Dissertation

Information Asymmetry and Information Dissemination in High-Frequency Capital Markets

Thomas Pöppe

Dissertation

Vom Fachbereich für Rechts- und Wirtschaftswissenschaften,
Technische Universität Darmstadt (D17)
zur Erlangung des akademischen Grades Doctor rerum politicarum (Dr. rer. pol.)
genehmigte Dissertation

Information Asymmetry and Information Dissemination in High-Frequency Capital Markets

vorgelegt von

Dipl.-Wirtsch.-Inf. Thomas Pöppe
geboren in Frankfurt, Deutschland

Einreichungsdatum: 03.06.2015
Datum der mündlichen Prüfung: 15.10.2015
Erstgutachter: Prof. Dr. Dirk Schiereck
Zweitgutachter: Prof. Dr. Alexander Benlian

Darmstadt, 2016

Overview

Table of Contents

List of Tables

List of Tables in Appendix

List of Figures

List of Abbreviations and Frequently Used Symbols

alpha $\boldsymbol{\alpha}$: Parameter of the market microstructure trading model described in chapter 2. Denotes the probability of the existence of new information.

ASX: Australian Stock Exchange

BVC: Bulk volume classification

CLNV: Reference to the trade classification algorithm developed by Bidisha Chakrabarty, Bingguang Li, Vanthuan Nguyen and Robert A.Van Ness.

CME: Chicago Merchantile Exchange

delta $\boldsymbol{\delta}$: Parameter of the market microstructure trading model described in chapter 2. Denotes the probability of new information containing bad news.

EMO: Reference to the trade classification algorithm developed by Katrina Ellis, Roni Michaely, Maureen O'Hara.

epsilon $\boldsymbol{\varepsilon}$*:* Parameter of the market microstructure trading model described in chapter 2. Denotes the probability that an uninformed trader actually trades.

FWB: Frankfurter Wertpapierboerse

LR: Reference to the trade classification algorithm developed by Charles Lee and Mark Ready.

mu $\boldsymbol{\mu}$: Parameter of the market microstructure trading model described in chapter 2. Denotes the probability that a trade comes from an informed trader.

NASDAQ: National Association of Securities Dealers Automated Quotations. Electronic trading platform in the US.

NYSE: New York Stock Exchange

PIN: Probability of Informed Trading

SD: Standard deviation

VPIN: Volume-synchronized Probability of Informed Trading

XETRA: Electronic trading system of Deutsche Boerse.

1 Introduction

This dissertation is concerned with information asymmetry and information dissemination in high-frequency capital markets. The concepts of information asymmetry and information dissemination are elusive and ambiguous, on the one hand. At the same time, however, they are ubiquitous in financial research concerning capital markets and asset pricing. Market microstructure research provides the complementary tools to specify and concretize these concepts and quantify them to measure trends, changes and (causal) relationships.

At the intersection of information dissemination and asymmetry with market microstructure, this dissertation pursues three major goals. We propose enhancements to market microstructure methodology to be able to empirically conduct research on information dissemination and asymmetry on recent, high-frequency trading data. Second, we empirically evaluate related microstructure methodology to test its robustness and guide researchers in its application. Third, we employ the proposed methodology to evaluate the efficacy of different information channels, both traditional, legislation-based and new, technology-based channels. Is one of those channels superior in disseminating information efficiently and effectively? Is information asymmetry between different market participants actually reduced by the release of new information, and if so, by how much and for how long? How fast does information travel, how long require capital market to incorporate new information? What types of traders react to what type of information? When do investors react, right after or already before an official announcement? These are a few exemplary research questions that we intend to answer. We aim to advance the knowledge of our understanding of how information is transmitted to financial markets, what channels are relevant to investors, how market participants react and, finally, what methods are appropriate to conduct research in this field.

Understanding how information is incorporated into prices is crucial not only for researchers but also for everyone investing, directly or indirectly, in securities. Given this importance, this topic has been at the forefront of many researchers already for decades, of course. However, two major trends affect capital markets and both require reviewing and rethinking previous theoretical and empirical evidence. First, trading volume has been growing exponentially over the last decades. As depicted in figure 1-1, the New York Stock Exchange (NYSE) executed about 19 million trades in 1990. In the following exponential rise over almost three decades, this number crossed the mark of 1 trillion in 2007. Only the financial crisis in 2008 could temporarily halt this trend.

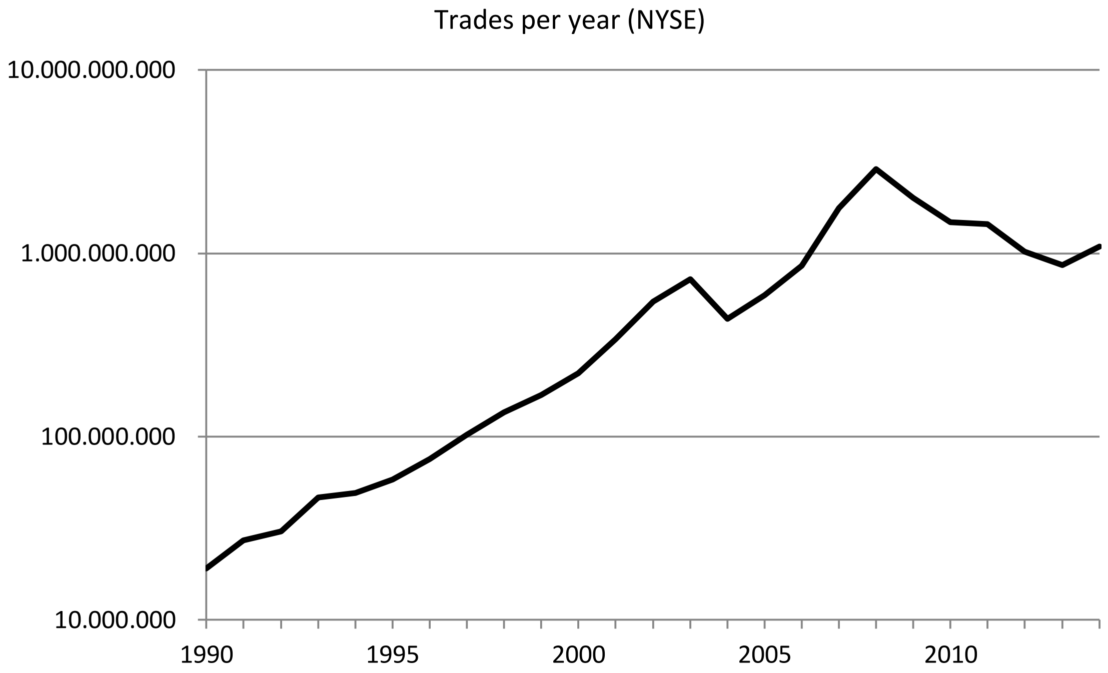

Figure 1-1: Historical trade count NYSE Group

This figure shows the aggregated number of trades per year handled by the NYSE Group from 1990 up to 2015. The vertical axis is in logarithmic scale and starts at 10 Mio. Source: Own analysis based on data from the NYSE Transactions, Statistics and Data Library: https://www.nyse.com/data/ transactions-statistics-data-library

The second trend is the increasing speed with which information is generated and disseminated, propelled through the technological revolution of the internet and mobile communications. The social media and micro-blogging platform Twitter.com, which enables the distribution of short messages within the blink of an eye to millions of people around the world, released its first message, a so-called “tweet”, in 2006. In 2007 on average 5000 of messages were tweeted per day. By 2013, the number of messages released reached an average of 500 million *per day* (see figure 1-2). Twitter is just one example, but stands representative for the rise of social media and Web 2.0 where users switch from content consumers to content creators, enabled by standardized platforms on the internet.

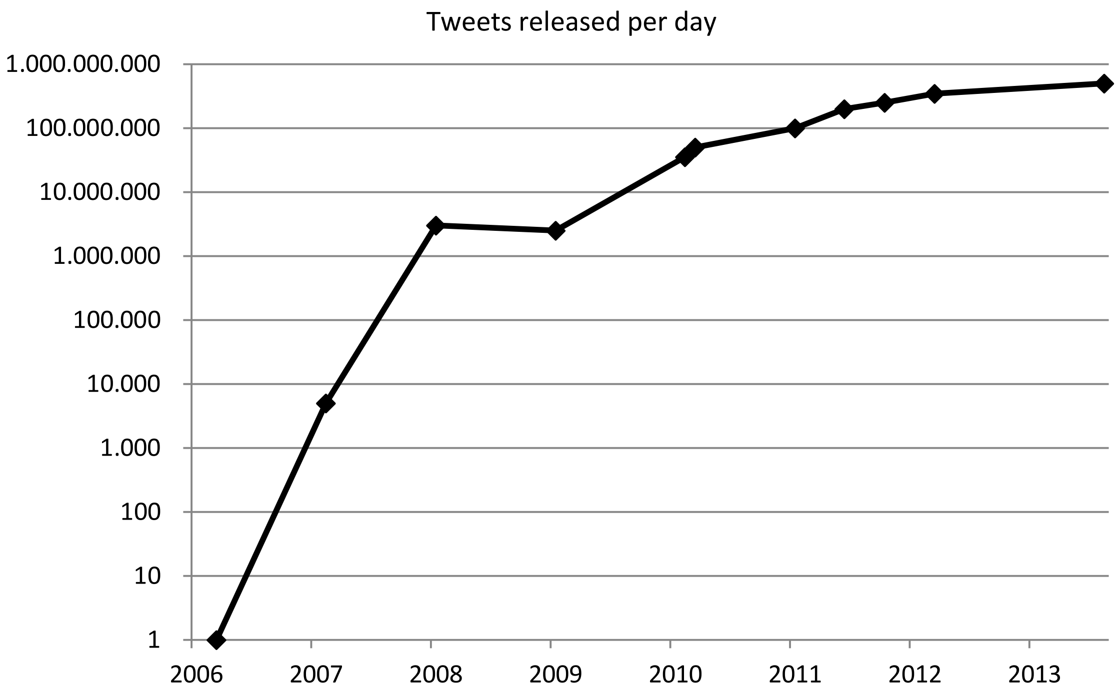

Figure 1-2: Tweets per day since start of Twitter

This figure shows the evolution of the number of tweets sent per day on the platform Twitter.com since its Go-Live on 2006. The vertical axis is in logarithmic scale. Source: http://www.internetlivestats.com/Twitter-statistics/

At the same time, not only information dissemination became easier and faster, but also information research. To “google” something has become a widely used phrase. Empirical evidence confirms that investors embrace these new technologies, as Google Search Volume is a very good indicator of investor attention (Da, Engelberg, & Gao, 2011; Pöppe, Schiereck, & Zielinski, 2014).

The emergence of market microstructure research for capital markets in the form covered in this dissertation and applied today in financial research is roughly four decades old. None of the foundational papers that we discuss in the next chapter could have imagined and incorporated high-frequency trading or the ubiquity of information retrieval through the digitalization of our society. In the study from Easley, Engle, O'Hara, and Wu (2008), for example, several hundred trading days, even from companies such as Enron or JP Morgan, need to be dropped from the sample due to inactivity. Hence, previous findings as well as methodology and theory must be challenged and adapted to this new environment. This is exactly where this dissertation starts off.

Our results provide insight for the research community and likewise offer important implications for investors, regulators and listed companies. Results are of both empirical and theoretical nature. Large, granular, high-frequency intraday datasets of trades and news are combined with state-of-the-art econometric analysis. Let us outline in the following paragraphs what the succeeding chapters accomplish.

Chapter 3 develops a new intraday estimation procedure for the microstructure trading model initially proposed by Easley, Kiefer, and O'Hara (1997b) to show how official announcements stipulated in German insider trading legislation significantly reduce information asymmetry upon public disclosure. Using a full year of intraday trading data for the top 100 German stocks, we demonstrate how the new estimation procedure eliminates or significantly reduces the shortcomings of the original approach in recent, high-frequency trading environments. Convergence rates are above 95% for the most liquid stocks and the intraday probability of informed trading can be applied in short horizon event studies. Further, the model's underlying assumptions of independence are fulfilled to a much higher degree.

The fourth chapter evaluates the recently published metric VPIN (Easley, Lopez de Prado, & O'Hara, 2012b) which aims to detect and predict the toxicity of order flow. We study empirically the sensitivity and robustness of VPIN to the choice of the trade classification scheme, the major input to compute VPIN. We compare deterministic trade-by-trade classification approaches with results computed with the newly proposed heuristic approach, bulk volume classification. On all levels up the ladder of aggregation, that is, for trade classification, for order imbalance, for VPIN and for identifying "toxic periods", we report substantial differences in the results. The detection of toxic periods does not yield consistent results more often than in 60% of all cases. Regression analysis identifies volume and return volatility as parameters that contribute to a higher sensitivity.

Chapter 5 comprises the third major empirical analysis in this dissertation and focuses on a new, very fast way of information generation and dissemination - the social media platform Twitter. Twitter was a key enabler in the Arab spring, made politicians resign and its fiercest discussions sometimes even make it into headline evening news. This study is the first to assess the relevance of Twitter for the information dissemination in financial markets on a per-stock basis and a comprehensively exhaustive sample. For the top 100 German stocks, we include all Twitter feeds that mention a company. In particular, our sample is not a random subsample of the full Twitter stream, not filtered by $- or €-tags and not diluted by an arbitrary choice of sentiment extraction. Instead, we measure the true attention for a stock based on its worldwide tweet activity. Besides common metrics for trading activity and market quality, we measure the relation of Twitter and an intraday version of the Probability of Informed Trading (PIN). Results indicate Twitter to be a platform for near real-time post-processing of information but with very limited, if any at all, predictive power regarding future volume or volatility. Thus, attention on Twitter follows the market and not vice versa.

The following chapter 2 will introduce the reader to the microstructure trading model that forms the foundation for the empirical work in chapter 3, is referred to in chapter 4 and employed once more in chapter 5, respectively. We start with explaining the microstructure trading model by Easley et al. (1997b), step-by-step, including the maximum likelihood estimation required to gain actual parameter estimates for quantitative analysis. Next, since every model is only a simplified projection of the true world, we discuss the core assumptions of the model regarding their validity. The subsequent section looks backward and forward in the literature to show where the model by Easley et al. (1997b) originates from and how scholars have progressed in its evolution.

2 The Microstructure Trading Model by Easley et al. (1997)

Market microstructure in terms of the finance domain analyzes and models the trading of financial assets, such as equity, bonds or derivatives. To do so, it studies the processes of how and why participants decide to trade, how they act to trade and how, finally, supply and demand are matched. Maureen O'Hara, a renowned researcher in the field, describes market microstructure as "the study of the process and outcomes of exchanging assets under explicit trading rules" (O'Hara, 1995).

The works of Easley, Kiefer, O'Hara, and Paperman (1996a) and Easley et al. (1997b) belong to the most prominent and most-cited in the market microstructure domain[1]. These well-published models serve as foundation for both theoretical and empirical work in this dissertation. In the following paragraphs we will refer to Easley et al. (1997b) and Easley et al. (1996a) as EKO97 and EKOP96, respectively. The two share the same foundations (Easley & O'Hara, 1987, 1992; Glosten & Milgrom, 1985) and are almost identical apart from two key differences: in EKOP96, the arrival of traders is modeled in continuous time with Poisson process while the model in EKO97 operates in discrete time and uses arrival probabilities. On the other hand, the EKO97 model integrates one property of the trading process that is not represented in EKOP96: the concept of deriving information from the absence of trades (Easley & O'Hara, 1992). We think this property is not only quite interesting from a theoretical and modeling point of view but also a more realistic approximation of the true trading process and hence describe and work with the EKO97 model, although description, implications and probably results based on the EKOP96 model would be very similar.

2.1 The trading process and market participants

This section describes the sequential, discrete-time microstructure trading model formulated in EKO97. Investors trade a single risky asset and money with a market maker. The market maker is risk neutral and competitive. These assumptions imply zero profit for the market maker and allow to ignore the effect of inventory cost on prices. Further, the competitive setting implies the possibility of competitors to enter the market, hence the assumption of just one market maker is not a restriction. The market maker quotes prices at which he or she is willing to buy or sell the asset. Traders, also risk-neutral, arrive sequentially, check the quote and decide whether to trade or not.

[1] We will give an overview of the model's numerous applications in empirical finance research in chapter 3.

Let the trading days be indexed by $t = 1 \ldots T$. The true value of the asset, represented by the random variable V, depends on the arrival of new information, which is defined as a signal ψ about V. Signals are assumed to arrive independently at the start of every trading day with probability α (and $1 - \alpha$ for no signal). A signal conveys bad news with probability δ and good news with probability $1 - \delta$, denoted by $\psi_t = L$ and $\psi_t = H$, respectively. The occurrence of no signal is denoted by $\psi_t = 0$. Let the true value of the asset in trading day t conditional on good news be $\overline{V_t}$, conditional on bad news be $\underline{V_t}$ and be V_t^* if no signal occurred. Of course it must hold that $\underline{V_t} < V_t^* < \overline{V_t}$. The signal $\psi_t$and the true value $\underline{V_t}, V_t^*$ or $\overline{V_t}$ of the asset will become public knowledge at the end of each trading session.

Signals are only observed by a fraction of traders, the informed traders. As they know the true value of the asset, they will trade as long as the true value is outside the current quote set by the market maker and will not trade if there is no signal. All other traders are uninformed traders, trading for liquidity reasons such as timing of consumption or portfolio adjustments, independent of the occurrence of a signal. The market maker also does not observe any signals, nor is he able to distinguish informed and uninformed traders.

The trading process is modeled from the perspective of the market maker, who tries to infer the true value of the asset to adjust his quotes accordingly. The process runs in discrete time where traders arrive sequentially, one at a time, check the quote and decide to buy, sell or not to trade given the quoted bid and ask. Explicitly modeling the no-trade trade outcome allows that quotes can adjust without any trade happening. This is the invention of the EKO97 model over the continuous model EKOP96. The trading session is therefore divided into intervals $i = 1 \ldots I$, long enough to accommodate one trade (or count as no-trade interval).

Each trading day is a realization of firstly a random draw deciding on the occurrence of a signal, followed by the realization of a random process where for each trading interval a buy, a sell or a no-trade is realized. Conditional on a signal occurring, informed traders always want to trade in the direction of the signal and will be able to do so with probability μ[2]. With probability $1 - \mu$ uninformed traders will be given the chance to trade. They

[2] In addition to assuming risk-neutrality, we also rule out strategic behavior to influence prices, which ensures that the informed will always trade on one side of the market if the price level has not yet adjusted.

decide to trade with probability ε, or decide not to trade with probability $1-\varepsilon$. If there is no signal, only uninformed traders decide to trade with probability ε, or decide not to trade with probability $1-\varepsilon$. Uninformed traders are assumed to split their trades equally into buys and sells. The structure described so far is depicted in figure 2-1.

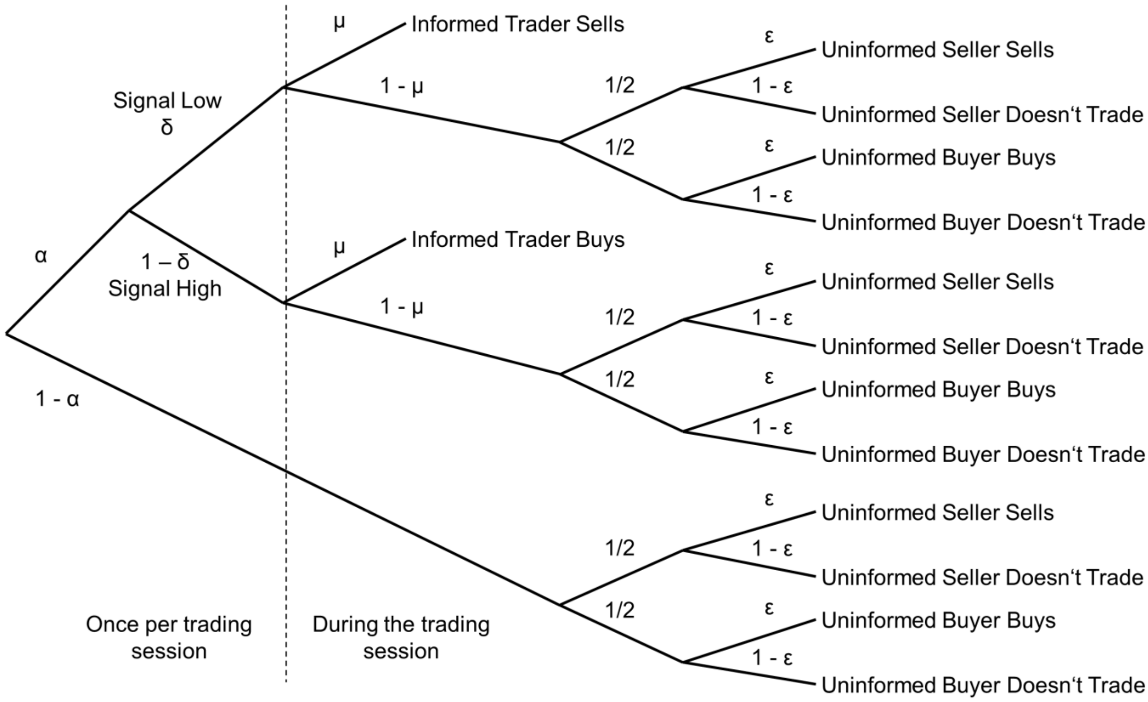

Figure 2-1: Tree diagram of the trading process

This figure gives the structure of the trading process. The part of the tree left to the dotted line occurs at the beginning of each trading session. The part to the right of the dotted line is iterated for each trading interval. The whole tree is iterated several times per trading day (or session), depending on the length of a bucket. Alpha is the probability of an information event, delta is the probability of information being bad news, mu is the probability that the trade comes from an informed trader, epsilon is the probability that an uninformed trader actually trades.

