(-1.83)	(-1.18)
SMB		-0.0002	-0.0001		0.0002	0.0003
		(-0.48)	(-0.30)		(0.45)	(0.60)
HML		-0.0015***	-0.0004		-0.0015***	-0.0008*
		(-3.88)	(-1.17)		(-3.64)	(-1.66)
UMD			0.0013***			0.0009**
			(3.52)			(2.53)
Alpha	0.00079***	0.00079***	0.00077***	0.00077***	0.00076***	0.00074***
	(6.89)	(7.04)	(7.07)	(6.48)	(6.48)	(6.36)
R^2	0.0159	0.042	0.0728	0.019	0.0424	0.0565

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 4.22 reports the performance of the trading strategy on Twitter sentiment for all stocks in the sample using global risk factors. We ignore transaction costs in this analysis and these strategies are based on open-to-open returns (see Table A.8 in Appendix A for the results on open-to-close returns). Open-to-close rebalancing increases transaction costs that can counteract the profit. Transaction costs can be decreased by holding the portfolio using open-to-open prices (Url-9). Column 1 in Table 4.22 reports that the alpha of the sentiment-based investment strategy with Capital Asset Pricing Model (CAPM) is 7.9 (7.7) basis points per day for equal-weighted (value-weighted) portfolios. Column 2 and Column 3 report the results on Fama French Three Factor and Carhart Four Factor Models. These columns show that the daily average raw returns of these long-short strategies are 7.9 (7.6) and 7.7 (7.4) basis points, respectively.

Griffin (2002) supports that local factor models explain returns better and have lower pricing errors than the global factor model and Cakici et al. (2013) give an evidence that local factors perform better suggesting emerging market segmentation. As reference to these studies, we report the performance of models using local factors. Table 4.22 reports the performance of trading strategy on Twitter sentiment in region basis using US, Europe and emerging markets risk factors for S&P 500, S&P 350 Europe and S&P EM Core stocks respectively. We ignore transaction costs in these cases and these strategies are based on open-to-open returns.

Table 4.23 : Trading strategy on Twitter sentiment (region portfolios).

	Panel A. US portfolio					
	Equal-weighted			Value-weighted		
	(1)	(2)	(3)	(1)	(2)	(3)
MKT	-0.0005***	-0.0005***	-0.0003*	-0.0001	-0.0002	-0.00008
	(-3.10)	(-2.87)	(-1.78)	(-1.11)	(-1.29)	(-0.48)
SMB		0.00002	0.0004		0.0004	0.0006**
		(0.07)	(1.27)		(1.43)	(2.16)
HML		-0.0010***	-0.00006		-0.00124***	-0.0006**
		(-3.15)	(-0.18)		(-4.15)	(-2.19)
UMD			0.0016***			0.0009***
			(6.37)			(4.32)
Alpha	0.00064***	0.00064***	0.00061***	0.00052***	0.00052***	0.00050***
	(4.43)	(4.49)	(4.61)	(3.58)	(3.66)	(3.65)
R^2	0.0131	0.0317	0.1219	0.0015	0.0321	0.0618
	Panel B. Europe portfolio					
	Equal-weighted			Value-weighted		
	(1)	(2)	(3)	(1)	(2)	(3)
MKT	-0.0002	-0.000002	0.0002	-0.0004***	-0.0002	-0.0001
	(-1.31)	(-0.01)	(0.93)	(-2.60)	(-1.35)	(-0.49)
SMB		-0.0002	-0.0002		-0.0004	-0.0004
		(-0.61)	(-0.49)		(-0.89)	(-0.84)
HML		-0.0025***	-0.0015***		-0.0025***	-0.0017***
		(-6.24)	(-3.84)		(-5.04)	(-3.24)
UMD			0.0015***			0.0012***
			(4.45)			(3.47)
Alpha	0.00089***	0.00088***	0.00083***	0.00056***	0.00056***	0.00051***
	(6.25)	(6.21)	(5.91)	(3.56)	(3.52)	(3.31)
R^2	0.0032	0.0689	0.0999	0.0088	0.0595	0.0746
	Panel C. Emerging markets portfolio					
	Equal-weighted			Value-weighted		
	(1)	(2)	(3)	(1)	(2)	(3)
MKT	-0.0891**	-0.0829*	-0.0829*	-0.0338	-0.0218	-0.0214
	(-1.97)	(-1.83)	(-1.86)	(-0.64)	(-0.41)	(-0.40)
SMB		-0.0411	-0.0436		0.00615	0.00884
		(-1.00)	(-0.89)		(0.14)	(0.21)
HML		-0.0397	-0.0402		-0.0599*	-0.0596**
		(-1.57)	(-1.43)		(-1.94)	(-2.13)
UMD			-0.0190			0.0215
			(-0.32)			(0.39)
Alpha	0.00124***	0.00120***	0.00121***	0.00158***	0.00154***	0.00154***
	(3.90)	(3.75)	(3.74)	(4.05)	(3.90)	(3.92)
R^2	0.0078	0.01	0.0105	0.0007	0.0037	0.0041

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

In Table 4.23, Panel A, B and C shows US, Europe and emerging markets portfolio results, respectively. Column 1 reports the multivariate analysis of the sentiment-based investment strategy with CAPM and Column 2 and Column 3 report Fama French

Three Factor and Carhart Four Factor Models, respectively. Column 3 shows that the daily average raw returns of this long-short strategies in the US, Europe and emerging markets are 6.1 (5), 8.3 (5.1) and 12.1 (15.4) basis points per day for equal-weighted (value-weighted) portfolios, respectively. These results indicate that this trading strategy is more profitable in emerging markets stocks and less profitable in US stocks. These results are consistent with the literature stating that emerging markets have high information asymmetries and social media content measures tend to have higher predictive power for returns in emerging markets.

Table 4.24 : Trading strategy on Twitter sentiment with trading costs.

Panel A. 1 day holding								
	Equal-weighted				Value-weighted			
	Global	US	Europe	EM	Global	US	Europe	EM
Trading Costs (bps)	Abnormal Annualized Returns (%)				Abnormal Annualized Returns (%)			
0	22.00	17.25	23.90	36.61	21.06	14.00	14.29	48.74
10	-3.80	-8.55	-1.90	10.81	-4.74	-11.80	-11.51	22.94
20				-14.99				-2.86
Panel B. 5-days holding								
	Equal-weighted				Value-weighted			
	Global	US	Europe	EM	Global	US	Europe	EM
Trading Costs (bps)	Abnormal Annualized Returns (%)				Abnormal Annualized Returns (%)			
0	22.15	17.37	24.06	36.84	20.31	13.52	13.81	46.74
10	16.95	12.17	18.86	31.64	15.11	8.32	8.61	41.54
20				26.44				36.34

We perform additional analysis on the performance of trading strategy on Twitter sentiment with trading costs. We follow the methodology of Tetlock et al. (2008) and assume round-trip trading costs of 10 basis points for total buy and sell for US and Europe stocks. Round trip trading costs refer to the sum of all the costs incurred in single securities transaction such as commissions exchange fees, bid-ask spreads, market impact costs, and taxes. For emerging market stocks, Lesmond (2005) states that higher incremental political risk is related to 10 basis points increase in transaction costs. Thus, we add 20 basis points round trip costs into the analysis to examine the higher transaction effects in emerging market portfolio.