The market maker now needs to set his quotes such that the losses incurred to the informed traders are compensated by the earned spread from trades with the uninformed traders. The market maker is assumed to know this trading structure and rationally update his believes in Bayes' manner after every trade. His knowledge includes specific believes about the values of the parameters $\alpha, \delta, \varepsilon$ and μ, which he keeps constant at least for a full trading session. He does not know the current realization of the two reoccurring random events - signal and trade originator. However, knowing the trading structure allows him to gradually infer the existence and the direction of potential information. For instance, a buy is most likely if there is positive news, hence he will raise the quotes. A no-trade makes the absence of information more likely. Given a no-trade after a buy, this increases the probability that the previous buy was from an uninformed trader. Hence, the market maker would partially reverse his initial response to the buy and slightly lower quotes again. In general, the time between trades or the absence of trades provides information to the marker maker about the *existence* of a signal. Whether an actual trade is a buy or a sell provides information about the *direction* of a signal[3].

2.2 Estimation of model parameters and PIN

The result of the market makers thought process, which is the quoted spread, and the inputs to his thinking, which are buys, sells and no-trade intervals, allow us to derive the parameters underlying the trade process via maximum likelihood estimation. Following Easley, Kiefer, & O'Hara (1997b) we specify the probability of observing a combination of buys B, sells S and no-trades N in terms of the parameters of the model.

For a day without a signal, that is

$$\Pr\{B,S,N|\psi=0\}=[1/2\,\varepsilon]^B[1/2\,\varepsilon]^S[(1-\varepsilon)]^N \tag{2-1}$$

Similar, for a trading day with a positive signal the probability is given by:

$$\Pr\{B,S,N|\psi=H\}=[\mu+(1-\mu)\,1/2\,\varepsilon]^B[(1-\mu)\,1/2\,\varepsilon]^S[(1-\mu)(1-\varepsilon)]^N \tag{2-2}$$

Multiplying these conditioned probabilities with the probability that the conditioned event is occurring gives the unconditional probability. Summing these three products yields the likelihood function of observing a certain combinations of buys, sells and no-trades for a single day, conditional on the parameters $\alpha, \delta, \varepsilon$ and μ:

$$\begin{aligned}\mathrm{P}\{\mathrm{B},\mathrm{S},\mathrm{N}|\alpha,\delta,\varepsilon,\mu\}=\alpha(1-\delta)\,\mathrm{P}\{B,S,N|\psi=H\}+\alpha\delta\ \mathrm{P}\{B,S,N|\psi=L\}+\\(1-\alpha)\,\mathrm{P}\{B,S,N|\psi=0\}\end{aligned} \tag{2-3}$$

The assumption of information events arriving independently between the trading days and the assumed stationarity of the parameter set over a set of trading days allows to form the likelihood function over a period of *K* days simply as product over the single days:

$$P\{(B_t,S_t,N_t)_{t=1}^T|\ \alpha,\delta,\varepsilon,\mu\}=\prod_{t=1}^{T}P(B_t,S_t,N_t)|\ \alpha,\delta,\varepsilon,\mu\} \tag{2-4}$$

[3] The concept of deriving information from the absence of trades was introduced in Easley and O'Hara (1992).

A log transformation, simplifying and dropping a constant term results in the following maximum likelihood function[4]

$$\sum_{t=1}^{T} log\left[\alpha(1-\delta)\left(1+\frac{\mu}{x}\right)^{B} + \alpha\delta\left(1+\frac{\mu}{x}\right)^{S} + (1-\alpha)\left(\frac{1}{1-\mu}\right)^{S+B+N}\right] + \sum_{t=1}^{T} log\left[\left((1-\mu)(1-\varepsilon)\right)^{N} x^{S+B}\right] \tag{2-5}$$

where $x = (1-\mu)\frac{1}{2}\varepsilon$.

The probability of informed trading was defined as "the probability that any trade that occurs at time t is information-based" (Easley et. al. 1996):

$$PIN96 = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \tag{2-6}$$

The PIN measures the share of informed trades relative to all trades. In other words, it is the probability of an informed trade conditional on the probability that some trade occurs. The above formula was given in the continuous model of the PIN, where the arrival of informed and uninformed is modeled by independent Poisson processes that run in parallel (Easley et. al., 1996). In the discrete model presented here, if new information exists, the trade of an uninformed trader must be conditioned on the informed trader not being selected. As can easily be read from the tree diagram, the probability of an uninformed trade at any time t is given by

$$\begin{aligned} P\{uninformed\ trade\} &= P\{uninformed\ trade|\psi = 0\} \\ &+ P\{uninformed\ trade|\psi \neq 0\} = (1-a)\varepsilon + a(1-\mu)\varepsilon \end{aligned} \tag{2-7}$$

The chance of a trade occurring at any time t is therefore:

$$\begin{aligned} &P\{informed\ trade\} + P\{uninformed\ trade\} \\ &= \alpha\mu + (1-\alpha)\varepsilon + \alpha(1-\mu)\varepsilon \\ &= \alpha\mu + \varepsilon(1-\alpha\mu). \end{aligned} \tag{2-8}$$

This results in a slightly different PIN formula:

$$PIN97 = \frac{\alpha\mu}{\alpha\mu + \varepsilon(1-\alpha\mu)} \tag{2-9}$$

[4] The derivation in the original paper EKO97 provides more detail in certain steps than presented here.

The missing scalar 2 of ε is negligible. The main difference to the PIN96 formular is ε being lowered by the factor $(1 - \alpha\mu)$, which is due to the sequential setting of the model that rules out "parallel" arrival of informed and uninformed traders.

2.3 Discussion of key assumptions

The discussed model is based on a number of key assumptions whose validity requires discussion, as not all of them look plausible at first sight. In this chapter, we discuss the fundamental assumptions qualitatively from a theoretical point of view and argue with results from the previous literature. In chapter 3, these assumptions will be tested empirically. We summarize the restrictions imposed by the authors on the trading process as follows:

Assumption 1: Information events or signals occur randomly (a) and outside of trading hours (b).

Assumption 2: All information is incorporated into prices at the end of the trading day.

Assumption 3: The arrival of traders, informed or uninformed, is exogenous, and buy and sell orders arrive independently.

The random arrival of information signals as part of assumption 1 seems reasonable. The parameter *alpha* can vary between stocks and estimation periods but is constant within each estimation period. The studies from Easley et al. evaluate *alpha*. In EKO97, the independence of information signals is tested with a runs test, where the realization of *alpha* from the maximum likelihood estimation serves as a threshold to assign each trading day in the estimation window as a day with an information signal or a day without an information signal. The resulting series of days with a signal and days without a signal is then tested for independence. For the dataset used in EKO97, the assumption of independence is not rejected. In related work, Easley, O'Hara, and Paperman (1998b) test whether the number of analysts following a certain stock influences the probability of information signals. The hypothesis is rejected in the cited work, which is another indicator for the exogeneity of information signals. The second part of the assumption restricts new signals to occur only outside of trading hours, which is obviously contradictory to reality. In this case, however, the authors argue that the relaxation of this assumption is actually possible without loss of generalization: “this complicates the analysis, but does not change the results” (Easley & O'Hara, 1987).

Assumption 2 expects the market maker to extract information about the existence and the direction of a new information signal from the order flow, based on which he adjusts prices accordingly to the new true value of the traded asset. The process of price adjustment should be completed within a trading day. The literature on the speed of information processing strongly supports part (a) of this assumption and even suggests time spans shorter than a day (see for example Barclay & Litzenberger, 1988; Green, 2006; Muntermann & Guettler, 2007; Patell & Wolfson, 1984). A different theoretical solution is proposed by Admati and Pfleiderer (1988), who assume that the previously private information becomes public at the end of the trading day. Hence, also this second assumption is acceptable without significant loss of generality.

The independence of trading days as noted in assumption 3 is required to derive the maximum likelihood function as a product of the probabilities of the single trading days within the observation period. Without this independence, the parameters could not be estimated. Similar to the independence of news, EKO97 design an empirical test, which does not reject the hypothesis of buys and sells being independent across trading days. We will also evaluate this property once more on recent trading data in chapter 3.

Another crucial property of assumption 3 is the exogeneity and independent arrival of orders according to fixed arrival probabilities. Thereby strategic behavior from all types of traders is ruled out. In today’s markets, however, strategic behavior, whether it is based on information, like slicing orders to hide one’s true intention, or simply based on speed and volume, like pump-and-dump schemes, marking the close or other algorithmic trading strategies, are clearly present. The theoretical literature predicts various possibilities for strategic behavior. In a fundamental work, Kyle (1985) describes a trading model where strategic behavior is incorporated and investors endogenously decide how to trade. A monopolistic insider tries to maximize her profit by choosing the right order size to hide information from the noise traders and market makers, who all observe the order flow. Uninformed investors may also trade strategically. In a model by Allen and Gale (1992), uninformed investors generate profit by mimicking the behavior of an informed trader.

One more restriction implicitly following from assumption 3 is the uniform arrival of orders throughout the day. This restriction may yet as well be breached by the existence of strategic trading. Notable theoretical and empirical research suggests that intraday trading follows a u-shaped pattern (Admati & Pfleiderer, 1988). Strategic investors, it is argued, prefer to execute their trades when the market is most liquid to avoid price impact. The start and the end of day are natural hot spots for trading activity because before the start of trading, information has accumulated over time and right before the end of trading is the last chance to trade for all investors until the next trading day. As we have outlined in the previous paragraphs, strategic behavior by investors clashes for a number of reasons with the assumptions required for the PIN model to be valid. Easley et al. draw on the argument of competitiveness to uphold the generality of their model, but do not provide empirical evidence in this case. At the end of the day, no model can fully reflect realistic market conditions. Whenever possible, key assumptions require thorough empirical testing. This is what EKO97 and subsequent work by these authors tries to accomplish. We will continue on this path in chapter 3, when we present a modified estimation procedure of this microstructure trading model and empirically test the robustness of the underlying assumptions.

2.4 The context of the EKO models

2.4.1 Earlier models

David Easley and his colleagues were all but the first to model the trading process between investors and a market maker. The concept of information asymmetry instead of inventory cost lying at the core of the bid ask spread was not new. However, they substantially advanced the field of market microstructure models that deal with information asymmetry. In this section, we give an overview of related work to differentiate the foundations from the new contributions.

2.4.1.1 Glosten and Milgrom (1985)

Glosten and Milgrom (1985) were the first to examine the phenomenon of the bid-ask spread not from the usual perspective of inventory cost but instead as an adverse selection problem the market maker faces in the presence of informed and uninformed traders. In their model, inventory and other fixed or variable transaction cost are set to zero. The market maker is primarily concerned with recouping his losses to informed traders by gains in trades with liquidity traders. Formalizing this model allows deriving the key result that the spread is positively correlated with the degree of information asymmetry and negatively correlated with overall liquidity, i.e. volume of trading.

The model of Glosten and Milgrom (1985) establishes the key elements of the successional PIN model: a risk-neutral, competitive market maker acts in a competitive market and hence expects zero profit from any transaction. Orders arrive sequentially. Between each trade, the market maker can adjust the buy and sell quotes. Only one unit of the security is traded per trade. Investors approach the market maker anonymously and sequentially to decide to trade at the current quote set by the market maker. Investors are partitioned into informed and uninformed traders. Uninformed traders sell or buy with equal chance of 50% as the trade purely for liquidity reasons. Informed traders possess private information about the true value of the traded security. Competition among market makers forces the market maker to narrow his spread, while the requirement to minimize and offset losses in trading with informed traders lets him increase the spread. This basic model was the first to be able to explain the existence of a positive spread solely on the grounds of asymmetric information.

The described model is only a few steps away from the PIN models in EKOP96 and EKO97. One difference is the introduction of the parameter *alpha*. Glosten and Milgrom (1985) assume private information to be continuously available and informed traders to always be present. In the PIN model, however, an information event that affects the true value of the asset takes place only with probability *alpha*, and on days without news only uninformed traders trade. Including this probability brings this model closer to reality and circumvents a theoretical critique of the Glosten/Milgrom model: if informed traders are always present, informed traders may not want to trade at all but instead wait until their information is incorporated into prices. The second difference concerns the size of a trade, which is not considered in Glosten and Milgrom (1985), whereas Easley and O'Hara (1987) and succeeding work, at least partially, do consider the size of a trade.

2.4.1.2 Innovations in subsequent models by Easley et al.

There are three publications from Easley et al. that already build upon and extend the work from Glosten and Milgrom (1985) but still precede the final PIN models in EKO97 and EKOP96. The first one, Easley and O'Hara (1987), is the first purely theoretical paper. It evaluates the effect of the size of a trade by distinguishing small and large buy or sell orders. The modeling predicts two equilibrium outcomes. In the separating equilibrium, all large orders originate from informed traders, as they prefer to trade large quantities with their superior knowledge. Hence, a positive spread exists for large orders, as the market maker needs to be compensated and can only gain if the direction of the information changes. The spread is zero for small orders, which are always uninformed. In the pooled equilibrium, informed traders use both order types, hence the spread for both small and large orders is positive. The difference between the spread for large orders and the spread for small orders depends on the relative volume of the two order types but is always larger for the larger order type. The price mark-up of very large orders can still be observed on today's markets and Easley and O'Hara (1987) provide an explanation that goes beyond the linear relationship of order volume with the cost of inventory.

The second major innovation building on the Glosten/Milgrom model, which is also a key feature in the '97 PIN model, is introduced in Easley and O'Hara (1992): not only trades but also time contains information. In particular, the time between trades provides information. If there is new private information, informed traders will want to trade along with uninformed traders. Without new information only uninformed traders trade. Easley and O'Hara (1992) explicitly model the time between trades as another input to the market maker's decision process, which previously contained only the flow of buy and sell orders. Therefore, a trading day is divided into short intervals to be able to actually measure a "no-trade". For the market maker, the absence of trades provides information about the existence of new information. The trades itself provide information about the direction of the signal. This publication is also the first to introduce the tree structure of the trading process with the diagram that is used in most subsequent publications. The third publication before the actual PIN variable is introduced in EKOP96 extends the previously introduced tree structure of the trading process by additional branches to distinguish large and small orders (Easley, Kiefer, & O'Hara, 1997a). Further, one lag of history dependence is allowed for the uninformed traders.

The two innovations of trade size and the time between trades were both considered for the models in EKOP96 and EKO97 but only the model in EKO97 actually incorporated the time between trades. Trade size is regarded as irrelevant in the empirical analysis in both EKO97 and EKOP96. The authors justify the exclusion of this characteristic with the existence of a "pooled" equilibrium developed in Easley and O'Hara (1987) where informed traders trade in all order sizes.

2.4.2 Innovations in PIN methodology

Further work enhancing and improving the EKOP96 and EKO97 models has been pursued not only by the original authors. Enhancements by other scholars mostly share the motivation to advance the PIN method with a focus on higher convergence rates for more frequently traded stocks and more recent data as well as shorter estimation timeframes to enable day-by-day analyses[5]. David Easley, Maureen O'Hara and their co-authors, on the other hand, focus on the extension of the models to integrate different market and trading characteristics. An exception is the paper by Easley et al. (2008), which incorporates time-varying arrival rates of information into the originally static model. We will summarize notable work in the following subsections. The most current evolution, the volume-synchronized probability of informed trading (VPIN, Easley et al., 2012b), is discussed theoretically and evaluated empirically in chapter 4.

2.4.2.1 Enhancements to the PIN models by the original authors

There are five notable publications by the original authors that extend the basic microstructure trading model underlying the PIN variable in EKOP96 or EKO97. Three of those serve to accommodate broader market settings, one addresses numerical estimation problems of PIN and the final one allows for time-varying arrival rates of uninformed traders.

The first variant extends the EKOP96 model to cover two market places instead of just one to appropriately represent the large number of independent exchanges across the US (Easley, Kiefer, & O'Hara, 1996b). An extension of the final layer of the tree diagram discussed in section 2 lets a trade be executed at one or the other exchange. The authors employ the proposed variation to empirically prove a difference in the information content between two market places (NYSE and Cincinnati Exchange) and conjecture this difference to be due to cream-skimming of retail order flow by broker dealers. Adding a new level to the base tree of EKOP96 can also serve to differentiate between market and limit orders (Easley, O'Hara, & Saar, 2001). Equipped with this extended model, the authors analyze the effect of stock splits on the level of informed trading and the use of limit versus market orders. The EKO97 model has also been extended in similar fashion, that is, by fanning out the last level of the tree diagram to accommodate additional complexity. Easley, O'Hara, and Srinivas (1998a) let investors not only buy or sell a stock but also buy or sell a call or a put on the respective stock. To do so, new parameters are introduced that declare a fraction of informed or uninformed traders to choose to trade in the options instead of the security market.

[5] We will cover PIN's shortcomings and also remedies in more detail in the next chapter.

The fourth extension is of higher relevance to the active research community. The authors address the growing need to employ the PIN model in more recent, more frequent trading data (Easley, Hvidkjaer, & O'Hara, 2010). In the maximum likelihood estimation, computing efficiency suffers and truncation errors occur especially for stocks that have a large number of buys and sells. To mitigate these issues, the authors suggest a factorization of the joint likelihood function to facilitate numerical maximization, i.e. defuse the factorial terms by replacing them with a minimum/maximum construct (see Easley et al., 2010, p. 296f for details). Nevertheless, while all of the cited papers provide a significant contribution both theoretically and empirically to the finance literature, researchers relied mostly on the basic version of the EKOP96 or EKO97 model when employing PIN in their research questions, as anyone can easily verify when looking at the citation counts.

A key innovation that goes beyond changes in the last layer of the tree diagram is presented in Easley et al. (2008). The author's motivation is to allow time-varying arrival rates for informed and uninformed traders. We do not intend to repeat the derivation of the fairly complex model in the current work but instead highlight its innovations and differences compared to the setting used in EKOP96 and EKO97. The model is based on the first complete formulation of the PIN microstructure trading model in Easley and O'Hara (1992). A competitive market maker trades in continuous time with informed and uninformed traders. Orders of both types of trader follow a Poisson process with daily arrival rates. Information events occur outside of trading hours and determine the value of the asset. This model, as are EKOP96 and EKO97, is static as the realizations of information events and trade are drawn from identical and independent distributions. Now, this assumption is relaxed for the arrival rates of traders and replaced with a vector autoregressive process to allow arrival rates to depend on past observable values. The resulting specification is analogous to a GARCH equation (see Easley et al., 2008, p. 176f for details). While the arrival rate of informed traders is variable, the probability of an information event remains static, as the model formulation allows only one of them to vary. The authors argue that a varying arrival rate captures different importance of information events, which in turn attracts a varying degree of informed traders.

To retrieve parameter estimates, maximum likelihood estimation is still employed, but preceded by calculating forecasts of the parameters *alpha* and *mu*, based on the GARCH-type specification of the arrival process. Empirical results with this new model and estimation procedure clear up an empirical contradiction to the PIN model raised in a working paper by Benos and Jochec (2007). Using the EKO96 model with long estimation windows of 28 days, PIN does not increase before earnings announcements. Using the new model with shorter event windows, the hypothesized behavior of PIN rising before earnings announcements and falling afterwards is proven empirically. While the PIN model proposed by Easley et al. (2008) solves a number of criticisms of PIN, it is less widely employed than the EKOP96 or EKO97 models. The very high mathematical complexity and remaining uncertainty and issues with numerical estimation may deter researchers from applying this new version of PIN to their research questions.

2.4.2.2 Approaches to PIN estimation from other scholars

Given its widespread application in empirical finance research, other scholars propose improvements to PIN, in particular to its estimation procedure. One of the studies which kicked-off the discussion of PIN's shortcomings in more recent trading data is Yan (2009). He demonstrates that the joint market share of stocks with valid PIN estimates drops from 79% in 1983 to only 52% of the entire market in 2001 due to the steady increase in trading activity and the concurring convergence issues in the estimation process. For this particular reason, Yan (2009) propose a new way of estimating PIN without the need for maximum likelihood estimation. First, the parameters *alpha* and *delta* are identified with the use of event study methodology. An admittedly arbitrary cutoff identifies days with abnormal returns as event days. A simple average relating event days to the length of the observation period yields *alpha*. Given *alpha* and *delta*, the derivation of *epsilon* and *mu* is a straightforward transposition of the equations of the microstructure model.

Besides Yan's peculiar but innovative approach, other scholars concentrate on getting the maximum likelihood estimation to converge more often with less errors and higher accuracy, i.e. a higher chance of having found the actual optimum. This can be achieved with several approaches: First, factorization, that is, simplifying the maximum likelihood function to reduce the magnitude of numbers that have to be calculated. Second, better starting values for the model parameters before starting the maximum likelihood estimation. A simple factorization is suggested by Easley et al. (2010), where they successfully estimate PIN for a large sample of 2037 stocks. However, Lin and Ke (2011) document a bias in estimation results when relying on the procedure proposed in Easley et al. (2010). To resolve this bias, Lin and Ke (2011) draw on insights from the computer science and mathematics domain to re-organize the maximum likelihood function in a way that reduces the occasions of floating-point-exceptions. Further, they rely on a variant of the starting values proposed in a working draft of Yan and Zhang (2012).

Yan and Zhang (2012), in turn, build on the work of all three afore mentioned studies. They detect a bias in PIN estimates regardless of using the factorization from Easley et al. (2010) or the intended correction by Lin and Ke (2011). Yan and Zhang (2012) propose a new algorithm to find better starting values, which, in combination with the factorization as suggested in Lin and Ke (2011), yields precise PIN estimates.

2.5 Summary

The current chapter introduces a core foundation of this dissertation: market microstructure models. We discuss the evolution of those models that form the basis for the composite variable Probability of Informed Trading (PIN), which is one of the key variables used in empirical analysis in the following chapters.

The PIN variable was first specified by Easley et al. (1996a) and adopted, enhanced and also criticized in future work by the original authors as well as other scholars. The model consists of a competitive market maker who faces the dilemma of trading with informed and uninformed traders, whom he cannot distinguish. The only way to avert losing to the informed traders is to establish a spread and adjust his quotes after every trade. This is because the market maker cannot observer the information events that determine the value of the traded asset. Only the informed traders do. Knowing the structure of the trading process, however, allows the market maker to infer from the order flow of buys and sells the probability of an information event taken place and also the direction of the required price adjustment. (Easley et al., 1996a) and (Easley et al., 1997b) demonstrate how researchers can estimate the parameters of this underlying trade process with maximum likelihood estimation based solely on the aggregate number of buys and sells per trading day. The key parameters of the model, that is, the probability of an information event *alpha*, the probability of informed traders trading *mu* and the probability of uninformed traders *epsilon*, form the composite variable PIN.

The next chapter will provide more background on PIN and its application. Its main contribution is the proposal of a new estimation procedure that allows the daily estimation of PIN. The following chapter empirically evaluates an alternative to PIN that has been developed in parallel to the research in chapter 3. The volume-synchronized probability of informed trading (VPIN) formally builds on the same microstructure model as PIN does, but applies a number of rough estimation and approximations to arrive at a new metric that is much faster and easier to calculate and enables intraday analysis as well. The final empirical chapter taps a very new news source – Twitter – and analyzes its informational value with the microstructure tool kit developed in the previous chapters.

3 The Intraday Probability of Informed Trading[6]

3.1 Introduction

The microstructure trading model whose best known application is a composite variable named Probability of Informed Trading (PIN) has evolved over the past twenty-five years, starting with the formulation of the basic model in Easley and O'Hara (1987), the introduction of the maximum likelihood estimation for the model’s parameters and the PIN variable in Easley et al. (1996a) and an extension to account for time between trades in Easley et al. (1997b). More recent development allows for time-varying arrival rates of traders (Easley et al., 2008). Especially the earlier PIN models from Easley et al. (1996a) and Easley et al. (1997b) have become and still are a workhorse in empirical financial research. The original authors prove its relevance in asset pricing by incorporating PIN in a multifactor asset pricing model (Easley, Hvidkjaer, & O’Hara, 2002; Easley et al., 2010). Further topics have been the informativeness of trade size to market participants (Easley et al., 1997a), the effect of analyst coverage on a stock’s share of information-based trading (Easley et al., 1998b) and the use of limit orders and the effect of stock splits on information asymmetry (Easley et al., 2001). Estimations of PIN have been used as indicator of market quality to compare different trading venues and evaluate market design (changes), as Heidle and Huang (2002) do for NASDAQ vs. NYSE, Grammig, Schiereck, and Theissen (2001) for electronic vs. floor trading or Hachmeister and Schiereck (2010) for post-trade anonymity – and the list of empirical applications of PIN could easily be extended[7]. But results in current empirical literature also spell doubt on the PIN model’s validity. The PIN metric does not capture the information leakage before merger announcements (Aktas, Debodt, Declerck, & Vanoppens, 2007), it is larger for treasury bills than for equities (Akay, Cyree, Griffiths, & Winters, 2012), PIN does not increase during periods when informed traders trade (Collin-Dufresne & Fos, 2012) and whether PIN is actually relevant for asset pricing or whether it is just a substitute for liquidity is a heavily discussed proposition[8].

---

[6] An abbreviated and modified version of the results presented in this chapter has been published in Pöppe, Aitken, Schiereck, and Wiegand (2015)

[7] For example, the speed of information processing in equities trading (Visaltanachoti & Yang, 2010), seasonality in trading activity(Kang, 2010) or the impact of informed trading on manger’s sensitivity of corporate investment to stock price (Chen, Goldstein, & Jiang, 2006) or cash flow (Ascioglu, Hegde, & McDermott, 2008).

[8] Fuller, Ness, and Ness (2009) can only partly confirm Easley et al.’s results when replicating their study on NASDAQ data. See Duarte and Young (2008) and successional literature for further references to name a few prominently published studies.

The first of two major contributions of this chapter is to provide an evolution of the estimation procedure for the trade model by Easley et al. (1997b), referenced EKO97 hereafter, that allows to estimate one PIN per trading day. With this modification, major shortcomings of PIN addressed by scholars can be solved or at least significantly mitigated: the application of PIN in a short horizon event study context, the problem of convergence of the maximum likelihood estimation in recent high-frequency trading data, the assumption of at most one information event per day and the independence assumptions about information events and the arrival of orders. Mitigating these issues and especially allowing short-term event studies may also help to explain the mentioned contradictory results in empirical literature.

The approach is simple, especially in comparison to more complex variations with similar goals (Lei & Wu, 2005; Tay, Ting, Tse, & Warachka, 2009), yet results are compelling. We slice the trading day into buckets of several minutes' length, assuming independent arrival of traders and news throughout the day. Improved trade classification algorithms provide the inputs of buys, sells and no-trade intervals for the maximum likelihood estimation. Results show that the intraday PIN and its single parameters exhibit the previously demonstrated and hence expected behavior. PIN is lower for the most liquid stocks and higher for the least liquid stocks. Results are robust against the newly introduced parameter setting - the length of an intraday bucket. The necessary assumptions of independence across information events and arrival of traders prove to be far more realistic than for the original approach in EKO97. Boundary solutions for the model's parameter estimates are almost not existent. The maximum likelihood estimation following our new specification reaches convergence rates of 95% even for the most liquid stocks on trading data for the top 100 German stocks from 2005. In contrast, the datasets used in related studies with similar goals from Tay et al. (2009), Lei and Wu (2005) and all studies on PIN from Easley et al. up to 2009 reach only up to 1995, 2002[9] and 1998, respectively.

Whereas the motivation to extend and improve the EKO97 model stems from the large amount of research that relies on it, the second objective of this chapter is motivated by the non-existence of empirical evidence on the effectiveness of insider trading regulation on a granular level. Existing studies on the topic usually analyze data on a highly aggregated country level (Bhattacharya & Daouk, 2002) or stock level (Fernandes & Ferreira, 2009).

---

[9] Note that in their sample of 611 stocks, Lei and Wu (2005) pick a random sample from the decile 8th downwards that omits the top 122 most liquid stocks.

The other large area of empirical studies deal with legal, disclosed trades by potential insiders (director's dealings). The current study uses the new intraday PIN in an event-study setting to evaluate the effectiveness of single, regulatory enforced actions that try to promote a lower level of information asymmetry. An event is based on the so-called "ad-hoc announcements" stipulated in German insider trading law as the single channel for listed companies to release price-sensitive information. According to the law, this type of information has to be published without delay or kept strictly confidential. Hence, the release of these ad-hoc announcements should turn formerly speculative private information into certain public knowledge.

After we adjust for the time-scale to adhere with current market conditions and speed of information processing, the new intraday PIN reacts significantly to the disclosure of price-sensitive information. Ad-hoc announcements significantly reduce the information asymmetry on the day of disclosure. Depending on the liquidity of the stock, it takes one to two days for the PIN to reach its pre-announcement level. None of the common, simple measures of information asymmetry is able to capture this effect.

Event-study type settings are a common field of application for research using PIN to evaluate the information processing of different kinds of information flows or news announcements in capital markets. Vega (2006) uses PIN along with public news data like the Wall Street Journal and the New York Times to determine the root cause of the post-announcement drift around earnings announcements. Easley et al. (2008) analyze earnings announcements and are able to confirm that informed trading is higher before and lower after the announcement – a result that was previously questioned in a similar study by Benos and Jochec (2007). Another type of news events, conference calls, are analyzed using PIN by (Brown, Hillegeist, & Lo, 2004), and the list could easily be extended. While earnings announcements are formal announcements where the event time is known ahead by everyone, just not the exact content, the news source employed in this chapter has unknown content and unknown timing.

The remainder of this chapter is organized as follows. Section 2 details the mentioned challenges the PIN model faces and how other approaches try to make the PIN model applicable to current research questions. Section 3 covers the research design, that is, the microstructure trading model and its modification for intraday applicability, robustness checks and the event study around ad-hoc announcements. Section 4 describes the data set. Section 5 discusses results. Section 6 summarizes and concludes.

3.2 An updated estimation procedure for PIN

The PIN model has a number of known limitations and its validity, explanatory power and robustness have been thoroughly questioned. High trading intensity complicates the maximum likelihood estimation of the original model's parameters. With increased trading intensity comes increased speed of information processing, which in turn shortens the appropriate time period to be chosen for an event study type research design. Combining PIN with this approach in empirical financial research is desirable for a number of research questions but happens to produce disputable results when using long event windows (Aktas et al., 2007). Third, the assumption of at most one information event per day needs to be relaxed. The variation presented in this study allows one estimation per day and thereby overcomes all three mentioned limitations.

The incomplete convergence of the maximum likelihood estimation is a technical, but crucial issue that leaves high volume stocks without a valid PIN estimate in recent, high-frequency trading data. With exception of their most recent PIN paper using data up to 2001 (Easley et al., 2010), all previous studies on PIN from Easley et al. rely on a dataset that reaches only up to 1998 where trading frequency is so low that a few hundred trading days even from large companies such as Enron or JP Morgan need to be dropped from the sample. Aktas et al. (2007) lose about 13%-20% of their M&A cases from the years 1995 to 2000 depending on the length of the event window. In a very large sample of PIN calculations containing all ordinary common stocks listed on the NYSE and AMEX for the years 1983 to 2001 the convergence failure rate builds up in the final years reaching 3.6% of stocks accounting for 23.7% of total market capitalization in 2001 (Easley et al., 2010)[10]. Yan and Zhang (2012) claim that the share of stocks with valid PIN estimate in the mentioned dataset is only 52% of total market capitalization and wonder how to obtain valid conclusions. In contrast, our approach achieves convergence rates of 95% for the most liquid stocks on 2005 data. Another attempt to solve the convergence issue is to simply linearly scale down the input data, but this has only been tested in simulations and not real world applications and does not allow daily PIN estimates (Jackson, 2013).

[10] These PIN estimates are publicly available on Soren Hvidkjaer's website and have been employed in a number of publications (https://sites.google.com/site/hvidkjaer/data).

The lack of short-term applicability of PIN is acknowledged by the original authors themselves in reply to results that contradict the intuitive interpretation of PIN[11]: "Our belief is that this occurs because the variation in trade based on private information occurs in short periods before and after announcements and using long periods to estimate PINs obscures this effect" (Easley et al., 2008). There are (at least) two other noteworthy publications with the same motivation to develop a one-PIN-per-day model based on Easley et al. (1996a). Lei and Wu (2005) extend the continuous trading model with a Markov process that allows both informed and uninformed traders to probabilistically switch between different levels of arrival rates. This design lets them also model a certain degree of dependence between trading days instead of the independence usually assumed. However, this design limits the estimation to a distinct number of different levels for the arrival rates, whereas in the approach in this chapter, arrival rates can take on any value[12]. They further require the informed traders to know the arrival rate of the uninformed. Tay et al. (2009) merge another mathematical tool with the PIN model by using an asymmetric autoregressive conditional duration model to estimate PIN. Their data, however, reaches only up to 1995 where the maximum number of trades per day per stock is 678 – a number that even many of the "illiquid" stocks in the sample used in this study surpass. What distinguishes this study from the two aforementioned alternatives and also the re-formulation by Easley et al. (2008) besides the up-to-date trading data and the event study application is the relatively simple, albeit remarkably robust modification of the estimation procedure with which it still solves the same issues mentioned before. Another approach extends a microstructure model from 1983 whose building blocks are very similar to EKO97 (Kumar & Popescu, 2014). That is, a single market maker trades an asset, whose value depends on the arrival of news, with three instead of two groups of heterogeneously motivated and informed traders. Key difference is the modeling of the bid and ask of the market maker as a free American option on the asset, out-of-the-money for the uninformed, but eventually in-the-money for the informed traders, depending on the arrival of news. The approach is computationally feasible and allows short-term, intra-day applications. However, their results correlate with the actual PIN to only 67% to 75%,

[11] PIN was estimated to be lower before the announcement than after in a sample of earnings announcements over 13 years for stocks on the NYSE (Benos & Jochec, 2007) and in a sample of mergers and acquisitions (Aktas et al., 2007). Both studies used long event windows of 28 and 60 days, respectively.

[12] Lei and Wu (2005) present their model and parameter estimates for two levels, high and low, but note that their model is extensible to more states. We would expect the ability to estimate PIN to suffer with rising complexity when imposing a number of different states (e.g., ten) for arrival rates to accurately measure differences between days.

while the approach in this study enables to calculate the true, original PIN intraday, i.e. with almost 100% accuracy, which should clearly be preferred.

Of course there are different approaches to extract the information content from the trading process. A vector autoregressive model on the time series of trades and quotes has been proposed by Hasbrouck (1991) and subsequently been refined by several others, e.g. Dufour and Engle (2000). Nyholm (2003) develops another model to identify information-based trades and reaches conclusions that are very similar to the ones drawn by Easley et al. (1996a). Leung, Rose, and Westerholm (2013) demonstrate that signed small trade turnover, i.e. the order imbalance of small trades, also serves as proxy for uninformed trading due to its incremental predictive power in a Fama-French three-factor model where PIN is also included.

In parallel to the research of this study, the original authors of the PIN model developed a new evolution of PIN designed for application in high-frequency trading environments (Easley et al., 2012b). While the basic setting is not changed[13], there are two major methodological characteristics that distinguish the so-called volume-synchronized PIN (VPIN) and our model. We slice the trading day by time, assuming independent arrival of news throughout the day. The VPIN approach slices the day in buckets depending on a fixed amount of accumulated volume. Thereby a strong positive correlation between arrival of (private) information and volume is assumed, a proposition where literature has conflicting views (Dufour & Engle, 2000; Wuensche, Theissen, & Grammig, 2011). The second difference is that the VPIN input parameters of buys and sells are calculated based on a heuristic bulk classification that replaces the common classification algorithm and its derivatives. We find that on our recent data from XETRA, the computer-based continuous trading system from Deutsche Boerse in Germany, most, if not all of the issues with trade classification that stem from delayed or incomplete data are resolved and hence prefer the precise classification over a heuristic one. Apart from technical aspects, VPIN has so far not been analyzed regarding its applicability on equities, as it has only been applied to ultra-high-frequency trading future contracts, like the E-mini S&P 500 future or WTI crude oil future (Easley et al., 2012b). We will return to VPIN in chapter 4 where we extensively validate VPINs robustness and consistency.

[13] Market makers are still at the core of the model, in the new version they are represented by high frequency trading firms that provide liquidity by placing passive orders in the electronic order book. They face the risk of adverse selection and are assumed to observe the order flow to try to infer existence and direction of information.

3.3 Research design

This section develops a new estimation procedure of the sequential, discrete-time microstructure trading model formulated by Easley et al. (1997b), with adaptions for the time-scale[14]. The original authors assumed new information events to arrive at maximum once per day. Instead, we allow for intraday arrival of information. This is accomplished by splitting the trading day into “buckets” of several minutes’ length. Recent literature indicates that 15 minutes or less should be long enough for market participants to incorporate new information into prices and short enough to accommodate the increasing frequency of both information releases and trading itself[15]. Patell and Wolfson (1984) find in their sample as old as the late 1970s that the trading profits after earnings and dividend announcements disappear within the first five to fifteen minutes after disclosure. Similar conclusion have been drawn for the announcement of seasoned equity offerings (Barclay & Litzenberger, 1988), initial analyst coverage on NASDAQ and NYSE/AMEX (Kim, Lin, & Slovin, 1997), analyst recommendations (Green, 2006) or stock market news on TV (Busse & Green, 2002), to name a few. An analysis of the intraday stock price reaction on XETRA to ad-hoc announcements, the market and the news source used in this chapter, found that the price reaction was completed within minutes (<15) or a few trades (<10) after the release of the announcement (Muntermann & Guettler, 2007). To show that the arbitrary choice of the length of the bucket does not affect our results, we will later present results for 8, 12 and 15 minutes length.

Whereas in chapter 2, a signal ψ about about the value of asset V arrives each trading day, now a signal may occur for every bucket within a trading day. Alternatively, one could talk of trading sessions being modeled, where a session could refer to a full day or just a part of day. The derivation of the model is identical to the one presented in chapter 2 and therefore restated only partially in this chapter. To differentiate the daily version from the intraday version we index the buckets by $k = 1 \dots K$ in contrast to the days being indexed by t$=$ $1 \dots T$.

---

[14] Given the variety of prominently published models by Easley et al. we want to explicitly state again that we build on the version of the trading model from Easley et al. (1997b) that uses discrete time and do *not* build on the trading model from Easley et al. (1996a) that uses continuous time and ignores the time between trades as source of information..

[15] The arbitrary choice of the length of the trading session is noted by the original authors themselves: "Certainly, our specification of a day is arbitrary. In active markets, prices could adjust to new information in minutes and new information events could occur quite frequently"(Easley & O'Hara, 1992).

The signals are assumed to arrive independently at the start of every trading session (a full day or intraday buckets) with probability α. A signal conveys bad news with probability δ and good news with probability $1-\delta$, denoted by $\psi_k = L$ and $\psi_k = H$, respectively. Conditional on a signal occurring, informed traders always want to trade in the direction of the signal and will be able to do so with probability μ. Uninformed traders will be given the chance to trade with probability $1-\mu$ and they decide to trade with probability ε. If there is no signal, only uninformed traders decide to trade with probability ε. Neither of the two parties may decide to trade within a certain amount of time, which is valuable information for the market maker as well. The trading session is therefore divided into a number of intervals long enough to accommodate one trade (or count as no-trade interval).We will later on present results for trading intervals of 5 to 30 seconds length, depending on the trading activity, and also show that results are independent of the actual choice as long as the trade interval stays within reasonable bounds.