We use two different holding periods, 1 day and 5 days upon the return reversal results. Table 4.24 reports daily alpha and annualized risk adjusted returns of the trading strategy using Carhart Four Factor Models before and after trading costs of 1 day and 5 days holding periods. Without trading costs, both equal-weighted and value weighted portfolios generate positive returns. After including trading costs, the returns are negative since the strategy iterates every trading day. The trading strategies with 5-days holding period generate significant positive returns, both before and after trading costs for equal-weighted portfolios. These results show that portfolio returns are statistically and economically significant for 5-days. Return reversals between 5 and 10 days suggest that holding in longer time horizons becomes challenging. These findings suggest that the predictive power is short sighted, and strategies might be formed only for short term investments. Table 4.24 Panel B shows that without trading costs, for value weighted portfolios, annualized return is 20.31% for global, 13.52% for the US portfolio, 13.81% for the Europe and 46.74% for emerging markets portfolio. After including 10 basis points trading costs for global, US and Europe portfolio, annualized returns are 15.11%, 8.32% and 8.61%, respectively. For emerging market portfolio, annualized return is 36.34% even after 20 basis points transaction costs. The portfolio results show that from a practical perspective, investors could potentially use social media sentiment in their trading strategies in global or regional basis. However, the predictive power is short sighted, and strategies might be formed only for short term investments.

5. CONCLUSION

The conclusion section of the thesis is based on two main hypotheses in the study. First, the section discusses the results of the impacts of investor attention measured by abnormal Google Search Volume Index on stock returns in Turkey. The first hypothesis claims that an increase in the investor attention increases the individual stock returns so that individual investors are net buyers of attention-grabbing stocks with short term buying pressure. The analysis regarding the first hypothesis indicate that the Google SVI likely and directly grabs the attention of individual investors, and stock prices tend to be driven by the behavioral factors due to the investor attention in stocks listed in Turkey. The second part of the conclusion provides evidences for the second hypothesis where the effects of stock-specific social media sentiment, mainly Twitter sentiment, is investigated in the perspective of an international investor. The second hypothesis claims that Twitter activity measured by the number of tweets is associated with stock returns and trading volume, and positive tone of the social media sentiment is associated with and predicts stock returns and trading volume. Different from first part of the thesis, the second part presents the impacts of the information-interactive social media platforms where the content of the tweets is also important besides the number of direct attention measures. The findings in the second part of the thesis provides better insights to explore the complex behavior of the investors with the expansion of the social media platforms in recent years. The results in the second part suggest that social media activity and sentiment provide new information about firms and social media has different impacts than traditional news media on firms' information environments. The results also show that the role of social media is to diffuse sentiment to investors who unintentionally make prices less efficient in the short term.

Search engines are an easy and effective source of information used by investors. Recent growing literature point outs the effectiveness of Google SVI in various fields of research on topics related to finance and stock market. We use Google SVI as a proxy of investor attention because search volume is likely to capture attention for two main reasons. First, individuals generally use Internet search engine to gather information meaning that Google SVI might represent the household interest on a topic in general. Second, Google SVI data provides better indication of investors' behaviors or decisions than other investor attention proxies such as turnover, news and abnormal returns, because searching for information on the web is more likely to be related to an action or buying.

We analyze whether Google search queries influence stock returns in a sample of stocks of the Borsa Istanbul all shares index over the period April 2013 and September 2017 in Turkey. As Barber and Odean (2008) stated that individual investors are net buyers of attention-grabbing stocks, we find evidence that an increase in ASVI is related to higher future returns. We find that firms attracting abnormally high attention earn higher returns and the price pressure effect of SVI is stronger among small stocks. The predictability of searches for return persists three weeks and ultimate price reversal occurs within a year which shows buying pressure from uninformed investors. This study shows that the predictability of abnormal returns with ASVI is significant for three weeks in the Turkish stock market, different from Da et al. (2011)'s study which shows that predictability continues two weeks for US stock market. This difference shows that the time for buying pressure from uninformed investors in emerging markets like Turkey is higher than the time for buying pressure in developed markets. Our findings suggest that the stock prices in Turkey are affected more by the behavior of individual investors and therefore the effects of stock return predictability continue longer due to investor attention with higher market inefficiency.

We formed equal and value weighted portfolios by sorting long position in the stocks with high abnormal search volume and short position in the stocks with low abnormal search volume. These results show that forming a portfolio sorting by attention levels creates a significant return premium in short term. We find that trading strategy with long in high attention stocks and short in low attention stocks has a significant positive alpha per week.

We conclude that, abnormal Google Search Volume Index likely grabs the attention of individual investors resulting in short-term buying pressure. Our findings reveal that stock prices tend to be driven by the behavioral factors due to the investor attention in Turkey.

Social media platforms have evolved in recent years and the expansion of the social media platforms enable researchers to explore the complex behavior of the investors. We use firm-specific Twitter sentiment to investigate the informational role of social media in stock markets. This study presents the impact of Twitter activity and sentiment on stock returns and trading activity in multi-country level.

We focus on S&P 500, S&P 350 Europe and S&P Emerging Markets Core index constituents between the sample period of 2015 to 2017 with the international investor perspective because investors are mostly active on Twitter for larger firms. Larger firms have more attention and analyst coverage; thus, sentiment information could be obtained for these firms. Using a large sample of stocks from international stock markets, we find that daily Twitter activity and sentiment are associated with trading volume and predicts next-day trading volume. We show that the daily number of tweets and Twitter sentiment is associated with higher raw and abnormal stock returns. We find that daily firm-specific Twitter sentiment contains information for predicting future stock returns, but no such relation exists in the number of tweets or Twitter activity. This predictive power remains significant after controlling the news sentiment. The positive tone of Twitter sentiment has more predictive power in small and emerging market firms. These results are consistent with the literature stating that small firms are hard to value, and emerging market firms contain high information asymmetry. Overall, these results suggest that social media activity and sentiment provides new information about firms and show that social media present different impacts than traditional news media on firms' information environments. The results show that the role of social media is to diffuse sentiment to investors who unintentionally make prices less efficient in the short term.

From a practical perspective, investors could potentially use social media sentiment in their trading strategies. The predictive power of Twitter sentiment for stock returns may influence market participants' trading decisions. We present the long-short portfolio that longs in the stocks with highest decile Twitter sentiment and shorts in the stocks with lowest decile. We show that a trading strategy formed on Twitter sentiment generates significant positive returns even after considering trading costs. A trading strategy with a 5-days holding period presents significant annualized returns even after trading costs. Due to return reversals, these findings suggest that the predictive power is short sighted, and strategies might be formed only for short term investments.

Twitter is available for public and can be used for an investable trading strategy. The increasing amount of sentiment information obtained from social media sources such as Twitter may provide a valuable information and proxies for investor behaviors and beliefs for financial market participants. Firms can monitor these platforms to manage firm-specific investor sentiment that may affect firm performance. Information embedded in social media sentiment can be applied to various industries providing a monitoring and comparison of sentiment levels.

This study has some limitations. We focus on aggregate level Twitter sentiment and this study does not allow to make inferences about individual investor activities. We only focus on large-cap stocks and blue-chip companies in the international S&P indexes with the aim of analyzing international investor perspective since Twitter information is mostly available for larger firms. Thus, we have no empirical data to claim whether the results apply to smaller firms and the effects of social media sentiment on stock returns and trading activity for smaller firms can be investigated as a future work with the availability of sentiment data for smaller firms.

REFERENCES

Antweiler, W., and Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance, 59* (3), 1259-1294.

Aouadi, A., Arouri, M., and Teulon, F. (2013). Investor attention and stock market activity: Evidence from France. *Economic Modelling, 35*, 674-681.

Azar, P. D., and Lo, A. W. (2016). The wisdom of Twitter crowds: Predicting stock market reactions to FOMC meetings via Twitter feeds. *Journal of Portfolio Management, 42* (5), 123-134.

Baik, B., Cao, Q., Choi, S., and Kim, J. M. (2016). Local Twitter activity and stock returns. (Working Paper Series). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2783670.

Baker, M., and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance, 61*, 1645–1680.

Baker, M., and Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives, 21* (2), 129-152.

Bali, T. G., Peng, L., Shen, Y., and Tang, Y. (2013). Liquidity shocks and stock market reactions. *Review of Financial Studies, 27* (5), 1434-1485.