The conditional probabilities from the perspective of the market maker are formulated identical to chapter 2. Buys, sells and no-trades are the key input to the market maker's thought process of which the quoted spread and in turn trades are the outcomes. Multiplying the conditioned probabilities to observe a combination of buys, sells and no-trades with the probability that the conditioned event is occurring gives the unconditional probability. Summing these three products yields the likelihood function for a single bucket, conditional on the parameters $\alpha, \delta, \varepsilon$ and μ:

$$\begin{aligned} &\mathrm{P}\{\mathrm{B}, \mathrm{S}, \mathrm{N} | \alpha, \delta, \varepsilon, \mu\} = \alpha(1-\delta)\, \mathrm{P}\{B, S, N | \psi = H\} + \alpha\delta\ \mathrm{P}\{B, S, N | \psi = L\} + \\ &(1-\alpha)\, \mathrm{P}\{B, S, N | \psi = 0\} \end{aligned} \tag{3-1}$$

As shown earlier, the maximum likelihood function of a whole day forms, as result of a log transformation, as sum over the buckets per day:

$$\begin{aligned} &\sum_{k=1}^{K} log\left[\alpha(1-\delta)\left(1+\frac{\mu}{x}\right)^{B} + \alpha\delta\left(1+\frac{\mu}{x}\right)^{S} + (1-\alpha)\left(\frac{1}{1-\mu}\right)^{S+B+N}\right] \\ &\qquad + \sum_{k=1}^{K} log\left[\left((1-\mu)(1-\varepsilon)\right)^{N} x^{S+B}\right] \end{aligned} \tag{3-2}$$

where $x = (1-\mu)\frac{1}{2}\varepsilon$.

The probability of informed trading was defined as "the probability that any trade that occurs at time t is information-based" (Easley et. al. 1996). The PIN measures the share of informed trades relative to all trades. We use the PIN97 formula derived in chapter 2 to be consistent with the model developed in Easley et al. (1997b), albeit results with the PIN96 formula are very similar.

$$PIN97 = \frac{\alpha\mu}{\alpha\mu + \varepsilon(1 - \alpha\mu)} \tag{3-3}$$

For the estimation of the maximum likelihood function the four parameters were restricted to (0,1) by a logit transformation. Best convergence results were achieved with a combination of the four algorithms Newton-Raphson, Berndt-Hall-Hall-Hausman, Davidon-Fletcher-Powell and Broyden-Fletcher-Goldfarb-Shanno were each algorithm is run for five iterations before the next algorithms continues.

Practically, the model presented in this chapter allows us to estimate one PIN per day and use this PIN in an event study type setting. The subsequent section will demonstrate that the underlying assumptions of the model hold as least as good as they do in the original formulation (see Section 5.3).

3.3.1 Robustness tests

The presented model relies on a number of assumptions that need explanation and, where possible, validation via robustness checks. Further, some arbitrary choices in parameter settings need to be verified.

For each stock index the trading activity determines feasible boundaries within which arbitrary variations of the parameter trade interval should not affect the results. The second model parameter externally specified by research design is the bucket size. The right bucket size depends on the arrival rate of information and the time required to process these, both of which are not directly observable from our trading data. Hence, we generate results for trade intervals of 5, 10, 20 and 30 seconds and combine them with bucket sizes of 8, 12 and 15 minute length, resulting in 12 possible combinations to compare the analysis for.

The maximum likelihood estimation of the model may fail to converge or produce boundary solutions. Both issues have been identified and analyzed in the literature, the former by William Lin and Ke (2011) and the latter by Yan and Zhang (2012). We therefore report convergence rates and check all model parameters for their distribution and occurrence of boundary solutions.

The microstructure model builds on two key assumptions: the independent arrival of information events across intraday buckets and the independent arrival of traders throughout the day. We adapt the independence test run in EKO97 to verify these assumptions. Their validity in the original model had been illustrated at the example of one stock in EKO97 and is implicitly presumed by all research building on EKO97 or Easley et. al. 1996.

The independence of news events is tested with a runs test for each stock and day. Information events as specified by the model are not directly observable. But the informed trade only on one side of the market and the uninformed are equally likely to buy or sell. The different probabilistic structure of the arrival of trades in buckets with and without news makes it possible to classify buckets based on trading activity. The estimated parameter α indicates how many buckets of a given day process new information. Accordingly, buckets with a trade count equal or below the $(1-\alpha)$-percentile of the trade count distribution on that day are classified as no-news buckets and buckets with a trade count above this threshold are news buckets. In equivalent manner we test the independence of good and bad news by classifying news buckets as good news if there are more buys than sells and as bad news otherwise. The original interday PIN, i.e. bucket length equals one trading day, is estimated over non-overlapping windows of 33, 43, and 65 business days as a benchmark (replicating the number of buckets per trading day on XETRA for 15, 12 and 8 minutes intraday bucket length).

The second crucial assumption is about single trades being independent, conditional on a known signal. Informed traders have no reason to observe the trading history as they have superior information anyway. Uninformed traders know that the quotes set by the market maker are the expected values of the asset conditional on the information revealed by the trading process. Hence there is also no rational reason for the uninformed traders to let their trades be guided by the trading history because the market maker already incorporates all information to be extracted from the trading process into his prices[16]. Nevertheless, to test the dependence of trades, we calculate the autocorrelations of buys, sells and no-trades per bucket and day for six lags and a lagged regression of net trades. We compare our model's assumptions by running a regression per stock for the daily net trades and a regression per day and stock over the net trades of the intraday buckets.

There are number of other proxies for information asymmetry suggested in the literature, we compare two of the most popular to the PIN calculated daily. These are the time-

---

[16] In a slightly different version of the current model, the authors allow for one lag of history dependence of uninformed trades (Easley et al., 1997a).

weighted relative percentage spread (TWRPS) and the relative order imbalance (OIB). The TWRPS is known to be affected by both liquidity and information asymmetry, amongst other factors (Huang & Stoll, 1997). The OIB was suggested as easy-to-calculate approximation of the PIN (Aktas et al., 2007) and is defined as the absolute order imbalance relative to the total number of trades. TWRPS and OIB are calculated once per day. Both metrics are independent of the parameters bucket size or trade interval required for the daily PIN.

We find three more arguments on constraints of the presented model worth a brief discussion. These are the assumption of a single market maker, the disregard of trade size and the accumulation of information overnight.

Although originally established from the perspective of a single, competitive market maker in EKO97 and Easley et al. (1996a), the model has been applied to automated, order-driven markets where the collectivity of liquidity providers or “portfolio managers” (Easley et al., 2008) or high frequency trading firms that provide liquidity by placing passive orders in the electronic order book (Easley et al., 2012b) substitute the market maker. Hence, we also do not consider this as an issue for the model presented here.

Another restriction is that the size of a trade is not considered. Earlier studies on this topic explicitly model trade size and find it a relevant factor in determining the information content of trades, however these results were based on a fairly small sample of six stocks (Easley et al., 1997a) or purely theoretical (Easley & O'Hara, 1987). EKO97 tests and concludes that trade size provides no additional information content beyond the trade itself. This finding is also confirmed in more recent work on a very large sample of stocks (Easley et al., 2008). In the continuous version of this microstructure model Easley et al. (1996a) also do not consider trade size. We rely on this evidence that at least a large share of the information content is captured with the trade alone and the additional information of size, if any, can be neglected.

A more technical issue is that given the long overnight pause of trading the chance that a signal occurs that influences the first trading session of the day could be higher than for the other intraday buckets. Similarly, the signal could also be stronger given the longer time over which information can accumulate. The opening auction in the markets we study is designed to absorb this surplus liquidity and provide a balanced starting point for the following trade process. We assume that the opening auction fulfills its purpose. Practically the trades in the opening auction are ignored in our estimations.

3.3.2 Event study with ad-hoc announcements

We exemplify the application of the proposed new PIN estimation procedure in an event study design by testing the effectiveness of disclosure requirements in German insider trading legislation. To perform any type of event study, three questions need to be answered: What is the event, what is the event window and how are changes in the observed metrics evaluated?

An event is based on the so-called “ad-hoc announcements” stipulated in German insider trading law as the single channel for listed companies to release price-sensitive information. Section 15 of the securities trading act [17] specifies that price-sensitive information has to be published without delay or kept strictly confidential until a publication does not harm the company’s interests. Given the information released by the ad-hoc announcements is price-sensitive, we assume that a fraction of traders, e.g. “informed” institutional investors, anticipate or speculate on the upcoming news prior to its disclosure. Therefore, we hypothesize the public release of this information to lower information asymmetry in the market. In other words, the release of these announcements turns formerly speculative private information into certain public knowledge. Accordingly, we expect a lower PIN upon the day of disclosure.

In the parameters of our model, we expect fewer informed traders, as their previously private information turned public. The uninformed traders may respond to the public news by adjusting their portfolio composition or simply trade due to the attention brought to the stock, in the notion of Barber and Odean (2008). The parameter δ indicates whether the signal is positive or negative. Without further semantic analysis of the content of the ad-hoc announcements, we cannot expect this parameter to move significantly in any direction, thereby assuming equal probability of good and bad news. The parameter α is more ambiguous. Given that it represents the arrival of new information, one could argue that α should go up. In the model, however, α relates to the arrival of information that is *not* public, whereas the events we are studying are arrivals of public information. Hence, α is expected to decrease on the event day, if there is any reaction at all.

We define a symmetric event window around the date of the announcement to evaluate these hypotheses. Given the evidence on quick information processing noted earlier, a post-announcement window of two days should be long enough to capture the

[17] http://www.bafin.de/SharedDocs/Aufsichtsrecht/EN/Gesetz/wphg_101119_en.html

hypothesized drop in the PIN. In line with common event study methodology, we keep the pre-announcement window at equal length of two days.

After having identified all event days, the PIN for each security/day combination in the event window is computed. Subsequently, the average PIN and median PIN are calculated for each event day. The behavior of PIN, its parameters, and the related information proxies within the event window are then examined with several non-parametric tests: First, we run the Kruskal-Wallis-Test to check whether the distributions of PIN and its parameters for the five event window days differ significantly from one another. Second and third, the Mann-Whitney test and the Mood's Median test are used to test for pairwise differences in PIN on all successive event days, e.g. day -1 to 0.

The general relevance of Ad-hoc announcements to the market has been confirmed by Muntermann and Guettler (2007). We validate the relevance of these announcements to the market in our sample by testing the differences in trading volume within the event window.

3.4 Data and descriptive statistics

This section explains what trading data from which venue we use, how we classify trades into buys and sells from the raw trading data and it gives a brief background on the ad-hoc announcements that we study later on.

3.4.1 Trading data

We obtain trading data for XETRA from the Securities Industry Research Centre of Asia-Pacific (SIRCA). XETRA is the electronic trading system of the Frankfurt Stock Exchange (FSE), which is operated by Deutsche Boerse, and captures the majority of equity trading volume (>90%) in Germany, hence omitting FSE floor trading (shut down in 2011 anyway) and the eight regional exchanges cannot be expected to bias our results.

We choose the year 2005 for our analysis as it is the most recent year likely to be only negligibly affected by either crisis or excessive boom. We limit the sample to the stocks trading in one of the top 3 stock indices which are DAX, MDAX and SDAX. The DAX is Germany`s major equity index, consisting of the 30 largest and most actively traded German equities. The MDAX (SDAX) comprise the next 50 stocks from traditional sectors that rank below the DAX (MDAX) stocks in terms of market capitalization and trading volume. Although our sample covers "only" 102 stocks, these stand for 85% of the total traded value on XETRA.

The data comprises time stamped trades with price, volume and quotes stating the best bid and best ask price of limit orders. Quotes released as indicative prices during auction phases are ignored as well as trades settled within an auction (opening, intraday, closing, volatility interruption). The index constitution was acquired from Thomson Reuters Tick History. The major indices in Germany are updated twice a year. For our analysis summarized by indices, a stock is grouped into the highest index to which it ever belonged during the observation period.

The trading activity of this sample is comparable in magnitude to the trading activity in the US. In the year 2005 on NYSE the traded volume of the 30 stocks in the Dow Jones was about 100 billion shares. The 30 stocks in the major index DAX at Deutsche Boerse during the same year accounted for 25 billion shares traded. Table 3-1 gives an overview of the trading activity of the selected stocks on XETRA aggregated by index membership. The three measures of trading activity presented in this table, which are trade count, traded value, traded volume, in totals and mean and maximum per group, show a clear separation between the stock indices. Trading activity grows exponentially from DAX to SDAX. While we did not include a separate filter for illiquid stocks, it is clear from this data that estimates for the low-liquid stocks in SDAX will be less meaningful given that stocks happen to trade only a few times per day.

The last two columns give a good indication for the trade interval parameter. The mean time between two consecutive trades is calculated by dividing the mean und maximum trade count by the seconds of the trading day. So even with maximum trading intensity in the DAX, there is on average 1 second between two trades and during mean trading intensity there are on average 14 seconds between two trades. Given the non-uniform distribution of trades throughout the trading day, for stocks in the DAX, a 5 or 10 seconds long interval should fit the trading activity. For the MDAX 10 to 20 seconds are certainly well enough to accommodate the average trading activity, given that the trade interval setting should rather be too short than too long. For stocks in the SDAX we expect the 30 seconds interval to produce the most reliable results.

Table 3-1: Trading activity on XETRA in 2005

This table shows the aggregated trading activity of all stocks in the sample, aggregated by index membership. Columns two to four show the totals of the number of trades, the traded value and traded volume aggregated over the year 2005. The first column of the sections daily trade count and daily trade value is the mean of the daily trade value or trade count. "Max" is the maximum of all daily values. The last two columns estimate the mean time in seconds between trades by dividing the mean and maximum number of trades by the number of seconds per trading day (30600 seconds).

	Totals 2005			**Daily Trade count per Stock**		**Daily Traded Value per Stock** (x 1,000 €)		**Mean time btw. trades** (in seconds)	
	Trade Count (x 1,000)	**Value** (Mio. €)	**Volume** (Mio.)	**Mean**	**Max**	**Mean**	**Max**	**Mean trade count**	**Max trade count**
DAX	15,877	891,388	25,188	2,227	25,326	124,445	2,418,170	14	1
MDAX	2,926	56,118	2,135	305	11,788	6	592,757	100	3
SDAX	556	4,668	467	63	2,174	528	35,748	488	14
Total	19,359	952,175	27,790	750	25,326	37	2,418,170	41	1

3.4.2 Buy/sell classification

The model estimation requires trades to be flagged as buyer- or seller-initiated, information that we do not have in the trading data. In an electronic limit order book theoretically all trades should be executed at the bid or ask, but this is not the case in our raw data for about 13% of all trades (table A.1). We overcome reporting delays of trades or quotes and the induced random time lag between quotes and corresponding trades with three pre-processing heuristics that leave only 4.2% of all trades not at bid or ask and 2.2% inside the spread. Firstly, all trades occurring within 5-milliseconds at the same price level, i.e. without intervening quote revisions, are aggregated. These trades are very likely to originate from one order matching several smaller limit orders at the same price level. Secondly, large orders clearing several layers of the limit order book (run-throughs) are aggregated into one trade. The median delay between trades that have been identified as a sequence of run-throughs is 2.3 milliseconds, which is well below our maximum threshold set at 50 milliseconds. These two steps eliminate 10% of all trades in the original data. In the third step, our trade-quote-match rule (TQM) overcomes the reporting delay where quotes are reported ahead of trades by trying to match trades in price and volume with the next previous quote a maximum of three seconds ahead of the reported trading time. The number of trades not at bid or ask is reduced by 8% from 4.5% to 4.2%. The median delay of late reported trades identified by the TQM rule averaged over all stocks is 304 milliseconds, well below the three seconds maximum threshold.

After this procedure we replicate the algorithm proposed by Ellis, Michaely, and O'Hara (2000), EMO, itself being a variation of the well-known Lee-Ready algorithm (Lee & Ready, 1991b). That is, the remaining trades at the ask are classified as buys, trades at the bid as sells and all other trades are classified by the (forward) tick test. On a sample of three month 2006 trading data from the Australian Stock Exchange that included flags indicating buyer or seller initiated trades the implementation of the EMO algorithm combined with our pre-processing heuristics achieved an accuracy of 97.1% (see table A.2). Any misclassification bias in our PIN calculation as analyzed by Boehmer, Grammig, and Theissen (2007), can hence be excluded.

After having identified "true" buys and sells, the no-trade intervals are counted as the number of non-overlapping 5-second-intervals (10, 20, 30 seconds respectively) without any trade. A trading day is finally sliced into buckets of 8, 12 or 15 minutes to prepare the data for the maximum likelihood estimation.

3.4.3 Ad-hoc announcements

The news source for the event study part of this chapter are the so-called "ad-hoc announcements" stipulated in German insider trading law[18]. Companies usually delegate the ad-hoc disclosure obligations to one of several news agencies, which arrange the timely transmission of the news to the supervisory authority Bafin, to all market operators where the security or its derivatives are trading and, with a delay of several minutes, to the public via appropriate media channels. The current study's news source are all ad-hoc announcements published by the "Deutsche Gesellschaft für Ad-Hoc Publizität" (DGAP) through its service in 2005 for the selected stocks. DGAP is by far the largest ad-hoc service provider in Germany. We collected the ad-hoc news for five DAX companies that were not included in DGAP's sample manually from the company websites to have a complete sample of the major index[19], but did not repeat this procedure for MDAX and SDAX stocks where the coverage of DGAP's sample reaches already 39 of 50 and 35 of 50 stocks respectively. Duplicate announcements published in a second language are eliminated and announcements within the same day (market close to market close) merged into one, leaving a total of 471 announcements for 102 distinct stocks in 2005.

[18] See section 15 of the securities trading act for details, available on the website of the supervisory authority Bafin: http://www.bafin.de/SharedDocs/Aufsichtsrecht/EN/Gesetz/wphg_101119_en.html

[19] Except for two companies: Schering's announcements were not available anymore after its merger with Bayer; Henkel did not publish any ad-hoc announcements in 2005.

Announcements published out of trading hours are shifted right before market opening of the next trading session so that the event day is always the day where the announcement affects trading. If between any two announcements were less than four days, both announcements were dropped from the sample to exclude interferences between different events. From the 471 unique announcements, 72 are dropped due to conflicting event windows.

3.5 Empirical results

The following paragraphs discuss the empirical results, which we group in three sections. First, we present results of the maximum likelihood estimation of the model parameters to see how the parameters behave in general. Second, we analyze how the intraday PIN behaves around the release of ad-hoc announcements. Finally we show results of the series of robustness tests outlined in section 3. We present results for the cross section of all stocks in each of the three indices DAX, MDAX and SDAX.

3.5.1 Maximum likelihood estimation and robustness of parameters

For each stock in the sample a maximum likelihood estimation is run for each trading day. Results are then aggregated over all trading days and all stocks in the respective index[20]. Table 3-2 presents mean, median and standard deviation for the model's parameters and the PIN variable per index for different trade intervals and a constant bucket size of 12 minutes. Histograms of the estimated model parameters are presented in figure 3-1, figure 3-2, and figure 3-3 for DAX, MDAX and SDAX respectively. With the exception of α, the bucket size had almost no effect on the parameter values (an extended version of table 3-2 can be found in table A.3).

[20] We also split all stocks into deciles based on trading activity for a more granular cross-sectional breakdown. The trends and results described in the analysis using indices grouping are confirmed in the decile split. Results are available on request, but we leave them out here for brevity.

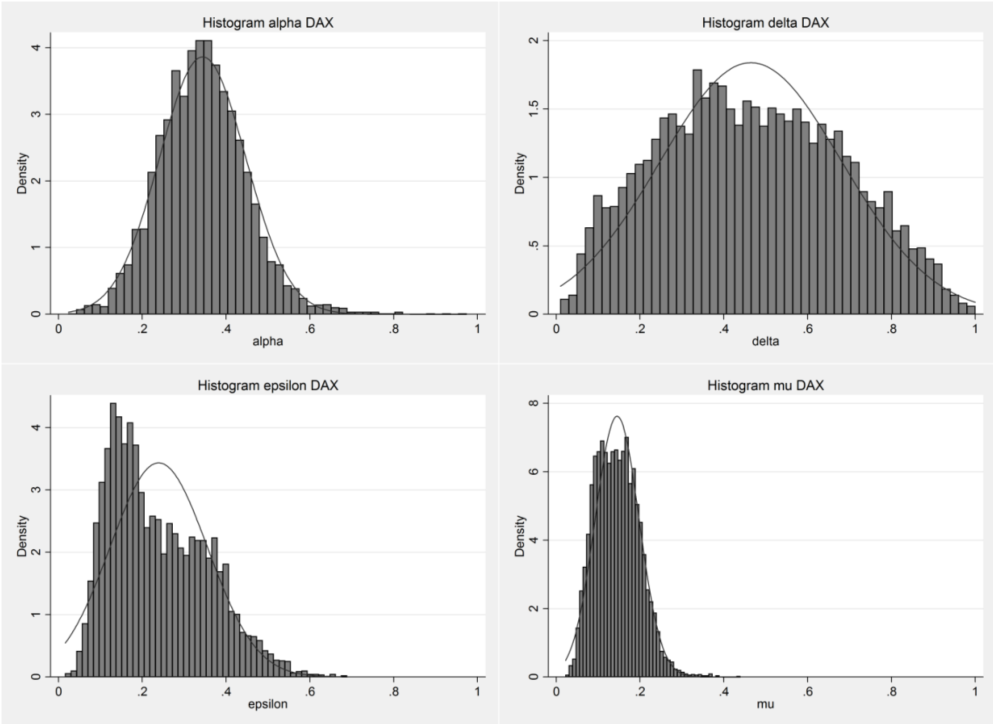

Figure 3-1: Histogram of alpha, delta, epsilon, mu for DAX stocks

This figure shows histograms of the estimates for the parameters alpha, delta, epsilon, mu for all stocks in the DAX and all trading days in 2005. Bucket size is 12 minutes, trade interval is 5 seconds.