Bank, M., Larch, M., and Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management, 25* (3), 239-264.

Barber, B. M., and Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies, 21* (2), 785-818.

Barber, B. M., Odean, T., and Zhu, N. (2009). Do retail trades move markets? *Review of Financial Studies, 22* (1), 151-186.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics, 49* (3), 307-343.

Barberis, N., Shleifer, A., and Wurgler, J. (2005). Comovement. *Journal of Financial Economics, 75* (2), 283-317.

Bartov, E., Faurel, L., and Mohanram, P. S. (2017). Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review, 93* (3), 25-57.

Behrendt, S., and Schmidt, A. (2018). The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking and Finance, 96*, 355-367.

Bekaert, G., and Harvey, C. R. (2002). Research in emerging markets finance: looking to the future. *Emerging Markets Review, 3* (4), 429-448.

Ben-Rephael, A., Da, Z., and Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies, 30* (9), 3009-3047.

Bijl, L., Kringhaug, G., Molnár, P., and Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis, 45*, 150-156.

Black, F. (1986). Noise. *Journal of Finance, 41* (3), 528-543.

Böhme, R., and Holz, T. (2006). The effect of stock spam on financial markets. (Working Paper Series). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=897431.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2* (1), 1-8.

Brzeszczyński, J., Gajdka, J., and Kutan, A. M. (2015). Investor response to public news, sentiment and institutional trading in emerging markets: A review. *International Review of Economics and Finance, 40*, 338-352.

Bukovina, J. (2016). Social media big data and capital markets - An overview. *Journal of Behavioral and Experimental Finance, 11*, 18-26.

Cakici, N., Fabozzi, F. J., and Tan, S. (2013). Size, value, and momentum in emerging market stock returns. *Emerging Markets Review, 16*, 46-65.

Calomiris, C. W., and Mamaysky, H. (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics, 133* (2), 299-336.

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

Chai, D., Dai, M., Gharghori, P., and Hong, B. (2019). Internet search intensity and its relation with trading activity and stock returns. *International Review of Finance*. doi: 10.1111/irfi.12268.

Chen, H., De, P., Hu, Y. J., and Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies, 27* (5), 1367-1403.

Cheng, L. Y., Yan, Z., Zhao, Y., and Gao, L. M. (2015). Investor inattention and under-reaction to repurchase announcements. *Journal of Behavioral Finance*, 16 (3), 267-277.

Crowley, R., Huang, W., and Lu, H. (2018). Discretionary disclosure on Twitter. (Working Paper Series). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3105847.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *Journal of Finance, 66* (5), 1461-1499.

Da, Z., Engelberg, J., and Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies, 28* (1), 1-32.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance, 52* (3), 1035-1058.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *Journal of Finance, 53* (6), 1839-1885.

Das, S. R., and Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science, 53* (9), 1375-1388.

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

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy, 98* (4), 703-738.

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics, 118* (3), 909-968.

Dimpfl, T., and Jank, S. (2016). Can internet search queries help to predict stock market volatility? *European Financial Management, 22* (2), 171-192.

Ding, R., and Hou, W. (2015). Retail investor attention and stock liquidity. *Journal of International Financial Markets, Institutions and Money, 37*, 12-26.

Drake, M. S., Guest, N. M., and Twedt, B. J. (2014). The media and mispricing: The role of the business press in the pricing of accounting information. *The Accounting Review, 89* (5), 1673-1701.

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

Ekinci, C., and Bulut, A.E. (2018). Google search and stock returns. *Global Business Research Conference 2018*. Istanbul, Turkey.

Engelberg, J. E., and Parsons, C. A. (2011). The causal impact of media in financial markets. *Journal of Finance, 66* (1), 67-97.

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

Fama, E. F. (1991). Efficient capital markets: II. *Journal of Finance, 46* (5), 1575-1617.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics, 49* (3), 283-306.

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 French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics, 105* (3), 457-472.

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

Fang, L., and Peress, J. (2009). Media coverage and the cross-section of stock returns. *Journal of Finance, 64* (5), 2023-2052.

Fenzl, T., and Pelzmann, L. (2012). Psychological and social forces behind aggregate financial market behavior. *Journal of Behavioral Finance, 13* (1), 56-65.

Fisher, K. L., and Statman, M. (2000). Investor sentiment and stock returns. *Financial Analysts Journal, 56* (2), 16-23.

Griffin, J. M. (2002). Are the Fama and French factors global or country specific? *Review of Financial Studies, 15* (3), 783-803.

Griffin, J. M., Hirschey, N. H., and Kelly, P. J. (2011). How important is the financial media in global markets? *Review of Financial Studies, 24* (12), 3941-3992.

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

Grullon, G., Kanatas, G., and Weston, J. P. (2004). Advertising, breadth of ownership, and liquidity. Review of Financial Studies, 17 (2), 439-461.

Guner, N., Onder, Z., and Rhoades, S. D. (2004). Underwriter reputation and short-run IPO returns: A re-evaluation for an emerging market. Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=495742

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. and Black, W. C. (1995). *Multivariate Data Analysis*. New York: Macmillan.

Hirshleifer, D., and Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics, 36* (1-3), 337-386.

Ho, C., and Hung, C. H. (2009). Investor sentiment as conditioning information in asset pricing. *Journal of Banking and Finance, 33* (5), 892-903.

Hou, K., Xiong, W., and Peng, L. (2009). A tale of two anomalies: The implications of investor attention for price and earnings momentum. (Working Paper Series). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=976394

Huang, D., Jiang, F., Tu, J., and Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies, 28* (3), 791-837.

Hur, J., and Singh, V. (2017). Cross-Section of expected returns and extreme returns: The role of investor attention and risk preferences. *Financial Management, 46* (2), 409-431.

Jiang, S., Chen, H., Nunamaker, J. F., and Zimbra, D. (2014). Analyzing firm-specific social media and market: A stakeholder-based event analysis framework. *Decision Support Systems, 67*, 30-39.

Joseph, K., Wintoki, M. B., and Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting, 27* (4), 1116-1127.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kaniel, R., Starks, L. T., and Vasudevan, V. (2007). Headlines and bottom lines: attention and learning effects from media coverage of mutual funds. (Working Paper Series). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=687103

Karabulut, Y. (2013). Can facebook predict stock market activity? (Working Paper Series). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2017099

Karagozoglu, A. K., and Fabozzi, F. J. (2017). Volatility wisdom of social media crowds. The Journal of Portfolio Management, *43* (2), 136-151.

Kim, S. H., and Kim, D. (2014). Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior and Organization, 107*, 708-729.

Korkmaz, T., Cevik, E. I., and Cevik, N. K. (2017). The relationship between investor attention and stock markets: An application on ISE-100 index. *Business and Economics Research Journal, 8* (2), 203-215.

Kumar, A., and Lee, C. M. (2006). Retail investor sentiment and return comovements. *Journal of Finance, 61* (5), 2451-2486.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315-1335.

Lawrence, A., Ryans, J., Sun, E., and Laptev, N. (2016). Yahoo Finance search and earnings announcements. (Working Paper Series). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2804353

Lee, W. Y., Jiang, C. X., and Indro, D. C. (2002). Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking and Finance, 26* (12), 2277-2299.

Lee, C. M., Shleifer, A., and Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *Journal of Finance, 46* (1), 75-109.

Lee, C. M., Sun, S. T., Wang, R., and Zhang, R. (2019). Technological links and predictable returns. *Journal of Financial Economics, 132* (3), 76-96.

Leitch, D., and Sherif, M. (2017). Twitter mood, CEO succession announcements and stock returns. *Journal of Computational Science, 21*, 1-10.

Lemmon, M., and Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies, 19* (4), 1499-1529.

Lesmond, D. A. (2005). Liquidity of emerging markets. *Journal of Financial Economics, 77* (2), 411-452.

Li, J., and Yu, J. (2012). Investor attention, psychological anchors, and stock return predictability. *Journal of Financial Economics, 104* (2), 401-419.

Li, T., van Dalen, J., and van Rees, P. J. (2018). More than just noise? Examining the information content of stock microblogs on financial markets. *Journal of Information Technology, 33* (1), 50-69.

Liew, J. K. S., and Wang, G. Z. (2016). Twitter sentiment and IPO performance: a cross-sectional examination. *Journal of Portfolio Management, 42* (4), 129-135.

Liew, J., and Budavari, T. (2017). The "Sixth" Factor—A Social Media Factor Derived Directly from Tweet Sentiments. *Journal of Portfolio Management, 43* (3), 102-111.

Lin, M. C., Wu, C. H., and Chiang, M. T. (2014). Investor attention and information diffusion from analyst coverage. *International Review of Financial Analysis, 34*, 235-246.

Lintner, J. (1975). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *In Stochastic Optimization Models in Finance* (pp. 131-155). Academic Press.

Liu, L., Wu, J., Li, P., and Li, Q. (2015). A social-media-based approach to predicting stock comovement. *Expert Systems with Applications, 42* (8), 3893-3901.

Lou, D. (2014). Attracting investor attention through advertising. *Review of Financial Studies, 27* (6), 1797-1829.

Mao, H., Counts, S., and Bollen, J. (2011). Predicting financial markets: Comparing survey, news, twitter and search engine data. *arXiv preprint* arXiv:1112.1051.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7* (1), 77-91.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance, 42* (3), 483-510.

Mondria, J. and T. Wu. (2011). Asymmetric Attention and Stock Returns. (Working Paper, No 134). Retrieved from Society for Economic Dynamics website: https://www.economicdynamics.org/meetpapers/2011/paper_134.pdf

Nelson, K. K., Price, R. A., and Rountree, B. R. (2013). Are individual investors influenced by the optimism and credibility of stock spam recommendations? *Journal of Business Finance and Accounting, 40* (9-10), 1155-1183.

Nofsinger, J. R. (2005). Social mood and financial economics. *Journal of Behavioral Finance, 6* (3), 144-160.

Newey, W.K., and West, K.D. (1987), A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica, 55,* 703–708.

Peltomäki, J., Graham, M., and Hasselgren, A. (2018). Investor attention to market categories and market volatility: The case of emerging markets. *Research in International Business and Finance, 44,* 532-546.

Preis, T., Moat, H. S., and Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports, 3*.

Qian, M., Sun, P. W., and Yu, B. (2017). High turnover with high price delay? Dissecting the puzzling phenomenon for China's A-shares. *Finance Research Letters, 22*, 105-113.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PloS one, 10* (9), e0138441.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking and Finance, 84*, 25-40.

Sabherwal, S., Sarkar, S. K., and Zhang, Y. (2008). Online talk: does it matter? *Managerial Finance*, *34* (6), 423-436.

Sayim, M., and Rahman, H. (2015). The relationship between individual investor sentiment, stock return and volatility: evidence from the Turkish market. *International Journal of Emerging Markets, 10* (3), 504-520.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance, 19* (3), 425-442.

Shleifer, A., and Vishny, R. W. (1997). The limits of arbitrage. *Journal of Finance, 52* (1), 35-55.

Shiller, R. J. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity, 1984* (2), 457-510.

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

Shleifer, A., and Summers, L. H. (1990). The noise trader approach to finance. *Journal of Economic Perspectives, 4* (2), 19-33.

Siganos, A., Vagenas-Nanos, E., and Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior and Organization, 107*, 730-743.

Siganos, A., Vagenas-Nanos, E., and Verwijmeren, P. (2017). Divergence of sentiment and stock market trading. *Journal of Banking and Finance, 78*, 130-141.

Siikanen, M., Baltakys, K., Kanniainen, J., Vatrapu, R., Mukkamala, R., and Hussain, A. (2018). Facebook drives behavior of passive households in stock markets. *Finance Research Letters, 27*, 208-213.

Skoulakis, G. (2008). Panel Data Inference in Finance: Least-Squares vs Fama and MacBeth. (Working Paper Series). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1108865

Solomon, D. H., Soltes, E., and Sosyura, D. (2014). Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics, 113* (1), 53-72.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., and Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20* (5), 926-957.

Sul, H. K., Dennis, A. R., and Yuan, L. (2017). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences, 48* (3), 454-488.

Sun, A., Lachanski, M., and Fabozzi, F. J. (2016a). Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. *International Review of Financial Analysis, 48*, 272-281.

Sun, L., Najand, M., and Shen, J. (2016b). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking and Finance, 73*, 147-164.

Swamy, V., Dharani, M., and Takeda, F. (2019). Investor attention and Google Search Volume Index: Evidence from an emerging market using quantile regression analysis. *Research in International Business and Finance, 50*, 1-17.

Takeda, F., and Wakao, T. (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. *Pacific-Basin Finance Journal, 27*, 1-18.

Tan, S. D., and Tas, O. (2019). Investor attention and stock returns: Evidence from Borsa Istanbul. *Borsa Istanbul Review, 19* (2), 106-116.

Tantaopas, P., Padungsaksawasdi, C., and Treepongkaruna, S. (2016). Attention effect via internet search intensity in Asia-Pacific stock markets. *Pacific-Basin Finance Journal, 38,* 107-124.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance, 62* (3), 1139-1168.

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information? *Review of Financial Studies, 24* (5), 1481-1512.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance, 63* (3), 1437-1467.

Url-1 <*http://www.tuyid.org/files/yayinlar/Borsa_Trends_Report_XXIII.pdf*>, date retrieved 26.06.2019.

Url-2 <*http://gs.statcounter.com/search-engine-market-share*>, date retrieved 26.06.2019.

Url-3 <*https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/*>, date retrieved 26.06.2019.

Url-4 <*https://about.stocktwits.com/*>, date retrieved 26.06.2019.

Url-5 <*https://gs.statcounter.com/social-media-stats*>, date retrieved 26.06.2019.

Url-6 <*https://faculty.chicagobooth.edu/john.cochrane/teaching* >, date retrieved 26.06.2019.

Url-7 <*https://trends.google.com/trends/*>, date retrieved 26.06.2019.

Url-8 <*https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/*>, date retrieved 26.06.2019.

Url-9 <*https://developer.twitter.com/content/dam/developer-twitter/pdfs-and-files/Bloomberg-Twitter-Data-Research-Report.pdf* >, date retrieved 26.06.2019.

Url-10 <*https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html* >, date retrieved 20.01.2019.

Url-11 <*https://www.bloomberg.com/company/press/trending-on-twitter-social-sentiment-analytics*>, date retrieved 20.01.2019.

Uygur, U., and Tas, O. (2014). The impacts of investor sentiment on returns and conditional volatility of international stock markets. *Quality and Quantity*, *48* (3), 1165-1179.

Van Bommel, J. (2003). Rumors. *Journal of Finance, 58* (4), 1499-1520.

Vlastakis, N., and Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking and Finance, 36* (6), 1808-1821.

Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking and Finance*, *41*, 17-35.

Wang, X. W., Yan, Z., Zhang, Q., and Gao, X. (2018). Investor attention and stock market under-reaction to earnings announcements: Evidence from the options market. *Journal of Futures Markets*, *38* (4), 478-492.

Wysocki, P. D. (1998). Cheap talk on the web: The determinants of postings on stock message boards. (Working Paper Series). Retrieved from Social Science Research Network website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=160170

Ying, Q., Kong, D., and Luo, D. (2015). Investor attention, institutional ownership, and stock return: empirical evidence from China. *Emerging Markets Finance and Trade*, *51* (3), 672-685.