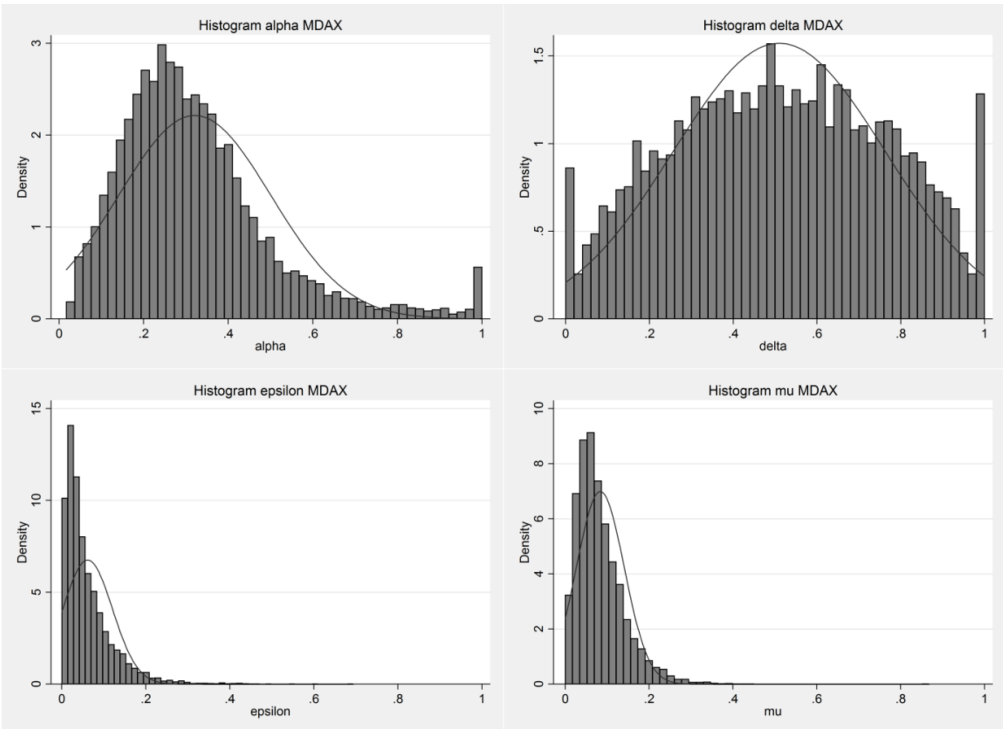

Figure 3-2: Histogram of alpha, delta, epsilon, mu for MDAX stocks

This figure shows histograms of the estimates for the parameters alpha, delta, epsilon, mu for all stocks in the MDAX and all trading days in 2005. Bucket size is 12 minutes, trade interval is 10 seconds.

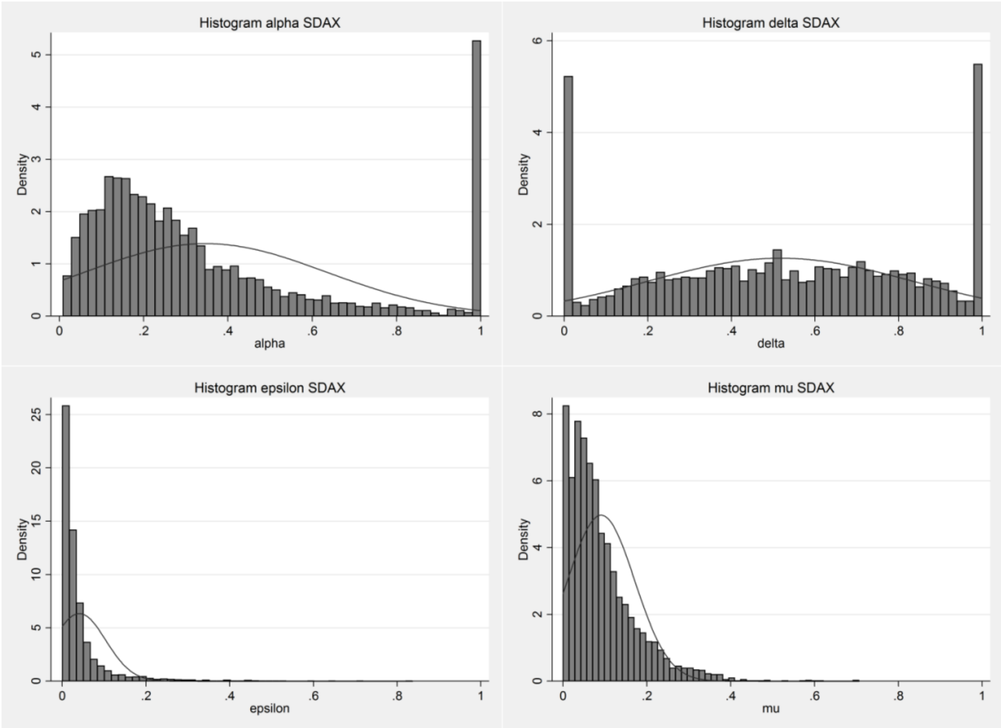

Figure 3-3: Histogram of alpha, delta, epsilon, mu for SDAX stocks

This figure shows histograms of the estimates for the parameters alpha, delta, epsilon, mu for all stocks in the SDAX and all trading days in 2005. Bucket size is 12 minutes, trade interval is 30 seconds.

Let us first look at the arrival probabilities ε and μ. The parameter μ is the fraction of observations, including no trades, made by informed traders and ε is the probability that an uninformed trader trades. Both ε and μ vary heavily with the length of the trade interval, which is expected. Each trade interval is an observation with a trade or no-trade as outcome. Hence doubling the length of the trade interval should roughly double the probability to observe a trade in any interval. This is confirmed in the results. The relation holds perfectly for MDAX and SDAX where medians of ε and μ double from 5 to 10 to 20 seconds, and both ε and μ for 30 seconds are three times their values in the 10 second interval. In the DAX, however, 20 seconds of trade interval length is simply too large as the nonlinear behavior of ε and μ confirms and as we argued earlier based on the raw trading data. The large differences in liquidity also become apparent in the absolute values of ε and μ which are about ten and three times larger in DAX than in MDAX, respectively; similar for MDAX to SDAX.

Table 3-2: PIN summary by indices

This table presents means, medians and standard deviations of parameter estimates by index membership for different trade intervals (TI, in seconds) and a constant bucket length of 12 minutes. PIN is the probability of informed trading, a composite variable of epsilon, mu and alpha. Epsilon is the probability an uninformed trader trades, mu is the probability of an informed trade, alpha is the probability of an information event and delta the probability information events are bad news. Columns 2 shows the number of observations, i.e. number of stocks times number of trading days and columns three is the share of observations where the maximum likelihood estimation converged.

			PIN			**epsilon**			**mu**			**alpha**			**delta**		
TI	**obs**	**conv**	mean	med.	sd.	mean	med.	sd.	mean	med.	sd.	mean	med.	sd.	mean	med.	sd.
DAX																	
5	7196	95%	.185	.180	.047	.239	.216	.116	.145	.142	.052	.344	.341	.103	.465	.456	.217
10	7196	95%	.174	.168	.050	.415	.392	.170	.214	.214	.054	.360	.355	.104	.461	.451	.221
20	7196	93%	.158	.151	.053	.635	.637	.187	.267	.264	.055	.388	.381	.117	.454	.444	.231
30	7196	91%	.151	.144	.053	.753	.783	.165	.289	.284	.065	.407	.398	.133	.444	.424	.237
MDAX																	
5	10023	87%	.327	.310	.122	.031	.021	.031	.044	.036	.034	.321	.287	.181	.512	.509	.254
10	10023	87%	.324	.307	.121	.061	.043	.059	.084	.071	.057	.321	.288	.180	.510	.507	.254
20	10023	87%	.321	.303	.123	.118	.085	.105	.148	.136	.084	.326	.296	.177	.511	.510	.256
30	10023	87%	.315	.298	.123	.170	.127	.141	.198	.191	.100	.333	.303	.181	.510	.510	.260
SDAX																	
5	8995	61%	.406	.389	.181	.007	.004	.016	.017	.012	.018	.334	.244	.276	.525	.523	.305
10	8995	61%	.406	.388	.182	.014	.007	.025	.033	.024	.035	.332	.243	.274	.520	.519	.307
20	8995	61%	.405	.386	.182	.028	.014	.045	.064	.048	.060	.332	.244	.273	.526	.524	.305
30	8995	63%	.398	.375	.181	.041	.021	.063	.091	.070	.080	.347	.250	.287	.517	.516	.316

An effect attributable to the design of the model is the difference in the standard deviation of ε and μ compared to the standard deviation of α and δ. The lower precision of the α and δ estimates is attributable to “the difference in information accumulation rates” (Easley et al., 1997b). Intuitively, if you think of the tree diagram, the part of the tree defined by α and δ is run through exactly once every trading session. The part of the tree defined by ε and μ is run through on every trade. The confirmation of this behavior in the parameter estimates is another indication that the new approach chosen in this study is comparable with the original model formulation.

While ε and μ both increase heavily with liquidity from SDAX to DAX, relative to all trades the uninformed traders account for an increasing larger share of the trading activity. This is exactly what the composite variable PIN captures. PIN is strongly decreasing from low to high volume stocks, starting with a median of 39% for SDAX stocks, about 31% for MDAX with and decreasing to a mean of around 17% for the DAX stocks. That informed trading or insider trading is less likely to happen in the very large and very liquid stocks relative to the less liquid stocks is a result also found in earlier PIN studies (Easley et al., 1996) and related literature. Although in absolute values informed trading alone, measured by the parameter μ, is much larger in DAX than in MDAX than in SDAX, this is more than offset by the higher arrival probabilities of uninformed traders. The mass of informed traders in the DAX can effectively hide in the liquidity provided by the uninformed. It is clear to see in table 3-2 how the mean and median PIN is stable across different trade interval (and bucket) settings. This gives assurance that further analysis, especially the event study, should not be affected by the exact choices of bucket size and length of trade interval. Only in the DAX is PIN slightly decreasing with lower values for the trade interval parameter, because the long intervals are not applicable given the high trading frequency.

Figure 3-4 shows a histogram of the PIN for each index and a constant bucket and trade interval. On first sight, the histograms appear to roughly resemble a normal distribution. At single stock level a few stocks per index “pass” the normality test, i.e. the null hypothesis is not rejected, but for the whole sample all applicable tests reject the null hypothesis, hence we will use only non-parametric tests. Unsurprisingly when looking at the histograms, the differences in the means of PIN between indices are highly significant when applying a Kruskal-Wallis test or a Wilcoxon-Mann-Whitney test on successive indices (results omitted for brevity).

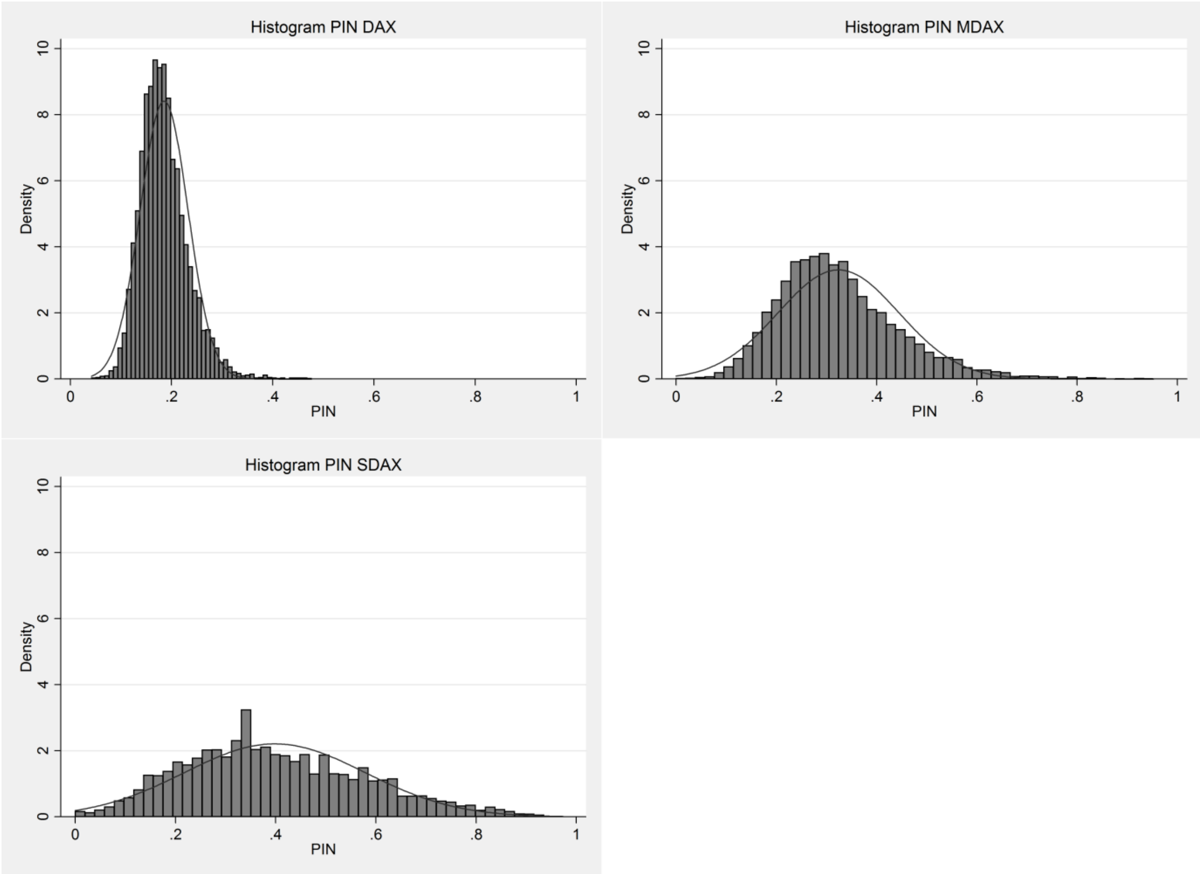

Figure 3-4: Histogram of PIN for DAX, MDAX, and SDAX

This figure shows histograms of the PIN estimates per stock and day, for DAX, MDAX and SDAX (clockwise starting top left). Bucket size is 12 minutes, trade interval is 5 seconds for DAX, 10 seconds for MDAX and 30 seconds for the SDAX.

The median of α ranges from 0.25 in SDAX to 0.29 in MDAX and about 0.35 in the DAX. A higher frequency of information arrivals in the high volume stocks is reasonable, given the higher attention these companies usually attract. The absolute values of α say that each company is affected by on average 10 to 15 different information events per day. This seems reasonable given the research building on Grossman and Stiglitz (1980), where private research constantly and randomly "creates" private information about a company. The parameter α is meant to also capture the creation of private information, not only public information events. Any announcement of certain relevance is usually accompanied by subsequently released pieces of information such as analyst's or competitor's comments or background reports and also preceding information events like rumors and speculative forecasts. Additional information events can also arise from the cross-sectional influence of information.

There is no reason for the probability of a news event, α, to vary with the length of the trade interval, which can be verified in the results. A larger bucket size, however, should increase α as the information events are distributed among fewer buckets. Figure 3-5 plots the median of α against the bucket size for DAX, MDAX and SDAX. The trend of α increasing with the bucket size is visible, although the absolute differences especially in the DAX are not linear in bucket size. A Kruskal-Wallis test within each index on the values of α between buckets is significant at the 1% level for all indices, confirming also statistically the expected behavior of this parameter.

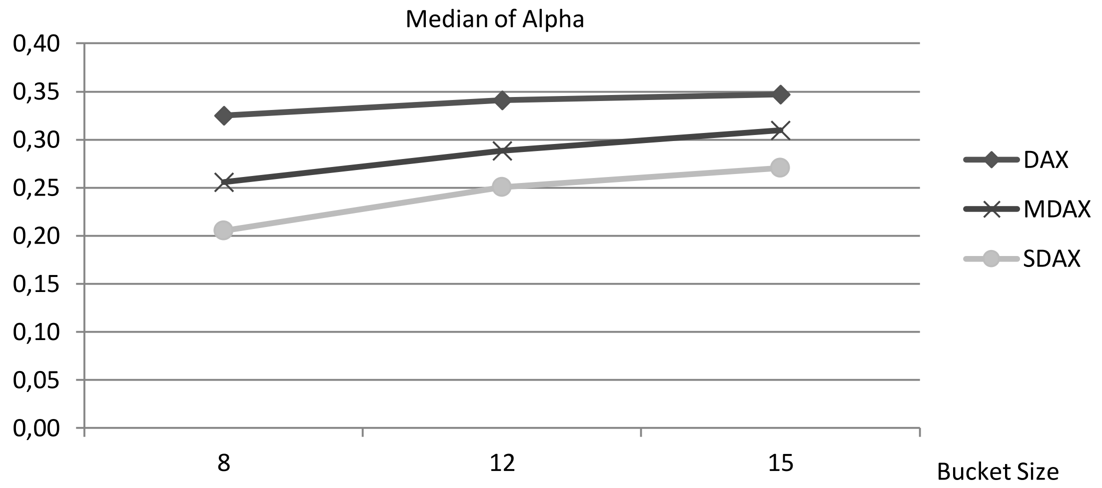

Figure 3-5: Alpha on bucket size

This figure plots the median of alpha, the probability of a news event, per index against the three different bucket sizes 8, 12 and 15 minutes. The trade interval is 5 seconds for DAX, 10 seconds for MDAX and 30 seconds for the SDAX.

The chance of good or bad news, the parameter δ, is expected to range around 50%. This is not exactly true in our results. While δ is not affected by either bucket size or trade interval, for the highly liquid DAX stocks δ is slightly below 0.5 and for the medium and low liquidity stocks in MDAX and SDAX δ is slightly above 0.5. Given that the parameter δ is not part of the PIN variable, this deviation cannot affect any of the following results and hence we did not investigate this issue further.

Convergence rates are very high for the very liquid stocks in the DAX. A little surprising at first sight, convergence rates decrease with decreasing trading frequency. Given the lower convergence rate of only around 61% in the SDAX, convergence of the maximum likelihood function in our setting is rather a problem for very illiquid stocks or illiquid days. In addition, some stocks in the SDAX simply do not trade at all on some days. What is important for the applicability in further empirical research is that the very liquid DAX stocks converge in up to 95% of the cases with the appropriate trade interval setting.

While the results of the estimations and their convergence rates sound plausible in the aggregate, we want to further analyze the preciseness of every single maximum likelihood estimation. We calculate the share of estimates that are significantly different from zero at the 1% and 5% level in table 3-3. For ε and μ, if the maximum likelihood estimation converges, basically all estimates are significantly different from zero for all stocks in all indices and estimation settings. This supports the validity of our estimation procedure itself and also of the PIN variable, where ε and μ are the major influential variables. Moreover, it also supports our interpretation from the previous paragraph that a minimum degree of trading activity is required for the estimation to converge. The interpretation of these values is different for α (and δ). The share of significant α increases with shorter bucket length, which is an expected but rather technical behavior given that more buckets per trading sessions mean more observations, hence increasing the likelihood of the parameter being statistically significant. It is not plausible for α do be different from zero all the time, as there can well be trading sessions with no or little new information and mainly uninformed trading. Even less significance is expected for the parameter δ. The parameter δ should be different from zero only if there are information events, and even then, the direction may not be clear.