Yu, Y., Duan, W., and Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, *55* (4), 919-926.

Zhang, B., and Wang, Y. (2015). Limited attention of individual investors and stock performance: Evidence from the ChiNext market. *Economic Modelling*, *50*, 94-104.

Zhang, W., Li, X., Shen, D., and Teglio, A. (2016). Daily happiness and stock returns: Some international evidence. *Physica A: Statistical Mechanics and its Applications*, *460*, 201-209.

Zhang, W., Shen, D., Zhang, Y., and Xiong, X. (2013). Open source information, investor attention, and asset pricing. *Economic Modelling*, *33*, 613-619.

Zhang, X., Fuehres, H., and Gloor, P. A. (2011). Predicting stock market indicators through twitter “I hope it is not as bad as I fear”. *Procedia-Social and Behavioral Sciences*, *26*, 55-62.

APPENDICES

APPENDIX A: Country breakdown in the sample, Twitter activity, sentiment and stock returns (open-to-close returns)

APPENDIX A

Table A.1 : Country breakdown in the sample.

Index	Country	Number of constituents	Constituents (%, total)	Total market capitalization (US dollars, million)	Market capitalization (%, total)
S&P 350 Europe	United Kingdom	93	8.75%	1,706.36	6.48%
S&P 350 Europe	France	48	4.52%	1,083.44	4.12%
S&P 350 Europe	Germany	45	4.23%	1,093.17	4.15%
S&P 350 Europe	Switzerland	33	3.10%	1,043.50	3.96%
S&P 350 Europe	Netherlands	25	2.35%	468.17	1.78%
S&P 350 Europe	Spain	25	2.35%	597.48	2.27%
S&P 350 Europe	Sweden	25	2.35%	315.58	1.20%
S&P 350 Europe	Italy	21	1.98%	266.77	1.01%
S&P 350 Europe	Denmark	14	1.32%	234.84	0.89%
S&P 350 Europe	Belgium	10	0.94%	242.83	0.92%
S&P 350 Europe	Finland	9	0.85%	132.34	0.50%
S&P 350 Europe	Ireland	8	0.75%	114.06	0.43%
S&P 350 Europe	Norway	6	0.56%	107.89	0.41%
S&P 350 Europe	Luxembourg	4	0.38%	39.30	0.15%
S&P 350 Europe	Austria	3	0.28%	24.77	0.09%
S&P 350 Europe	Portugal	3	0.28%	18.55	0.07%
S&P 500	US	552	51.93%	15,732.47	59.76%
S&P EM Core	China	52	4.89%	1,820.97	6.92%
S&P EM Core	India	19	1.79%	381.17	1.45%
S&P EM Core	South Africa	15	1.41%	159.50	0.61%
S&P EM Core	Brazil	10	0.94%	245.85	0.93%
S&P EM Core	Mexico	8	0.75%	127.02	0.48%
S&P EM Core	Malaysia	7	0.66%	45.52	0.17%
S&P EM Core	Chile	6	0.56%	26.98	0.10%
S&P EM Core	Russia	6	0.56%	148.93	0.57%
S&P EM Core	Philippines	4	0.38%	37.28	0.14%
S&P EM Core	Thailand	4	0.38%	35.92	0.14%
S&P EM Core	Indonesia	3	0.28%	56.02	0.21%
S&P EM Core	Poland	3	0.28%	9.75	0.04%
S&P EM Core	Colombia	1	0.09%	3.58	0.01%
S&P EM Core	Czech Republic	1	0.09%	7.10	0.03%
		1063	100.00%	26,327.09	100.00%

Table A.2 : Twitter activity and returns (open-to-close raw returns).

Panel A. Contemporaneous regressions						
	Ret (open-to-close)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Ntweet	0.00012***	0.00019***	0.00012***	0.00024***	0.00004	0.00016**
	(4.38)	(4.56)	(3.97)	(6.16)	(1.24)	(2.04)
Ret[−5,−1]	0.0089***	0.0092***	0.0091***	0.0071***	0.0146***	0.0115***
	(5.52)	(5.57)	(4.78)	(3.30)	(6.20)	(4.80)
Vola[−5,−1]	-1.388***	-1.299***	-1.465***	-1.531***	-2.200***	-1.085***
	(-4.67)	(-4.32)	(-4.22)	(-3.52)	(-4.87)	(-3.00)
Illiq[−5,−1]	-0.0330	0.0246	0.0161	-0.204**	-0.0319	0.0111
	(-1.28)	(0.67)	(0.45)	(-1.97)	(-0.56)	(0.62)
$AbTurn_{t-1}$	0.0001***	0.0001	0.0002***	0.0002***	0.0001	0.00003
	(2.75)	(1.54)	(2.80)	(2.82)	(1.57)	(0.21)
Size	-0.0001***	0.0002*	-0.0003***	-0.0007***	-0.0001**	-0.00008
	(-3.65)	(1.94)	(-4.45)	(-5.40)	(-2.51)	(-1.15)
Intercept	0.0011***	-0.0012*	0.0008	0.0030***	0.0015**	0.0003
	(3.71)	(-1.79)	(1.58)	(4.99)	(2.51)	(0.51)
Region dummy	Yes	Yes	Yes			
R^2	0.1264	0.1379	0.1448	0.1067	0.1022	0.1155
N	722,795	360,803	361,781	380,704	237,693	104,398
Time periods	781	781	781	780	781	781
Panel B. Predicting returns						
	Ret (open-to-close)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
$Ntweet_{t-1}$	-0.00002	-0.00001	-0.0000008	-0.00002	-0.000004	0.00005
	(-1.03)	(-0.48)	(-0.03)	(-0.71)	(-0.12)	(0.64)
Ret[−5,−1]	0.0089***	0.0093***	0.0092***	0.0072***	0.0147***	0.0114***
	(5.51)	(5.65)	(4.82)	(3.35)	(6.23)	(4.75)
Vola[−5,−1]	-1.248***	-1.135***	-1.292***	-1.036**	-2.193***	-1.040***
	(-4.14)	(-3.70)	(-3.69)	(-2.28)	(-4.84)	(-2.88)
Illiq[−5,−1]	-0.0237	0.0192	0.0195	-0.211**	-0.0263	0.0125
	(-0.91)	(0.52)	(0.55)	(-2.02)	(-0.46)	(0.69)
$AbTurn_{t-1}$	0.0002***	0.0001*	0.0002***	0.0003***	0.0001	0.00004
	(3.16)	(1.96)	(3.02)	(3.43)	(1.64)	(0.28)
Size	-0.00006	0.0002**	-0.0001**	-0.0002**	-0.0001**	-0.00003
	(-1.41)	(2.49)	(-2.18)	(-2.01)	(-2.15)	(-0.44)
Intercept	0.0004	-0.0012*	0.0001	0.0016***	0.0013**	0.00002
	(1.50)	(-1.94)	(0.27)	(2.88)	(2.29)	(0.03)
Region dummy	Yes	Yes	Yes			
R^2	0.1258	0.1369	0.1440	0.1046	0.1018	0.1148
N	723,457	361,210	362,247	381,201	237,812	104,444
Time periods	782	782	782	781	782	782

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table A.3 : Twitter activity and returns (open-to-close abnormal returns).