In summary, the PIN and the single parameters of the model are independent of the subjective choices of trade interval and buckets size or vary with these parameters as expected from the model. The discrete setting is applicable in the high frequency trading data and thereby proves to not be inferior to the continuous model. The new model may even be superior regarding the convergence of the parameter estimation. The PIN estimates exhibit a strong size effect where the risk of information-based trading is significantly higher for low liquid stocks and lowest for the most actively traded stocks.

Table 3-3: Significance of ML estimates

This table shows the share of maximum likelihood estimations where the estimated parameters are different from zero at significant levels of 1% and 5%, by index membership for selected combinations of bucket size (BS) and trade intervals (TI). The eight columns show results for all four model parameters and two significance levels for each parameter. Robust standard errors were used to determine significance.

			alpha		delta		epsilon		mu	
BS	**TI**	**obs**	1%	5%	1%	5%	1%	5%	1%	5%
DAX										
	5	7,004	38%	57%	7%	19%	99%	99%	100%	100%
8	10	6,977	32%	52%	8%	22%	95%	96%	100%	100%
	20	6,932	20%	36%	10%	24%	94%	96%	99%	99%
	5	6,871	23%	42%	3%	14%	99%	99%	100%	100%
12	10	6,834	19%	38%	4%	16%	94%	95%	100%	100%
	20	6,722	12%	27%	5%	18%	93%	95%	99%	100%
	5	6,673	18%	35%	2%	10%	99%	99%	100%	100%
15	10	6,694	15%	32%	3%	12%	93%	95%	100%	100%
	20	6,489	10%	24%	4%	14%	92%	94%	100%	100%
MDAX										
	10	8,880	38%	54%	4%	12%	100%	100%	100%	100%
8	20	8,910	39%	54%	4%	13%	100%	100%	99%	100%
	30	8,920	37%	52%	5%	13%	99%	99%	98%	99%
	10	8,760	26%	41%	2%	9%	100%	100%	100%	100%
12	20	8,756	25%	40%	3%	9%	100%	100%	100%	100%
	30	8,763	24%	39%	3%	9%	98%	99%	99%	99%
	10	8,516	20%	34%	2%	7%	100%	100%	100%	100%
15	20	8,585	19%	34%	2%	7%	100%	100%	100%	100%
	30	8,591	18%	33%	2%	8%	98%	99%	99%	100%
SDAX										
	10	5,655	37%	48%	3%	7%	100%	100%	100%	100%
8	20	5,698	36%	49%	3%	7%	100%	100%	100%	100%
	30	5,766	36%	47%	3%	8%	100%	100%	100%	100%
	10	5,467	27%	39%	3%	6%	100%	100%	100%	100%
12	20	5,486	27%	40%	3%	6%	100%	100%	100%	100%
	30	5,628	27%	39%	3%	6%	100%	100%	100%	100%
	10	5,363	22%	35%	2%	5%	100%	100%	100%	100%
15	20	5,421	23%	35%	2%	5%	100%	100%	100%	100%
	30	5,524	21%	33%	3%	5%	100%	100%	100%	100%

3.5.2 *Event study with ad-hoc announcements*

After we have analyzed the PINs composition and general behavior, this section demonstrates the application of the new PIN model to evaluate the effectiveness of ad-hoc announcements. We discuss results for those trade intervals that have been shown in the previous section to fit each stock group's trading activity, i.e. 5 or 10 seconds for DAX, at least 10 seconds for MDAX and 30 seconds for the SDAX. Announcements without valid PIN estimates on any day in the event window are dropped. This requirement eliminates less than 10% of the announcements from DAX stocks, but of course a little more for MDAX and SDAX. In absolute numbers for the DAX stocks up to 102 events remain in 2005, up to 108 events for the MDAX stocks and up to 52 events in the SDAX.

The announcements we are studying are actually relevant and do affect the market. The first panel in table 3-4 shows the mean, median and standard deviation of the change in trading volume within the event window aggregated over all observable announcements within each index group [21]. Throughout all stocks in all indices and parameter combinations, the volume increases significantly from day -1 to day zero and then immediately falls from day zero to day +1. The clear spike in volume visible in the descriptive statistics is also confirmed by non-parametric tests. For the DAX stocks the significant volume growth starts already from day -2 to day -1 and also the decline after the event lasts one day longer (the volume change is not significant anymore from day -3 and day +3 onwards). Test statistics of multiple and pair-wise comparison of the volume change between event days in the last column of table 3-4 are all highly significant, so the announcements do convey information that affects the market and makes market participants trade.

The behavior of PIN during the event study by index membership is shown in table 3-5. Throughout all parameter settings, the PIN drops heavily on the event day in the DAX and MDAX. In the following two days, it recovers to its previous level. Mean or median PIN drop about half a standard deviation on day zero, recover almost completely on day +1 and reach the full pre-event level two days after the announcement. The SDAX does not exhibit a consistent movement during the event window. The Kruskal-Wallis test displayed in the last column of table 3-5 confirms the significance of the varying levels of PIN during

[21] Traded volume and the parameters bucket size and trade interval are theoretically independent. However, the set of observable announcements depends on the chosen bucket size and trade interval, as these affect the convergence rate of the maximum likelihood estimation. We discuss results for only one set of parameters, as results do not differ.

Table 3-4: Event study - volume, spread, order imbalance

The first few columns of this table show the mean of the change in trading volume, the time-weighted relative percentage spread and the relative absolute order imbalance over the event window, aggregated per index. The following two columns show the test statistic and p-value of a Kruskal-Wallis test on differences over all 5 days in the event window. The remaining columns show results of the Mann-Whitney test and Mood's Median test applied on successive days in the event window. The characters "*", "**", and "***" denote statistical significance at the 10%, 5% and 1% confidence level respectively.

		Mean					**Kruskal-Wallis**		**Mann-Whitney Test** (z-stat.)				**Mood's Median Test** (χ2-stat.)			
Index	**obs**	-2	-1	0	1	2	$\chi 2$	p-Value	(-2, -1)	(-1, 0)	(0, 1)	(1, 2)	(-2, -1)	(-1, 0)	(0, 1)	(1, 2)
Volume Change																
DAX	102	7%	13%	138%	-25%	-5%	150	.00 ***	-1.69 *	-6.63 ***	9.96 ***	-3.80 ***	6.35 **	31.37 ***	85.41 ***	11.29 ***
MDAX	99	44%	35%	316%	-5%	-5%	127	.00 ***	-1.11	-6.12 ***	8.91 ***	-1.45	1.23	27.77 ***	60.31 ***	1.23
SDAX	49	90%	50%	2600%	24%	74%	29.4	.00 ***	1.66 *	-3.57 ***	4.57 ***	-0.49	2.00	9.18 ***	18.00 ***	1.02
Time-Weighted Relative Percentage Spread																
DAX	102	.035	.035	.035	.035	.036	0.22	.99	1.63	-.71	-.82	.18	.00	.08	.00	.00
MDAX	99	.150	.146	.148	.145	.140	2.53	.64	-.11	1.40	.17	-1.27	.51	.51	.18	.51
SDAX	49	.340	.355	.347	.351	.337	0.28	.99	-.05	-.33	-.16	1.09	.04	.04	.04	1.02
Order Imbalance																
DAX	102	.084	.072	.074	.081	.080	3.53	.47	-.34	.25	-.15	-.22	2.82 *	1.96	.08	.08
MDAX	99	.170	.170	.144	.144	.159	4.57	.33	.26	.56	.07	.57	.18	2.44	.02	3.41 *
SDAX	49	.212	.217	.216	.220	.197	1.23	.87	.25	-.06	.20	-.03	.04	1.02	.04	1.02

Table 3-5: Event study - PIN

This table shows the behavior of PIN during the event window by index membership for selected combinations of bucket size (BS) and trade intervals (TI). The third column contains the number of announcements having a valid PIN estimate on every day in the event window and no interfering announcements nearby. The following columns show the mean, median and standard deviation of the PIN estimates on every event day aggregated over all announcements in the respective index. The last two columns show test statistic and p-value of a Kruskal-Wallis test on differences in PIN over all 5 days in the event window. The characters "*", "**", and "***" denote statistical significance at the 10%, 5% and 1% confidence level respectively.

			Mean					Median					SD					Kruskal-Wallis		
BS	TI	obs	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2	$\chi 2$	p-Value	
DAX																				
	5	102	.199	.203	.184	.193	.200	.197	.196	.178	.184	.198	.047	.047	.043	.047	.049	11.44	.02	**
8	10	96	.190	.194	.168	.185	.191	.187	.186	.162	.175	.194	.048	.051	.043	.049	.053	18.39	.00	***
	20	90	.177	.172	.147	.169	.173	.169	.169	.146	.164	.167	.051	.055	.048	.052	.058	18.14	.00	***
	5	84	.181	.190	.170	.187	.191	.182	.185	.168	.172	.186	.045	.048	.042	.050	.046	11.30	.02	**
12	10	83	.173	.181	.158	.175	.177	.168	.179	.155	.164	.176	.046	.049	.040	.052	.047	12.02	.02	**
	20	80	.159	.162	.141	.157	.161	.151	.155	.138	.153	.158	.047	.051	.044	.052	.049	11.39	.02	**
	5	70	.184	.180	.171	.184	.185	.182	.178	.165	.178	.179	.046	.043	.043	.048	.045	6.18	.19	
15	10	71	.180	.171	.155	.172	.170	.178	.171	.148	.169	.164	.047	.041	.044	.049	.045	14.82	.01	***
	20	62	.166	.158	.145	.158	.166	.167	.156	.145	.157	.163	.050	.046	.047	.047	.047	6.89	.14	
MDAX																				
	10	106	.341	.336	.301	.301	.308	.347	.322	.285	.295	.298	.123	.117	.093	.096	.100	12.14	.02	**
8	20	104	.324	.326	.290	.292	.303	.313	.314	.276	.286	.295	.117	.111	.096	.095	.104	9.52	.05	**
	30	108	.329	.315	.277	.287	.299	.310	.314	.266	.282	.287	.126	.115	.099	.099	.100	13.66	.01	***
	10	95	.319	.316	.278	.290	.290	.301	.315	.263	.279	.295	.122	.116	.086	.094	.092	11.31	.02	**
12	20	99	.327	.321	.276	.287	.295	.314	.318	.266	.279	.297	.119	.107	.089	.095	.090	17.17	.00	***
	30	96	.309	.319	.265	.281	.285	.286	.315	.255	.267	.278	.111	.108	.088	.100	.102	15.91	.00	***
	10	88	.321	.327	.277	.288	.295	.295	.314	.272	.282	.295	.128	.124	.084	.091	.096	10.11	.04	**
15	20	96	.315	.322	.265	.291	.300	.292	.299	.251	.276	.293	.119	.133	.085	.098	.103	12.58	.01	**
	30	90	.317	.318	.274	.289	.281	.304	.314	.257	.270	.280	.127	.128	.101	.107	.095	10.93	.03	**
SDAX																				
8	30	52	.401	.392	.386	.410	.357	.385	.359	.357	.375	.346	.174	.165	.154	.143	.148	3.47	.48	
12	30	37	.405	.384	.428	.371	.378	.354	.350	.394	.341	.355	.184	.136	.162	.156	.160	2.68	.61	
15	30	49	.404	.384	.383	.384	.349	.356	.343	.338	.355	.334	.226	.175	.158	.179	.151	1.52	.82	

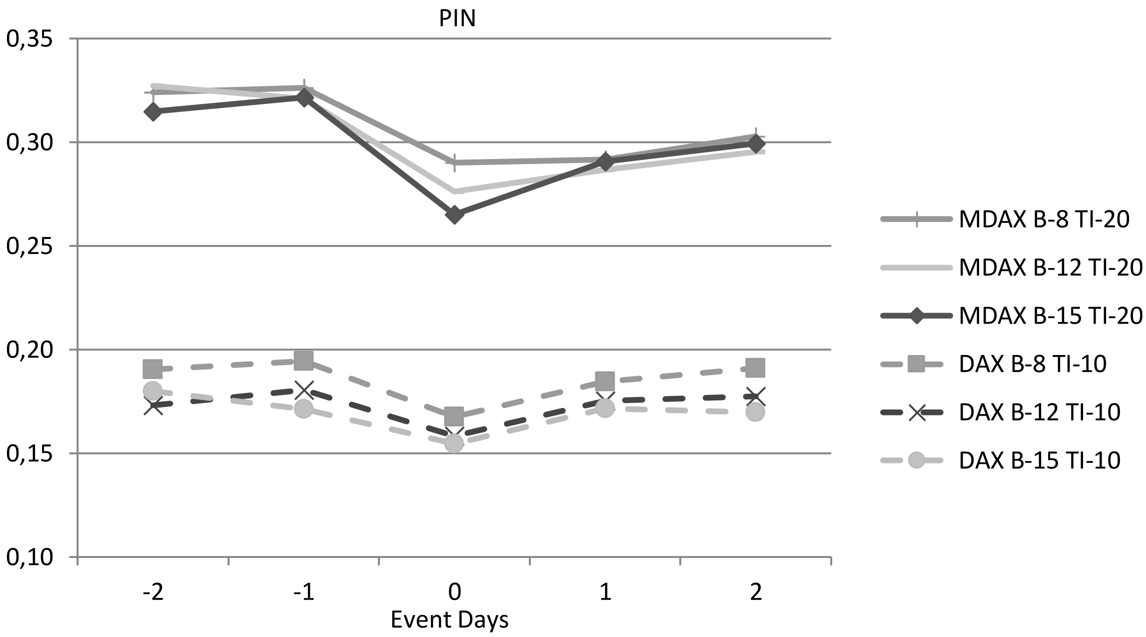

Figure 3-6: PIN over event window for different buckets

This figure plots the mean PIN within the event window for DAX and MDAX stocks; buckets size (B) is 8, 12 and 15 minutes, trade interval (TI) is constant at 10 seconds for the DAX and 20 seconds for the MDAX.

the event window for the DAX and MDAX at high levels of significance. Figure 3-6 illustrates the PIN movement for DAX and MDAX stocks. Ad-hoc announcements seem to be able to lower the information asymmetry in the market.

In both DAX and MDAX, the reaction of the PIN is also statistically significant for all feasible combinations of bucket size or trade interval length if we look at single transitions from day to day. Table 3-6 contains the test statistics of the Wilcoxon-Mann-Whitney tests and Mood's Median test for each transition within the event window. While volume already starts to increase from day -2 to day -1, the PIN stays flat on the day before the announcement and then drops significantly at the 1% confidence level on day zero. PIN subsequently increases in a similar fashion to its decrease from before. Both the Mann-Whitney test and the Mood's Median test identify the drop of PIN in the DAX and the MDAX as statistically significant.

Table 3-6: Event study - pairwise comparison of PIN during event window

This table compares PIN on successive days in the event window with Mann-Whitney and a Mood's Median test, by index membership for selected combinations of bucket size (BS) and trade intervals (TI). The characters "*", "**", and "***" denote statistical significance at the 10%, 5% and 1% confidence level respectively.

			Mann-Whitney Test (z-stat.)				**Mood's Median Test** (χ2-stat.)			
BS	**TI**	**obs**	(-2, -1)	(-1, 0)	(0, 1)	(1, 2)	(-2, -1)	(-1, 0)	(0, 1)	(1, 2)
DAX										
	5	102	-.33	2.92 ***	-1.17	-1.21	.08	6.35 **	.71	5.02 **
8	10	96	-.30	3.68 ***	-2.33 **	-.88	.00	12.00 ***	1.33	2.08
	20	90	.60	3.07 ***	-2.67 ***	-.45	.00	4.36 **	3.20 *	.00
	5	84	-1.32	2.83 ***	-1.69 *	-1.13	.10	9.52 ***	.86	4.67 **
12	10	83	-1.06	3.06 ***	-1.95 *	-.79	1.95	8.70 ***	2.92 *	2.92 *
	20	80	-.48	2.80 ***	-2.24 **	-.49	.90	6.40 **	8.10 ***	.10
	5	70	.65	1.51	-1.68 *	-.33	.46	4.11 **	4.11 **	.00
15	10	71	1.20	2.80 ***	-2.34 **	.32	.03	12.42 ***	8.14 ***	.03
	20	62	.87	1.50	-1.41	-.80	1.16	1.16	1.16	.52
MDAX										
	10	106	.32	2.37 **	-.11	-.47	1.21	4.83 **	1.21	.08
8	20	104	-.21	2.38 **	-.16	-.72	.00	4.92 **	.31	.69
	30	108	.48	2.68 ***	-.95	-.87	.07	7.41 ***	.67	.07
	10	95	-.26	2.68 ***	-.87	-.23	.02	4.74 **	1.03	1.03
12	20	99	-.01	3.21 ***	-.89	-.76	.02	5.84 **	.18	.99
	30	96	-.96	3.54 ***	-1.05	-.22	1.33	8.33 ***	.75	.33
	10	88	-.37	2.74 ***	-.79	-.40	.82	5.82 **	.36	.82
15	20	96	-.16	2.96 ***	-1.68 *	-.72	.08	5.33 **	2.08	.75
	30	90	.09	2.49 **	-.96	.07	.36	5.69 **	.36	.36
SDAX										
8	30	52	.37	.04	-.90	1.85 *	1.38	.00	1.38	2.46
12	30	37	.16	-1.06	1.55	-.30	.05	1.35	2.65	.05
15	30	49	.39	-.14	.04	.86	.04	.04	.37	.04

In contrast to the quick reversal in the DAX the PIN in the MDAX seems to take longer to "recover" and does not fully reach its pre-event level within two days. The rise from day zero to day +1 is not significant. The effect of ad-hoc announcements to lower information asymmetry appears to be a little more sustainable in the MDAX compared to the DAX. In the SDAX the volume peaks significantly on the event day and thereby confirms the impact the ad-hoc announcements have on the market. However, PIN does not consistently react in any direction. The high variance of PIN and the low number of events may prevent an existing movement from showing up in our results.

Having documented the statistically significant drop of PIN on the event day, we want to understand where this drop comes from. Do more uninformed traders trade to adjust their portfolios based on the new information? Do informed traders refrain from trading as for the time of the public release of information their competitive advantage is eliminated or smaller? In terms of the model, how do the key parameters affecting PIN – ε and μ – behave? We hypothesized informed trading to decrease and uninformed trading to increase on the announcement day. Figure 3-7 exemplifies the evolution of ε, μ and PIN over the event window for the announcements in the DAX and the MDAX and table 3-7 shows corresponding descriptive statistics and statistical tests for ε and μ. What can be observed is not only an increase in ε, but also an increase in the arrival of informed trades, μ. The jump of both μ and ε on day zero is highly significant for the DAX and MDAX across all combinations of buckets or trade intervals. The fall back to normal is significant in MDAX, but not so in the DAX. Mann-Whitney, Mood’s Median and Kruskal-Wallis test all confirm these results (not all shown). The increase in uninformed traders is large enough to offset the arrival of informed traders and lower the PIN and thereby the overall share of information-based trading significantly in the DAX and MDAX.

What is the motivation behind the observed trading activity? Uninformed investors could buy because they are motivated by the mere attention the ad-hoc announcement is bringing to the stock (Barber & Odean, 2008). In addition, the increase of informed trading despite the public release of information could be due to the additional private information required to correctly interpret the news. In a recent study Lambert, Leuz, and Verrecchia (2012) argue that what we refer to as “just” information asymmetry actually consists of two concepts, information asymmetry and average information precision. Both concepts are highly related but may move in opposite directions and are difficult to distinguish empirically. For the case of ad-hoc announcements, information precision should increase while information asymmetry decreases, although we cannot provide an empirical proof.

The lack of analyst following and media coverage that would provide supplementary information and interpretation of the news may weaken the positive effect of ad-hoc news in the very low volume stocks. In the SDAX the effect of volume on the arrival probabilities is visible in the whole sample but results suggest that the disclosure of ad-hoc announcements still does not attract enough uninformed traders to lower the share of information-based trading in the low volume stocks.

Summing up the results, there is statistically significant empirical evidence that the release of ad-hoc announcements lowers the information asymmetry in the market as measured by PIN. This positive effect, however, is only short-term sustainable, as the PIN recovers to pre-announcement levels within one or two days, depending on trading intensity. The results are consistent and robust across various settings of research design parameters and hold for both major indices DAX and MDAX. A similar consistent behavior in SDAX cannot be observed for the PIN variable, only an increase in volume from both informed and uninformed traders is visible.

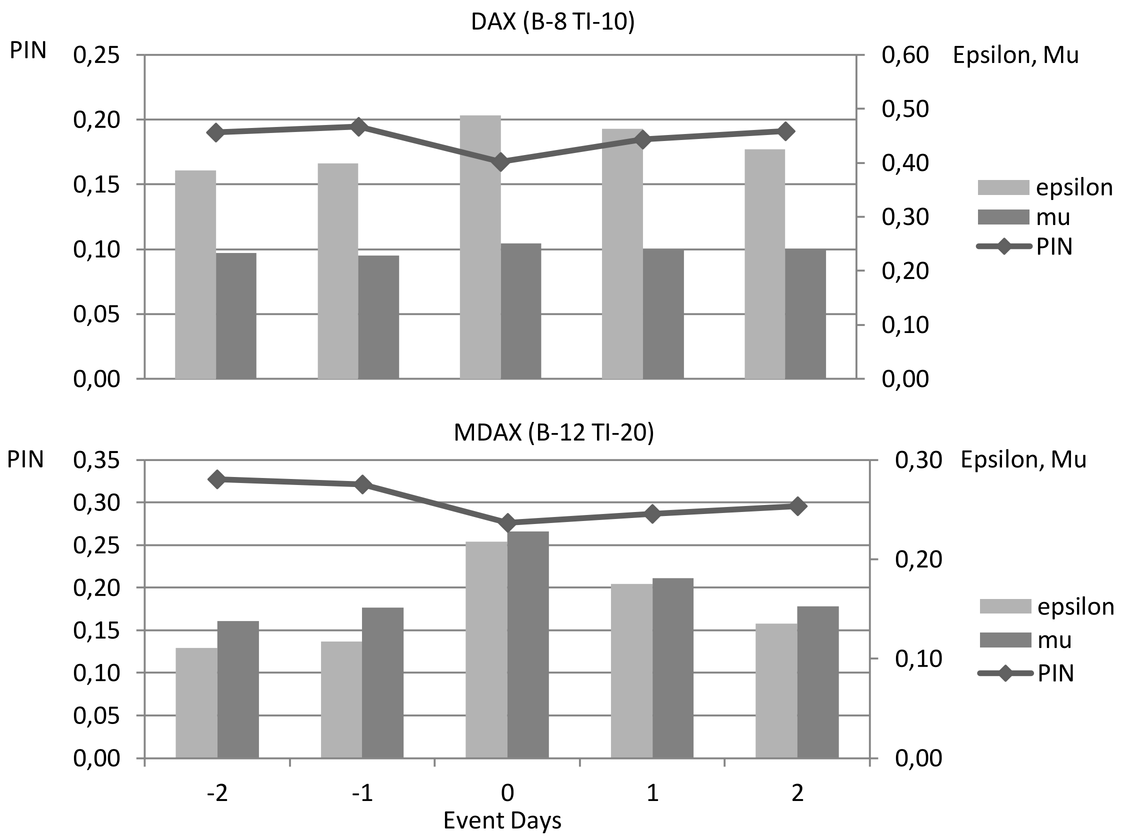

Figure 3-7: Parameters PIN, epsilon and mu in event window

This figure plots the mean PIN, epsilon and mu within the event window for DAX and MDAX stocks. The buckets size (B) is 8 minutes, the trade interval (TI) is 10 seconds for the DAX and 8 minutes and 20 seconds for the MDAX, respectively.

Table 3-7: Event study - behavior of model parameter epsilon and mu

This table shows the evolution of the arrival probabilities epsilon and mu over the event window, by index membership for combinations of bucket size (BS) and trade intervals (TI). The remaining columns contain the test statistics of Wilcoxon-Mann-Whitney tests on differences in parameter values on subsequent event days. *, **, *** denotes significance at the 10%, 5% and 1% confidence level respectively.

		Mean Epsilon					**Epsilon z-stat**				**Mean Mu**					**Mu z-stat**			
BS	**TI**	-2	-1	0	1	2	(-2, -1)	(-1, 0)	(0, 1)	(1, 2)	-2	-1	0	1	2	(-2, -1)	(-1, 0)	(0, 1)	(1, 2)
DAX																			
8	5	.234	.236	.292	.283	.250	-0.29	-2.88	0.68	1.43	.159	.159	.193	.173	.167	0.09	-3.77 ***	2.39 **	0.85
	10	.386	.399	.488	.463	.425	-0.56	-3.22 ***	1.04	1.23	.234	.228	.251	.239	.240	0.75	-2.74 ***	1.49	0.27
12	5	.208	.213	.260	.244	.225	-0.32	-2.73 ***	0.93	1.11	.137	.138	.163	.146	.143	-0.04	-2.71 ***	1.87 *	0.62
	10	.371	.378	.452	.439	.405	-0.27	-2.95 ***	0.65	1.14	.215	.203	.222	.220	.218	1.20	-2.23 **	0.25	0.47
15	5	.205	.209	.252	.244	.218	-0.40	-2.49 **	0.70	1.10	.127	.133	.151	.135	.133	-0.54	-2.07 **	1.60	0.28
	10	.365	.379	.445	.447	.420	-0.62	-2.43 **	0.22	0.70	.192	.201	.214	.204	.206	-0.98	-1.19	0.81	-0.03
MDAX																			
8	10	.059	.065	.127	.096	.073	-0.40	-4.75 ***	1.97 **	1.88 *	.100	.105	.171	.136	.109	-0.52	-5.62 ***	3.07 ***	2.58 ***
	20	.120	.130	.236	.187	.144	-0.35	-5.06 ***	2.16 **	2.23 **	.188	.198	.266	.231	.192	-0.53	-5.08 ***	2.99 ***	2.76 ***
	30	.176	.181	.303	.248	.205	-0.27	-4.78 ***	2.30 **	1.46	.236	.250	.323	.277	.252	-0.88	-5.04 ***	3.57 ***	1.74 *
12	10	.063	.068	.128	.099	.077	-0.41	-4.85 ***	2.25 **	2.23 **	.083	.091	.156	.111	.091	-0.67	-5.82 ***	4.11 ***	2.04 **
	20	.111	.117	.217	.175	.135	-0.30	-5.00 ***	2.01 **	2.04 **	.138	.151	.228	.181	.152	-0.97	-5.41 ***	3.54 ***	2.43 **
	30	.186	.190	.312	.253	.205	-0.07	-4.63 ***	2.23 **	1.85 *	.210	.218	.292	.238	.218	-0.74	-5.05 ***	4.14 ***	1.51
15	10	.067	.064	.126	.101	.078	0.12	-4.61 ***	1.82 *	1.77 *	.078	.082	.142	.108	.085	-0.14	-5.42 ***	3.12 ***	2.41 **
	20	.122	.119	.220	.173	.137	0.22	-5.11 ***	2.34 **	1.70 *	.144	.143	.231	.169	.142	0.10	-6.37 ***	4.75 ***	2.18 **
	30	.166	.164	.296	.235	.195	-0.04	-4.86 ***	2.14 **	1.35	.181	.192	.266	.213	.194	-0.74	-5.00 ***	3.95 ***	1.41
SDAX																			
8	30	.058	.062	.096	.077	.071	-0.05	-1.99 **	0.70	0.37	.156	.136	.206	.169	.157	0.41	-2.63 ***	1.28	0.59
12	30	.055	.055	.081	.069	.054	-0.50	-1.24	0.28	1.22	.118	.117	.158	.145	.127	-0.56	-1.86 *	0.56	0.88
15	30	.061	.063	.106	.091	.074	-0.14	-2.10 **	0.42	0.46	.100	.098	.158	.131	.116	-0.40	-2.67 ***	1.17	0.96

3.5.3 Robustness tests

Now that we have seen what research questions can be answered using the new estimation procedure for PIN we discuss the results of the proposed robustness tests. First, we stay within the event study setting and compare the behavior of PIN to other, simpler proxies of information asymmetry. Then we check whether the general assumptions of independence required to derive the maximum likelihood formula are met and also look at the share of boundary solutions produced by the maximum likelihood estimation.

3.5.3.1 Comparison with other information proxies

The original PIN model has been shown to be highly correlated with order imbalance and some scholars argue that PIN does not provide more insights than simple order imbalance itself (Aktas et al., 2007). The spread between bid and ask is also generally accepted as an indicator of information asymmetry (and liquidity). What is interesting to test is whether the simple metrics spread and OIB are able to capture the same movement that the intraday PIN discovered in the event study. If this is the case, then using these simple measures is much more effective and efficient as one does not have to go through tedious maximum likelihood estimation of a sequential microstructure trading model. If not, it is an indication of how useful the intraday PIN can be in future research. Of course this implication only holds if one acknowledges the general validity of the PIN model. Given the extensive, highly-regarded research with PIN, we allow ourselves to build our conclusions on this research and assume that PIN measures what it is meant to.

The lower two panels of table 3-4 provide descriptive results and test statistics for OIB and TWRPS equivalent to the event study analysis of PIN around the ad-hoc announcements. The TWRPS never shows any significant reaction. The relative order imbalance partly resembles the PIN movements, but not a single transition is significant, regardless of whether Kruskal-Wallis, Mann-Whitney or Mood's Median test are applied. The stability of the spread under the arrival of announcements with significant volume is a notable result. We interpret this as indication of substantial breadth, depth and resilience of the XETRA order book and the functioning of XETRA market design with roles like "Designates Sponsors" that ensure liquidity especially for illiquid stocks.

Overall, the PIN seems to be able to capture what none of the simpler measures like spread or OIB is able to. Neither does PIN simply resemble the volume volatility as there is no PIN reaction on the day -2 and day +2 where volume in DAX moves significantly.

3.5.3.2 Verification of model assumptions

Both trades and the arrival of news are assumed to be independent in the trading model. The first robustness check uses a runs test to test the independence of information events. We estimate the original model over non-overlapping windows of 33, 43, and 65 trading days as a benchmark to compare dependence with the intraday model presented in this chapter. This setting should give 8, 6 and 4 observations per stock for the year 2005. The trading activity with an average of 63 trades per day in the SDAX is very similar in magnitude to that of the one "very common stock" used in EKO97. Accordingly, the trade interval length is set to 5 minutes, equal to the one used in EKO97 and still acceptable given the low trading activity of these stocks. With these estimates we run a runs test as described in the preceding paragraph, simply replacing bucket by day. For the intraday PIN the length of the trade interval has no effect on α and is therefore kept constant at 5 seconds for the DAX, 10 seconds for the MDAX and 30 seconds for the SDAX. We would like to *not* reject the null hypothesis of news events being independent and hence present results with a relatively low confidence level of 10% that makes rejections more likely and lowers type-2 errors. The number of observations (i.e. security-day combinations) is reduced from the theoretical maximum by missing PIN estimates and, especially for the SDAX, also by very inactive stocks that partially prevent the application of the runs test[22].

Table 3-8 presents results. The interday estimation converges for stocks in SDAX and around 40% of MDAX stocks, but never does so for any stock in the DAX. For the intraday estimation, the independence assumption is rejected for the majority of stocks in the DAX. This ratio decreases to 24%-28% in MDAX and to just around 14% in the SDAX. The comparison with the original interday estimation in SDAX and MDAX, however, reveals that the dependence of news events is even higher in the interday model as it is in the intraday model. On a 10% confidence level, the share of observations where the assumption of independence is rejected in the MDAX with the interday estimation is 34% to 67% compared to 32% to 34% intraday. In the SDAX 80% of the old PIN estimates converge and the percentage of rejections of the independence hypothesis is clearly higher than in our new intraday estimation. Results of the test of independence of good and bad news are shown in the last two columns of table 3-8. Independence is not

[22] The run test is run with a continuity correction where in the case of fewer than 10 observations either above or below the threshold, more reliable critical values are used. This correction ensures for example that a sample with only one or two news buckets will always be rejected.

seriously violated as the share of rejections is well below 20% of all security-day combinations, for all stocks.

Table 3-8: Runtest on independence of information events

This table shows the results of a run test designed to test the independence of news events and the independence of good or bad news. The share of observations where the null hypothesis is rejected at two different significance levels is given per index and different bucket sizes. The trade interval is constant per index (DAX 5 sec, MDAX 20 sec, SDAX 30 sec). Results for the original interday PIN model with a 5 minute trade interval and a window length of 33, 43 and 65 trading days are also presented for comparison. In the DAX, the estimation of the original interday PIN model did not converge for any stock.

	Convergence		**Indep. of news events**		**Indep. of good and bad news**	
Index			% rejected		% rejected	
bucket	obs. converged	% of total	α = 5%	α = 10%	α = 5%	α = 10%
DAX						
8 min	7,004	97%	75%	80%	7%	14%
12 min	6,871	95%	67%	74%	6%	14%
15 min	6,673	93%	61%	71%	6%	16%
interday 33 days	0	0%	-	-	-	-
interday 43 days	0	0%	-	-	-	-
interday 65 days	0	0%	-	-	-	-
MDAX						
8 min	8,881	89%	28%	34%	6%	16%
12 min	8,761	87%	24%	33%	6%	15%
15 min	8,516	85%	24%	32%	6%	17%
interday 33 days	58	37%	27%	34%	4%	12%
interday 43 days	95	41%	33%	47%	5%	12%
interday 65 days	141	45%	59%	67%	5%	16%
SDAX						
8 min	5,768	64%	15%	20%	7%	14%
12 min	5,631	63%	13%	20%	6%	14%
15 min	5,524	61%	13%	18%	7%	16%
interday 33 days	194	85%	27%	34%	7%	18%
interday 43 days	254	84%	33%	41%	6%	12%
interday 65 days	136	89%	54%	65%	9%	18%

The independent arrival of trades throughout the day is tested by calculating the autocorrelations of buys, sells and no-trades per bucket and day for six lags and by running a lagged regression on the net trades. Table 3-9 shows the mean autocorrelation of buys and the share of correlation coefficients significantly different from zero for the three intraday buckets and the interday data for comparison, aggregated over the indices. The interday autocorrelation was calculated once per stock over the year 2005; the intraday autocorrelation was calculated once per day and stock, giving 257 observations per stock.

As we are interested in *not* rejecting the hypothesis that the coefficients are significantly different from zero, we test for significance at a lower confidence level of 10%. The final column displays the results of the Q-statistic, a chi-square type test on the joint hypothesis that all autocorrelation coefficients up to lag six are simultaneously equal to zero. Results for either buys, sells or no-trades were very similar, therefore we only discuss the results for buys and place corresponding tables for sells and no-trades in the appendix (See table A.4 for sells and table A.5 for no-trades).

Table 3-9: Autocorrelation of buys

This table shows the autocorrelation of the number of buys per intraday bucket of 8, 12 and 15 minutes and per day, for six lags. The daily autocorrelation was calculated once per stock over the whole year, the intraday autocorrelation was calculated once per day per stock, resulting in 257 values per stock for the year 2005. For each lag, the mean autocorrelation per index and the share of autocorrelation coefficients significantly different from zero at the 10% confidence level is displayed. The final column displays the share of observations where the joint hypothesis of all autocorrelation coefficients being simultaneously zero is rejected, measured by the Q-statistic with a 10% confidence level.

		Mean autocorrelation (Lags)						**% of AR coefficients sign. α=0.10**						**Q-Stat**
bucket	**obs**	1	2	3	4	5	6	1	2	3	4	5	6	**α=0.10**
DAX														
interday	28	.433	.245	.162	.167	.162	.111	100%	96%	68%	61%	75%	32%	100%
8 min	7,196	.378	.278	.227	.194	.161	.138	82%	65%	54%	46%	38%	31%	80%
12 min	7,196	.393	.279	.213	.167	.129	.093	75%	56%	42%	30%	20%	12%	70%
15 min	7,196	.406	.276	.200	.150	.098	.041	73%	50%	33%	21%	10%	3%	65%
MDAX														
interday	39	.415	.257	.216	.196	.178	.160	100%	95%	79%	74%	77%	64%	100%
8 min	10,023	.153	.094	.067	.052	.040	.026	36%	23%	17%	14%	12%	9%	33%
12 min	10,023	.159	.089	.059	.037	.021	.010	32%	18%	13%	10%	7%	6%	26%
15 min	10,023	.166	.089	.050	.027	.010	-.008	29%	16%	11%	7%	5%	4%	22%
SDAX														
interday	35	.393	.255	.188	.151	.145	.134	97%	82%	74%	63%	55%	58%	97%
8 min	8,995	.054	.027	.018	.011	.005	.001	16%	11%	10%	8%	7%	7%	16%
12 min	8,995	.055	.023	.010	.000	.00	-.01	15%	10%	8%	7%	6%	5%	13%
15 min	8,995	.054	.021	.003	-.01	-.01	-.02	14%	9%	7%	5%	4%	4%	11%

The mean intraday autocorrelations (AR) of buys for DAX stocks are marginally smaller than the interday AR, but both are at levels of 40% and 25% in the first and second lag that may conflict with the assumption of independence. Although relatively high in absolute values, the dependence in DAX is clearly lower intraday than interday when measured by the share of significant AR coefficients. In the medium and low liquid stocks in MDAX and SDAX, the mean intraday AR comes close to zero from lag two on and is always by magnitudes smaller than the interday AR. A longer bucket also brings down the autocorrelation, which gives more credibility to PIN results obtained with the 12 or 15 minute bucket.

Therefore, similar to the independence of news, the original interday model clearly violates the independence assumption for the trading data used in the current study. In contrast, the AR calculations for the new intraday model result in lower values and indicate that the number of buys in subsequent buckets of 8 to 15 minutes within a day can be regarded as independent, at least for medium to low liquid stocks and to some extent for the high liquid stocks.

Another approach to test for the dependence of trades is a lagged regression on net-trades, i.e. buys minus sells. Table 3-10 summarizes the results per index. Columns 3 to 15 display the mean of the coefficients for each lag and the percentage of regressions where the coefficient of that specific lag was significantly different from zero at a confidence level of 10%. The mean coefficients of the intraday model in the first lag of buys are about half that of the interday model. Almost all coefficients of lag 1 of the interday trade count are significant whereas this share is less than quarter of all regressions in intraday data. From lag 2 on, the mean coefficients fluctuate in a narrow band around zero and the share of significant coefficients drops to or even below 10%.

Table 3-10: Lag regression of net trades

This table shows the mean coefficients and the percentage of coefficients significantly different from zero of a regression of net trades (buys minus sells) on 6 lags. The net trades and the respective regression are based the intraday buckets of 8, 12 and 15 minutes and on daily data. For the daily net trades one regression per stock is run over the whole year. The intraday regressions are run once per stock and trading day and the results are averaged per stock and index. The second column for each lag denotes the percentage of regressions where the coefficient of that specific lag is significantly different from zero at the 10% level.

		Lag 1		**Lag 2**		**Lag 3**		**Lag 4**		**Lag 5**		**Lag 6**	
bucket	**obs**	øβ	sign	øβ	sign	øβ	sign	øβ	sign	øβ	sign	øβ	sign
DAX													
interday	28	.22	96%	.08	29%	.00	7%	.06	21%	.00	14%	.04	25%
8 min	7,196	.12	26%	.02	13%	.03	11%	-.01	10%	.01	10%	-.01	10%
12 min	7,196	.12	21%	.01	11%	.01	10%	-.02	9%	.00	9%	-.03	9%
15 min	7,196	.12	18%	-.01	11%	.01	10%	-.03	9%	-.01	8%	-.05	8%
MDAX													
interday	39	.19	74%	.04	23%	.03	18%	.01	21%	.04	23%	.01	5%
8 min	10,023	.10	24%	.01	12%	.02	11%	-.01	10%	.00	9%	-.02	9%
12 min	10,023	.09	19%	-.01	10%	.01	9%	-.03	8%	-.01	8%	-.03	8%
15 min	10,023	.09	16%	-.02	9%	.00	9%	-.04	8%	-.01	7%	-.04	7%
SDAX													
interday	35	.22	82%	.05	37%	.03	21%	.01	24%	.04	18%	.03	13%
8 min	8,995	.04	16%	-.01	12%	.01	10%	-.02	10%	.00	8%	-.02	9%
12 min	8,995	.03	14%	-.03	10%	.00	9%	-.03	9%	-.01	8%	-.03	8%
15 min	8,995	.03	13%	-.04	10%	-.01	8%	-.05	8%	-.02	7%	-.04	8%

In summary, the results of these robustness checks suggest that our intraday PIN model conforms with the crucial assumptions of independence to a much higher degree than the original interday model. If assumptions of independence have to be rejected for the intraday model, this happens to a much lower degree than in the original, widely applied and scientifically accepted interday model of PIN.