Panel A. Contemporaneous regressions						
	AbRet (open-to-close)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Ntweet	0.00012***	0.00019***	0.00012***	0.00024***	0.00004	0.00017**
	(4.38)	(4.56)	(4.04)	(6.14)	(1.17)	(2.12)
AbRet[−5,−1]	0.0087***	0.0090***	0.0091***	0.0071***	0.0147***	0.0113***
	(5.39)	(5.43)	(4.74)	(3.31)	(6.22)	(4.72)
Vola[−5,−1]	-1.382***	-1.280***	-1.483***	-1.464***	-2.109***	-1.144***
	(-4.72)	(-4.36)	(-4.28)	(-3.39)	(-4.83)	(-3.17)
Illiq[−5,−1]	-0.0305	0.0282	0.0149	-0.204*	-0.0348	0.0109
	(-1.19)	(0.77)	(0.42)	(-1.95)	(-0.61)	(0.61)
AbTurn_{t-1}	0.0001***	0.0001	0.0002***	0.0002***	0.0001	0.00004
	(2.67)	(1.38)	(2.75)	(2.70)	(1.49)	(0.28)
Size	-0.0001***	0.0002**	-0.0003***	-0.0007***	-0.0001**	-0.00009
	(-3.63)	(1.97)	(-4.43)	(-5.41)	(-2.44)	(-1.28)
Intercept	0.0008***	-0.0006	0.0009*	0.0027***	0.0014**	0.0004
	(2.68)	(-0.98)	(1.75)	(4.98)	(2.38)	(0.53)
Region dummy	Yes	Yes	Yes			
R^2	0.1071	0.1235	0.1203	0.1065	0.1020	0.1151
N	719,656	359,385	360,056	378,768	236,783	104,105
Time periods	776	776	776	776	776	776
Panel B. Predicting returns						
	AbRet (open-to-close)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Ntweet_{t-1}	-0.00002	-0.00001	0.000002	-0.00002	-0.000006	0.00006
	(-0.94)	(-0.39)	(0.09)	(-0.69)	(-0.19)	(0.85)
AbRet[−5,−1]	0.0087***	0.0091***	0.0092***	0.0073***	0.0147***	0.0113***
	(5.38)	(5.48)	(4.78)	(3.36)	(6.21)	(4.66)
Vola[−5,−1]	-1.244***	-1.119***	-1.309***	-0.969**	-2.092***	-1.073***
	(-4.18)	(-3.73)	(-3.74)	(-2.15)	(-4.78)	(-2.97)
Illiq[−5,−1]	-0.0198	0.0246	0.0195	-0.210**	-0.0293	0.0139
	(-0.77)	(0.67)	(0.55)	(-2.00)	(-0.51)	(0.77)
AbTurn_{t-1}	0.0002***	0.0001*	0.0002***	0.0002***	0.0001	0.00005
	(3.07)	(1.75)	(2.99)	(3.32)	(1.54)	(0.34)
Size	-0.00006	0.0002**	-0.0001**	-0.0002**	-0.0001**	-0.00004
	(-1.41)	(2.54)	(-2.18)	(-2.04)	(-2.08)	(-0.64)
Intercept	0.0004	-0.0012*	0.0003	0.0013**	0.0012**	0.00006
	(1.24)	(-1.83)	(0.76)	(2.56)	(2.15)	(0.08)
Region dummy	Yes	Yes	Yes			
R^2	0.1064	0.1222	0.1195	0.1044	0.1015	0.1144
N	720,467	359,875	360,592	379,283	237,005	104,179
Time periods	777	777	777	777	777	777

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table A.4 : Twitter sentiment and returns (open-to-close raw returns).

Panel A. Contemporaneous regressions						
	Ret (open-to-close)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Twitter	0.00283***	0.00336***	0.00209***	0.00383***	0.00143***	0.00233***
	(23.69)	(19.89)	(17.16)	(22.10)	(9.92)	(5.16)
$Ret[-5,-1]$	0.0083***	0.0084***	0.0092***	0.0058***	0.0143***	0.0107***
	(5.07)	(5.00)	(4.82)	(2.65)	(6.06)	(4.31)
$Vola[-5,-1]$	-1.180***	-1.054***	-1.230***	-1.020**	-2.049***	-0.663*
	(-3.77)	(-3.33)	(-3.40)	(-2.26)	(-4.38)	(-1.70)
$Illiq[-5,-1]$	-0.0343	-0.0152	0.0218	-0.155	-0.0473	-0.0138
	(-1.28)	(-0.40)	(0.60)	(-1.52)	(-0.82)	(-0.40)
$AbTurn_{t-1}$	0.0001***	0.0001*	0.0002***	0.0002***	0.0001	0.00006
	(2.68)	(1.72)	(2.71)	(2.98)	(1.64)	(0.37)
Size	-0.0001**	0.0001	-0.0001**	-0.0002***	-0.0001**	-0.00007
	(-2.58)	(1.59)	(-2.40)	(-2.60)	(-2.58)	(-1.03)
Intercept	0.0002	-0.0015**	0.0004	0.0013***	0.0013**	0.0003
	(0.85)	(-2.13)	(1.28)	(2.76)	(2.48)	(0.51)
Region dummy	Yes	Yes	Yes			
R^2	0.1284	0.1407	0.1421	0.1104	0.0971	0.1111
N	775,091	376,798	398,067	404,043	274,382	96,666
Time periods	781	781	781	780	781	781

Panel B. Predicting returns						
	Ret (open-to-close)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
$Twitter_{t-1}$	0.00061***	0.00056***	0.00058***	0.00031**	0.00090***	0.00154***
	(6.08)	(4.23)	(4.83)	(2.58)	(6.33)	(3.99)
$Ret[-5,-1]$	0.0087***	0.0089***	0.0095***	0.0069***	0.0143***	0.0108***
	(5.30)	(5.28)	(5.00)	(3.16)	(6.04)	(4.33)
$Vola[-5,-1]$	-1.276***	-1.189***	-1.321***	-1.140**	-2.103***	-0.682*
	(-4.05)	(-3.69)	(-3.64)	(-2.51)	(-4.50)	(-1.76)
$Illiq[-5,-1]$	-0.0356	-0.00543	0.0167	-0.198*	-0.0419	-0.0132
	(-1.32)	(-0.15)	(0.46)	(-1.93)	(-0.72)	(-0.38)
$AbTurn_{t-1}$	0.0002***	0.0001**	0.0002***	0.0003***	0.0001*	0.00008
	(3.05)	(2.15)	(2.90)	(3.39)	(1.70)	(0.54)
Size	-0.0001***	0.0002**	-0.0001***	-0.0003***	-0.0001***	-0.00008
	(-2.78)	(2.03)	(-2.69)	(-3.16)	(-2.61)	(-1.20)
Intercept	0.00006	-0.0011	0.0005	0.0017***	0.0013**	0.0005
	(0.21)	(-1.64)	(1.43)	(3.62)	(2.55)	(0.68)
Region dummy	Yes	Yes	Yes			
R^2	0.1263	0.1383	0.1409	0.1053	0.0969	0.1109
N	775,755	377,200	398,555	404,564	274,490	96,701
Time periods	782	782	782	781	782	782

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table A.5 : Twitter sentiment and returns (open-to-close abnormal returns).

	Panel A. Contemporaneous regressions					
	AbRet (open-to-close)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Twitter	0.00281***	0.00332***	0.00208***	0.00382***	0.00140***	0.00205***
	(23.80)	(19.75)	(17.16)	(21.94)	(9.87)	(5.40)
AbRet[−5,−1]	0.0083***	0.0083***	0.0091***	0.0058***	0.0144***	0.0105***
	(5.00)	(4.94)	(4.77)	(2.66)	(6.11)	(4.20)
Vola[−5,−1]	-1.171***	-1.016***	-1.263***	-0.966**	-1.957***	-0.750*
	(-3.76)	(-3.24)	(-3.50)	(-2.15)	(-4.25)	(-1.95)
Illiq[−5,−1]	-0.0326	-0.0150	0.0220	-0.154	-0.0503	0.0141
	(-1.22)	(-0.40)	(0.61)	(-1.50)	(-0.87)	(0.72)
AbTurn_{t-1}	0.0001***	0.0001	0.0002***	0.0002***	0.0001	0.00008
	(2.64)	(1.60)	(2.70)	(2.88)	(1.55)	(0.54)
Size	-0.0001***	0.0001	-0.0001**	-0.0002***	-0.0001**	-0.00006
	(-2.62)	(1.53)	(-2.42)	(-2.62)	(-2.55)	(-0.88)
Intercept	0.0002	-0.0010	0.0001	0.0010**	0.0012**	0.0001
	(0.86)	(-1.54)	(0.44)	(2.26)	(2.32)	(0.25)
Region dummy	Yes	Yes	Yes			
R^2	0.1090	0.1260	0.1175	0.1101	0.0970	0.1100
N	771,908	375,415	396,262	401,989	273,497	96,422
Time periods	776	776	776	776	776	776
	Panel B. Predicting returns					
	AbRet (open-to-close)					
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
Twitter_{t-1}	0.00057***	0.00054***	0.00054***	0.00029**	0.00088***	0.00138***
	(5.86)	(4.13)	(4.62)	(2.42)	(6.19)	(3.90)
AbRet[−5,−1]	0.0086***	0.0088***	0.0095***	0.0070***	0.0144***	0.0105***
	(5.21)	(5.24)	(4.97)	(3.17)	(6.06)	(4.21)
Vola[−5,−1]	-1.256***	-1.105***	-1.337***	-1.078**	-2.008***	-0.749*
	(-4.02)	(-3.52)	(-3.70)	(-2.39)	(-4.37)	(-1.94)
Illiq[−5,−1]	-0.0330	-0.0073	0.0183	-0.197*	-0.0437	0.0124
	(-1.23)	(-0.20)	(0.50)	(-1.91)	(-0.75)	(0.64)
AbTurn_{t-1}	0.0002***	0.0001**	0.0002***	0.0002***	0.0001	0.0001
	(2.99)	(1.98)	(2.90)	(3.30)	(1.61)	(0.67)
Size	-0.0001***	0.0002*	-0.0001***	-0.0003***	-0.0001**	-0.00007
	(-2.81)	(1.94)	(-2.68)	(-3.17)	(-2.56)	(-1.10)
Intercept	0.0004	-0.0013*	0.0003	0.0014***	0.0012**	0.0003
	(1.30)	(-1.87)	(0.88)	(3.18)	(2.37)	(0.43)
Region dummy	Yes	Yes	Yes			
R^2	0.1067	0.1229	0.1162	0.1051	0.0968	0.1096
N	772,736	375,905	396,831	402,534	273,715	96,487
Time periods	777	777	777	777	777	777

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table A.6 : Twitter sentiment and returns with news sentiment (open-to-close returns).

Panel A. Raw return						
Ret (open-to-close)						
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
$Twitter_{t-1}$	0.0004***	0.0004***	0.0004***	0.0002**	0.0007***	0.0012***
	(5.22)	(3.51)	(4.05)	(2.32)	(5.65)	(3.21)
News [-1]	0.0005***	0.0006***	0.0004***	0.0001*	0.0007***	0.0016***
	(6.96)	(6.50)	(4.99)	(1.83)	(5.74)	(7.18)
$Ret[-5,-1]$	0.0084***	0.0085***	0.0094***	0.0069***	0.0139***	0.0102***
	(5.13)	(5.04)	(4.91)	(3.14)	(5.90)	(4.07)
$Vola[-5,-1]$	-1.257***	-1.193***	-1.283***	-1.130**	-2.070***	-0.669*
	(-3.99)	(-3.69)	(-3.55)	(-2.49)	(-4.43)	(-1.71)
$Illiq[-5,-1]$	-0.0347	-0.0037	0.0140	-0.201*	-0.0382	-0.0111
	(-1.29)	(-0.10)	(0.39)	(-1.96)	(-0.65)	(-0.32)
$AbTurn_{t-1}$	0.0001***	0.0001*	0.0002***	0.0002***	0.0001	0.00008
	(2.82)	(1.92)	(2.70)	(3.34)	(1.62)	(0.55)
Size	-0.0001***	0.0002*	-0.0001***	-0.0003***	-0.0001***	-0.0001
	(-2.96)	(1.91)	(-2.81)	(-3.20)	(-2.77)	(-1.48)
Intercept	0.0002	-0.0008	0.0007*	0.0017***	0.0013**	0.0004
	(0.88)	(-1.29)	(1.72)	(3.61)	(2.40)	(0.63)
Region dummy	Yes	Yes	Yes			
R^2	0.1281	0.1409	0.1441	0.1077	0.1017	0.1219
N	774,575	376,722	397,853	404,064	274,170	96,341
Time periods	782	782	782	781	782	782
Panel A. Abnormal return						
AbRet (open-to-close)						
	All	Small	Big	S&P 500	S&P 350E	S&P EM
	(1)	(2)	(3)	(4)	(5)	(6)
$Twitter_{t-1}$	0.00046***	0.00044***	0.00043***	0.00025**	0.00074***	0.00109***
	(4.96)	(3.47)	(3.81)	(2.18)	(5.50)	(3.05)
News [-1]	0.00057***	0.00066***	0.00048***	0.00014*	0.00078***	0.00166***
	(6.92)	(6.53)	(4.90)	(1.82)	(5.67)	(7.20)
$AbRet[-5,-1]$	0.0083***	0.0084***	0.0094***	0.0069***	0.0140***	0.0099***
	(5.04)	(5.02)	(4.87)	(3.16)	(5.90)	(3.94)
$Vola[-5,-1]$	-1.237***	-1.106***	-1.300***	-1.071**	-1.971***	-0.736*
	(-3.97)	(-3.52)	(-3.60)	(-2.37)	(-4.29)	(-1.90)
$Illiq[-5,-1]$	-0.0324	-0.0058	0.0154	-0.200*	-0.0401	0.0137
	(-1.21)	(-0.16)	(0.42)	(-1.94)	(-0.68)	(0.70)
$AbTurn_{t-1}$	0.0001***	0.0001*	0.0002***	0.0002***	0.0001	0.0001
	(2.76)	(1.76)	(2.69)	(3.24)	(1.53)	(0.63)
Size	-0.0001***	0.0001*	-0.0001***	-0.0003***	-0.0001***	-0.00009
	(-2.99)	(1.82)	(-2.81)	(-3.22)	(-2.72)	(-1.39)
Intercept	0.0002	-0.0013**	0.0002	0.0014***	0.0012**	0.0003
	(0.72)	(-1.97)	(0.58)	(3.18)	(2.23)	(0.39)
Region dummy	Yes	Yes	Yes			
R^2	0.1084	0.1255	0.1194	0.1075	0.1016	0.1205
N	771,585	375,448	396,137	402,058	273,400	96,127
Time periods	777	777	777	777	777	777

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table A.7 : Return reversals at different time horizons (all sample, open-to-close returns).

	All						
	AbRet (open-to-close)						
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 1-5	Day 6-10
$\mathrm{Twitter}_{t-1}$	0.00057***	-0.00006	0.00008	0.00009	-0.00002	0.00066**	-0.00051*
	(5.86)	(-0.68)	(0.99)	(1.08)	(-0.31)	(2.11)	(-1.69)
$\mathrm{Ret}[-5,-1]$	0.0086***	0.0042***	0.0034**	0.0025	0.0015	0.0204***	0.0119*
	(5.21)	(2.70)	(2.11)	(1.58)	(0.94)	(3.43)	(1.95)
$\mathrm{Vola}[-5,-1]$	-1.256***	-1.25***	-1.26***	-1.09***	-1.03***	-5.85***	-5.75***
	(-4.02)	(-3.88)	(-3.68)	(-3.17)	(-2.91)	(-4.26)	(-3.81)
$\mathrm{Illiq}[-5,-1]$	-0.0330	-0.0245	-0.0156	-0.0240	-0.0170	-0.116	-0.0515
	(-1.23)	(-0.89)	(-0.56)	(-0.83)	(-0.61)	(-1.08)	(-0.47)
AbTurn_{t-1}	0.0002***	0.00005	-0.00003	-0.0001*	-0.0001***	-0.00008	-0.00003
	(2.99)	(0.78)	(-0.60)	(-1.94)	(-2.67)	(-0.45)	(-0.21)
Size	-0.0001***	-0.0001***	-0.00009**	-0.00009**	-0.00008**	-0.0004***	-0.0004***
	(-2.81)	(-2.70)	(-2.32)	(-2.13)	(-1.97)	(-2.87)	(-2.61)
Intercept	0.0002	0.0006*	0.0003	0.0003	0.00006	0.0024**	0.0011
	(0.66)	(1.91)	(1.10)	(1.18)	(0.19)	(2.16)	(1.00)
Region dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.1067	0.1020	0.1014	0.1009	0.1000	0.1043	0.1028
N	772736	771674	770612	769550	768489	768489	763184
Time periods	777	776	775	774	773	773	768

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table A.8 : Trading strategy on Twitter sentiment (open-to-close returns, global portfolio).

	Equal-weighted			Value-weighted		
	(1)	(2)	(3)	(1)	(2)	(3)
MKT	-0.0006**	-0.0005**	-0.0002	-0.0006***	-0.0005**	-0.0003
	(-2.47)	(-1.99)	(-1.05)	(-3.17)	(-2.30)	(-1.43)
SMB		-0.0001	-0.00004		0.0002	0.0003
		(-0.32)	(-0.11)		(0.49)	(0.64)
HML		-0.0016***	-0.0006*		-0.0015***	-0.0007*
		(-4.94)	(-1.87)		(-4.31)	(-1.82)
UMD			0.0012***			0.0009***
			(3.93)			(3.04)
Alpha	0.000677***	0.000674***	0.000649***	0.000547***	0.000537***	0.000519***
	(6.37)	(6.46)	(6.48)	(5.43)	(5.35)	(5.22)
R^2	0.0206	0.062	0.1004	0.0231	0.0553	0.0749

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table A.9 : Trading strategy on Twitter sentiment (open-to-close returns, region portfolios).

	Panel A. US portfolio					
	Equal-weighted			Value-weighted		
	(1)	(2)	(3)	(1)	(2)	(3)
MKT	-0.0006***	-0.0005***	-0.0003**	-0.0003*	-0.0003*	-0.0001
	(-3.61)	(-3.37)	(-2.26)	(-1.93)	(-1.93)	(-1.02)
SMB		0.00007	0.0004		0.0001	0.0004
		(0.23)	(1.55)		(0.64)	(1.59)
HML		-0.0010***	-0.0001		-0.0011***	-0.0005**
		(-3.80)	(-0.48)		(-4.51)	(-2.04)
UMD			0.0015***			0.0010***
			(7.32)			(5.84)
Alpha	0.000301**	0.000296**	0.000290***	0.000157	0.000151	0.000147
	(2.50)	(2.52)	(2.62)	(1.31)	(1.29)	(1.29)
R^2	0.0189	0.0448	0.1546	0.0056	0.0378	0.0878
	Panel B. Europe portfolio					
	Equal-weighted			Value-weighted		
	(1)	(2)	(3)	(1)	(2)	(3)
MKT	-0.0004***	-0.0001	0.000009	-0.0005***	-0.0002	-0.0001
	(-2.59)	(-1.06)	(0.05)	(-3.11)	(-1.54)	(-0.69)
SMB		-0.0003	-0.0002		-0.0003	-0.0003
		(-0.77)	(-0.67)		(-0.76)	(-0.71)
HML		-0.0026***	-0.0017***		-0.0026***	-0.0019***
		(-7.23)	(-4.76)		(-6.44)	(-4.55)
UMD			0.0013***			0.0011***
			(4.46)			(3.83)
Alpha	0.000970***	0.000961***	0.000912***	0.000627***	0.000619***	0.000579***
	(7.30)	(7.37)	(7.26)	(4.58)	(4.50)	(4.34)
R^2	0.0126	0.1027	0.1343	0.0159	0.0916	0.109
	Panel C. Emerging markets portfolio					
	Equal-weighted			Value-weighted		
	(1)	(2)	(3)	(1)	(2)	(3)
MKT	-0.0894**	-0.0895**	-0.0893**	-0.0463	-0.0446	-0.0443
	(-2.17)	(-2.17)	(-2.20)	(-1.02)	(-0.97)	(-0.97)
SMB		-0.0261	-0.0287		0.0069	0.0066
		(-0.61)	(-0.55)		(0.16)	(0.15)
HML		-0.0049	-0.0055		-0.0068	-0.0071
		(-0.18)	(-0.18)		(-0.24)	(-0.25)
UMD			-0.0188			-0.0010
			(-0.31)			(-0.02)
Alpha	0.00105***	0.00103***	0.00105***	0.00106***	0.00105***	0.00106***
	(3.77)	(3.69)	(3.73)	(3.21)	(3.16)	(3.19)
R^2	0.0102	0.0107	0.0113	0.0017	0.0018	0.0018

The t-statistics are in parentheses. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table A.10 : Trading strategy on Twitter sentiment with trading costs (open-to-close returns).

Panel A. 1 day holding								
	Equal-weighted				Value-weighted			
	Global	US	Europe	EM	Global	US	Europe	EM
Trading Costs (bps)	Abnormal Annualized Returns (%)				Abnormal Annualized Returns (%)			
0	18.22	7.77	26.51	31.10	14.32	3.87	16.11	31.43
10	-7.58	-18.03	0.71	5.30	-11.48	-21.93	-9.69	5.63
20				-20.50				-20.17
Panel B. 5-days holding								
	Equal-weighted				Value-weighted			
	Global	US	Europe	EM	Global	US	Europe	EM
Trading Costs (bps)	Abnormal Annualized Returns (%)				Abnormal Annualized Returns (%)			
0	18.35	7.83	26.69	31.30	14.43	3.89	16.22	31.64
10	13.15	2.63	21.49	26.10	9.23	-1.31	11.02	26.44
20				20.90				21.24

CURRICULUM VITAE

Name Surname	**:** Selin Düz Tan
Place and Date of Birth	**:** Edirne, 20.11.1986
E-Mail	**:** selinduztan@gmail.com

EDUCATION **:**

- **B.Sc.** **:** 2008, Hacettepe University, Faculty of Economics and Administrative Sciences, Department of Economics
- **M.Sc.** **:** 2010, Lund University, Sweden, School of Economics and Management, Department of Economics

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2019 – 2020, Erasmus University Rotterdam, The Netherlands, Erasmus School of Economics, Visiting Researcher
- 2015 – present, Borsa Istanbul, Specialist
- 2011 – 2015, Yapı Kredi Bank, Credit and Market Risk Internal Auditor
- 2010 – 2011, Vakıfbank, Research Analyst
- 2018 – 2019, Jean Monnet Research Scholarship

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Tan, S. D.,** Tas, O. (2019) Investor Attention and Stock Returns: Evidence from Borsa Istanbul. Borsa Istanbul Review, 19 (2), 106-116.
- **Tan, S. D.,** Tas, O., Firm-Specific Investor Attention and Stock Returns in Turkey, Istanbul Finance Congress, Bosphorus University, Istanbul, Turkey, November 2017.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- Tas, O., **Tan, S. D.** (2016) Corporate Governance, Foreign Ownership and Stock Price Synchronicity: Evidence from Borsa Istanbul. Journal of Economics Finance and Accounting, 3(2), 156-165.
- **Tan, S. D.**, Tas, O., Foreign Ownership and Stock Price Synchronicity: Evidence from Turkey, Global Business Research Congress, GBRC-2016, Istanbul, Turkey, May 2016.