3.5.3.3 Boundary solutions in maximum likelihood estimation

Finally, we check for the existence of boundary solutions and the empirical distribution of the models parameters. In none of the thousands of estimations (28 securities times 257 trading days results in 7196 observations in DAX, 10023 in MDAX, 8995 in SDAX) are ε or μ exactly equal to 1 or 0. This is different for α and δ, results are given in table 3-11. In the DAX, there is one single boundary solution for α and none for δ. In MDAX and SDAX the parameter α and δ do take on boundary values of 0 or 1, but the share is well below 1% in the MDAX and below 5% in the SDAX, which we deem acceptable. An explanation may be that the lower liquidity in these stocks makes a day with some activity and a price gain in any direction likely to produce a boundary solution for δ.

When we look at histograms of the estimated model parameters in Figure 3-1, Figure 3-2 and Figure 3-3 for DAX, MDAX and SDAX respectively, the picture stays the same for the DAX and the MDAX, but in the SDAX it becomes visible that “boundary solutions” do not fall exactly on 1 or 0 but rather pool close to 1 or 0. Still, in the DAX and MDAX, where the majority of conclusions of this research are drawn from, all parameters, except the rather unimportant δ in MDAX, behave very well and the number of boundary solutions is either zero or far away from being able to bias the sample.

Table 3-11: Boundary solutions of parameters alpha and delta

This table shows the relative and absolute frequencies of boundary solutions, i.e. a parameter being zero or 1, in the maximum likelihood estimation of the trading model.

		alpha				delta			
BS	**TI**	α == 1	%	α == 0	%	δ == 1	%	δ == 0	%
DAX									
8	**5**	0	0%	0	0%	37	1%	41	1%
	10	1	0%	0	0%	44	1%	63	1%
	20	99	1%	0	0%	87	1%	113	2%
	30	576	9%	0	0%	135	2%	154	2%
12	**5**	0	0%	0	0%	110	2%	95	1%
	10	2	0%	0	0%	122	2%	117	2%
	20	133	2%	0	0%	172	3%	173	3%
	30	662	11%	0	0%	223	4%	234	4%
15	**5**	1	0%	0	0%	138	2%	140	2%
	10	1	0%	0	0%	179	3%	191	3%
	20	127	2%	0	0%	204	4%	225	4%
	30	679	12%	0	0%	245	4%	287	5%
MDAX									
8	**5**	36	0%	0	0%	585	6%	289	3%
	10	30	0%	0	0%	616	7%	335	4%
	20	35	0%	0	0%	635	7%	370	4%
	30	46	0%	0	0%	660	7%	388	4%
12	**5**	35	0%	0	0%	686	8%	376	4%
	10	32	0%	0	0%	753	8%	414	5%
	20	42	0%	0	0%	822	9%	462	5%
	30	49	1%	0	0%	828	9%	510	5%
15	**5**	46	1%	0	0%	789	9%	447	5%
	10	61	1%	0	0%	869	10%	487	5%
	20	73	1%	0	0%	891	10%	543	6%
	30	74	1%	0	0%	949	10%	595	6%
SDAX									
8	**5**	41	1%	0	0%	1,159	17%	771	11%
	10	38	1%	0	0%	1,212	17%	817	12%
	20	41	1%	0	0%	1,235	17%	835	12%
	30	64	1%	0	0%	1,237	17%	861	12%
12	**5**	49	1%	0	0%	1,246	18%	809	12%
	10	59	1%	0	0%	1,277	18%	818	12%
	20	75	1%	0	0%	1,307	18%	847	12%
	30	97	1%	0	0%	1,310	18%	889	12%
15	**5**	77	1%	0	0%	1,300	20%	805	12%
	10	80	1%	0	0%	1,323	19%	840	12%
	20	95	1%	0	0%	1,377	20%	883	13%
	30	127	2%	0	0%	1,418	20%	930	13%

3.6 Conclusion

We have shown in this chapter how the estimation procedure for the microstructure model by Easley et al. (1997b) can be modified such that the PIN variable can be estimated once per day instead of having one static PIN over at least 30 trading days. This intraday PIN overcomes the problem of convergence for high frequency trading data, conforms with the crucial assumptions of independence to a much higher degree and enables the PIN to be applied in a short horizon event study context. Further, we find our approach appealing because of its simplicity. It still solves the issues mentioned before that other variations of the PIN model aim to solve (Lei & Wu, 2005; Tay et al., 2009), but do so with more complex mathematical concepts that limit intuitive interpretation.

The new intraday PIN and the significance of the presented results are robust to the two somewhat arbitrary choices that have to be made for estimating the model's parameters - the length of an intraday bucket and the length of a trade interval. Feasible ranges are given by the speed of information processing and the trading frequency, respectively.

We estimated the model on trading data from a sample of 102 German stocks in 2005, all of them belonging to one of the three major indices DAX, MDAX and SDAX. The probability of an information-based trade is highest in the most actively traded stocks and declines with decreasing volume, confirming previous literature. The probability of the arrival of an information event is also higher for high volume stocks, albeit the differences between stock segments are less pronounced. It is the large surplus in the arrival of uninformed traders, however, that lets the share of information-based trading relative to all trading activity (i.e. PIN) be lowest for high volume stocks and lets the PIN be significantly higher in medium and low volume stocks.

Applied in an event study around official announcements stipulated in German insider trading legislation, the complex composite variable PIN captures aspects that none of the simpler measures like spread or relative order imbalance were able to. The PIN decreases significantly on the day of disclosure for high and medium liquid stocks. Accepting the PINs validity allows drawing the conclusion that ad-hoc announcements fulfill their mission of disclosing price-sensitive information to all market participants simultaneously and thereby lower the information asymmetry in the market. This positive effect is albeit relatively short-lived as the PIN in the DAX reaches its pre-announcement level within a day and within two days for medium liquid stocks. We find that the valuable insights to be gained by employing the intraday PIN justify the considerable effort required to estimate the model parameters.

4 The Sensitivity of VPIN to the Choice of Trade Classification Algorithm

4.1 Introduction

In the trading of any security, some trades are motivated by private information, while others are purely liquidity-driven. Research in market microstructure models trading participants and their behavior stemming from different motivations. Based on these models, empirical estimation procedures aim to extract from the observed trading activity the degree to which trades were motivated by information. The model by Easley et al. (1996a) and its derivatives are widely employed examples, but they lack the applicability in high frequency trading environments. In the previous chapter, we have demonstrated how an improved estimation procedure can mitigate these issues. In parallel to this research, Easley et al. (2012b) present a new evolution of this model intended to measure the degree of information asymmetry especially in high frequency trading environments - the volume-synchronized probability of informed trading (VPIN). They show that VPIN is a predictor of short-term volatility and that it would have been able to signal the “flash crash” on May 6, 2010, hours ahead of the actual crash (Easley, Lopez de Prado, & O'Hara, 2011; Easley et al., 2012b).

In this chapter, we test the robustness and sensitivity of VPIN to one of its key design parameters - trade classification - to guide researchers in the application of VPIN. The computation of VPIN requires the separation of buys from sells in the raw trading data, which is a common prerequisite for the estimation of microstructure models and hence well studied. Instead of relying on the established deterministic algorithms, however, Easley, Lopez de Prado, and O'Hara (2012a) also introduce a new method to classify trades into buys and sells, called bulk volume classification (BVC). Unlike traditional classification algorithms, which operate on a trade-by-trade level, BVC defines a fraction of volume as either buyer or seller initiated simply based on the price movement within a short period of time – an inherently heuristic and potentially less accurate approach. Since the seminal work of Lee and Ready (1991a), the performance of trade classification algorithms has steadily improved and reaches accuracies above 90% (Chakrabarty, Li, Nguyen, & Van Ness, 2007; Ellis et al., 2000; Odders-White, 2000). On a proprietary subsample of data provided by Deutsche Boerse, we confirm this finding also for the recent trading data employed in this study: the complex algorithms that use both trade and quote data achieve a classification accuracy of 90.5%.

Our analysis is motivated in the first place by asking whether the switch to heuristics is critical, if deterministic methods are so accurate. Additionally, if we assume that VPIN is a robust metric and bulk volume classification fulfills its purpose, we examine whether every reasonable choice of trade classification theme does in fact produce roughly the same results. There is one caveat in the reasoning so far. Easley et al. (2012b) argue that informed (high frequency) traders do not trade aggressively on their information with market orders anymore but instead create price pressure through fast, sliced limit order entries combined with immediate cancellations or similar trading strategies designed to hide true intention. Bulk volume classification, through its focus on price changes, is intended to capture this new characteristic better than precise “theoretical” order imbalance. This crucial assertion, however, is not empirically proven. Even if it is true to some extent, there is certainly a difference to what degree single securities and especially different market places are affected by this change. Hence, evaluating VPIN’s robustness to, in other words, varying hypothesis of how informed traders trade, is of high interest. We further address this concern by incorporating analysis on VPIN’s incremental predictive power for future volatility.

To test the sensitivity of the VPIN model to the choice of trade classification algorithm, we compare the outcomes of the different approaches on all levels up the ladder of aggregation towards empirically relevant metrics. We start with basic trade classification results, aggregate the data to compute order imbalance, next compute VPIN and finally use the different versions of VPIN to compare the occurrence of “toxic periods” (Easley et al., 2012b), the major proposed application of VPIN. A practical example applies different VPIN versions to a crash of a blue chip stock that dropped by 24% within a day.

Only few studies have dealt with the robustness of VPIN, least VPIN’s sensitivity to the choice of trade classification algorithm. But the ones that have, stir a controversial debate around VPIN’s benefits and potential flaws, such as its incremental predictive power in comparison to autocorrelation of existing measures of volatility (Andersen & Bondarenko, 2013, 2014a, 2014b, 2015; Easley, Lopez de Prado, & O'Hara, 2014; Wu, Bethel, Gu, Leinweber, & Ruebel, 2013b). Chakrabarty, Pascual Gascó, and Shkilko (2012b) use order imbalance estimation and the detection of toxic events to compare trade-by-trade algorithms with the actual BVC. Andersen and Bondarenko (2015) show on a sample from

S&P 500 futures with near 100% classification accuracy that transaction based classification yields better results than BVC. More importantly, they conclude that VPIN(BVC) is not a superior indicator of future volatility or toxicity and question VPIN's ability to detect events like the flash crash (Andersen & Bondarenko, 2014b). The controversial results suggest that more empirical evidence is needed.

The bottom-up approach employed in this study covers both the foundations and the applications of VPIN. In addition, this study enhances the existing evidence on VPIN in several dimensions. So far, the published empirical evidence on VPIN focuses on the ultra-high-frequency trading E-mini S&P 500 future on the CME Globex (Andersen & Bondarenko, 2014b; Easley et al., 2011, 2012b). This study instead provides empirical evidence for the applicability of VPIN in the high frequency trading of equities. Trading in single stocks instead of market-wide indices can be expected to have a much higher share of private information, hence research on the applicability of VPIN is greatly desirable. The trading data in this chapter comes from a large sample of the top 30 German stocks covering the full year 2012, in contrast to the two future contracts analyzed by Easley et al. (2012b) and one contract in Andersen and Bondarenko (2014b). And it originates from a pure electronic limit order book without any market maker interference, in contrast to the WTI futures contract used in Easley et al. (2012b) where a (small) share of volume is traded in OpenOutcry. Further, we evaluate not only the tick rule but also more complex and presumably more precise trade-by-trade classification algorithms. Extending the geography to Germany also has the advantage of the least fragmented order flow compared to leading equity markets in the US and UK. Our trading data captures almost 70% of all trading in the respective stocks, whereas in the US a single venue covers, on average, barely a third of total trading[23]. Further, we add a new example of a crash similar to the "flash crash" to VPIN's detection track record.

Our findings indicate that there is a high discrepancy between heuristic and deterministic trade classification already at the trade level. The discrepancy does not diminish in higher aggregated metrics, but instead increases when we compute order imbalances and especially VPIN. Even though the average VPIN values are similar, the correlations of the time series are just around 54%. In the detection of toxic periods, the major proposed application of VPIN, both approaches do not give consistent results more often than in 60% of the cases. Further, neither of the approaches is consistently faster or earlier in

[23] Source: fragmentation.fidessa.com/indexstats

detecting toxic periods. Regression analysis identifies volume and return volatility as parameters that contribute to a higher sensitivity of VPIN estimates to the choice of classification algorithm. These are exactly those trading conditions where VPIN is supposed to be most useful. An examination of a crash of K+S, a German blue chip stock, on July 30th 2013, where K+S' stock dropped 24% within the first few hours of trading, reveals that VPIN is able to predict this crash, but only if trade-by-trade classification is used and not bulk volume classification.

The remainder of this chapter is organized as follows. Section 2 reviews the trade classification algorithms and summarizes empirical results regarding their performance. Section 3 introduces the VPIN model, toxic periods and explains how we run the evaluation. Section 4 describes the data employed in this chapter. Section 5 presents results and section 6 concludes.

4.2 Trade classification algorithms and their performance

Two major trends affect the performance of trade classification algorithms in recent trading data. Computerization of trading with precise millisecond timestamps should benefit deterministic algorithms, as data inaccuracies such as delayed quotes become less of an issue. On the other hand, high frequency and algorithmic trading may make trade classification more challenging. What we call "traditional" algorithms are the trade-by-trade algorithms. They classify each single trade individually based on the preceding or succeeding trade price or, more commonly, use the vicinity to the prevailing best bid or best ask to determine the trade direction. In contrast, the bulk volume classification defines a proportion of the total volume traded within a given timeframe as buy or sell volume, based on the price movement within that timeframe.

4.2.1 Trade-by-Trade classification algorithms

The tick rule is the simplest classification algorithm due to its neglect of quote data. This is in turn its greatest advantage, since only trade prices are needed. To classify trades, the tick rule compares the current trade price with the preceding one. Trades with a price higher than the previous one (uptick) are classified as buys, trades settling at a lower price (downtick) than the previous trade are classified as sells. In case of equal prices the last preceding price change is used. An early application can be found in Holthausen, Leftwich, and Mayers (1987). The tick rule is part of every complex trade-by-trade classification algorithm. The reverse tick rule is an alternative version of the tick rule, which was first introduced by Hasbrouck (1988) to classify midpoint trades. In contrast to the tick rule, the change of the following price is used for classification. Lee and Ready (1991a) show that both methods achieve the same classification for price reversals but overall the tick rule outperforms the reverse tick rule. The tick rule can lead to misclassifications if quotes move contrarily to the trade direction or if traded prices move through the best bid or best ask levels.

The quote rule compares trade prices to the prevailing quote at the time of the trade. Trades at or above the ask are classified as buys, trades at or below the bid as sells. Trades inside the spread are classified based on their proximity to the bid and ask (Harris, 1989). Trades at the midpoint of the spread, however, cannot be classified with the quote rule. That is why more sophisticated algorithms combine quote and tick rule: trades at the midpoint are always classified by the tick rule and trades at best bid or best ask are always classified by the quote rule. The three most common algorithms differ only in the choice of how to split up the remaining trades between quote and tick rule, as illustrated in figure 4-1. The first of this group of algorithms uses the quote rule for all trades except midpoint trades (Lee & Ready, 1991a, "LR"). Then it took research almost ten years to turn this principle up-side-down and apply the tick rule to all trades except those at best bid and best ask, resulting in slightly better performance (Ellis et al., 2000, "EMO"). Another variation slices the spread in three parts to use the quote rule for trades close to bid and ask and apply the tick rule to the 20% band around the midpoint, as well as for trades outside the spread (Chakrabarty et al., 2007, "CLNV").

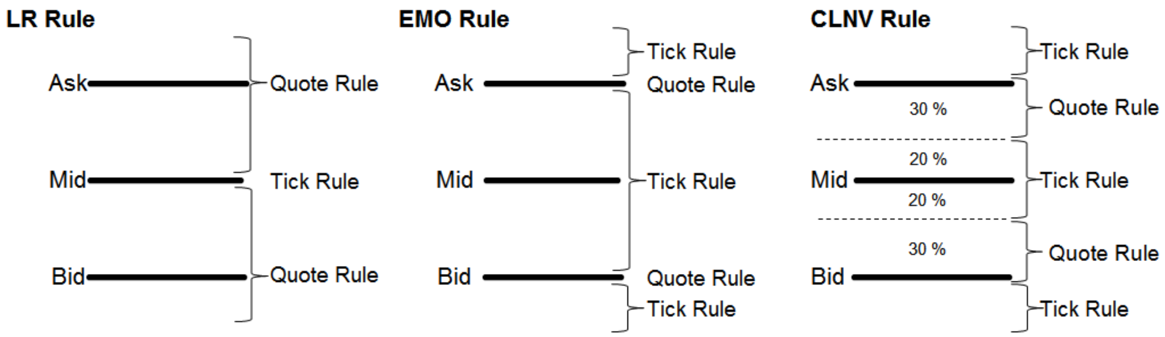

Figure 4-1: Classification algorithms

This chart illustrates the functioning of three different trade-by-trade classification algorithms: LR by Lee and Ready (1991), EMO by Ellis et al. (2002) and CLNV by Chakrabarty et al. (2007).

4.2.2 *Bulk classification*

Similar to the tick rule, bulk classification only needs tick data to infer trade direction. Instead of classifying trade by trade, however, bulk classification determines the share of buys and sells of a chunk of aggregated trading volume. Trades are aggregated in either time or volume bars[24]. Using one minute time bars, as proposed by Easley et al. (2012b), trades in each minute are aggregated by summing up their volume and by signing the bar with the last trade price in that minute. Bars with zero trading volume are excluded. Next, the price changes ΔP_i between time bars and the standard deviation of the price changes, $\sigma_{\Delta P}$ are calculated. Buy and sell volume of a time bar i are then defined as:

$$V_i^B = V_i \cdot Z\left(\frac{\Delta P_i}{\sigma_{\Delta P}}\right),$$
$$V_i^S = V_i \cdot \left(1 - Z\left(\frac{\Delta P_i}{\sigma_{\Delta P}}\right)\right) = V_i - V_i^B \tag{4-1}$$

where V_i^B and V_i^S is the respective buy and sell volume and V_i the volume of a bar. The cumulative distribution function (CDF) of the standard normal distribution, denoted by Z, weights the volume towards buys or sells. If the price change is zero, the volume is equally weighted between buy and sell volume ($Z\left(\frac{\Delta P_i}{\sigma_{\Delta P}} = 0\right) = 0.5$). If the price increases (decreases), it is more weighted towards buy (sell) volume, with the weight depending on the value of the price change. In the alternative approach that uses volume bars instead of time bars, trades are aggregated in bars with a constant amount of volume instead of time.

The procedure of bulk classification with one-minute time bars is illustrated in table 4-1. Panel A depicts a sample of raw tick data with time, price, and volume. These trades are then aggregated and classified in panel B. Table 4-1 clearly shows how the size of the price change influences the buy and sell volume. Relative to all price changes, $\Delta P_1 = -0.19$ is a large decline, and consequently the whole volume in the first time bar is classified as sell volume.

---

24 Easley et al. (2012b) prefer time bars, because vendors like Bloomberg often provide their data pre-compressed in time bars.

Table 4-1: Bulk volume classification of trading data

Panel A depicts a short excerpt of Adidas tick data from July 2, 2012 to illustrate bulk classification within one-minute time bars. Only a few trades at the start and end of every minute are reported. Panel B shows how the trades are aggregated into one minute time bars and classified using the cumulative distribution function previous price changes. P_i is the price of the last trade within the time bar. ΔP_i is the price change relative to the previous time bar. CDF is the normalized cumulative distribution function of the price change from bar *i-1* to bar *i* to estimate the share of buys and sells. V_i is the total volume traded within bar *i*. V_i^B and V_i^S are the resulting volume classified as buy and sell volume, respectively, within time bar *i*.

Panel A: Trading data			**Panel B: Time bar aggregation**								
Time	Price	Vol	Time bar	Start	End	P_i	ΔP_i	CDF	V_i	V_i^B	V_i^S
...more trades			2	09:02:00,001							
09:02:57,378	55,54	8	2		09:03:00,000	55,54	-0,19	≈ 0	9959	0,0	9959,0
09:03:00,125	55,51	73	3	09:03:00,001							
09:03:02,160	55,50	500	3								
09:03:02,364	55,52	228	3								
...more trades			3								
09:03:55,795	55,49	228	3								
09:03:55,795	55,49	92	3								
09:03:56,035	55,49	111	3		09:04:00,000	55,49	-0,05	0,136	7195	977,8	6217,2
09:04:00,517	55,51	230	4	09:04:00,001							
...more trades											
09:04:59,096	55,57	5	4		09:05:00,000	55,57	0,08	0,961	14240	13678,9	561,1

4.2.3 Performance of traditional trade classification algorithms

Motivated by the need for classified trading data to estimate different microstructure models, various algorithms have been proposed and their performance been evaluated on markets all over the world. The most commonly used algorithm was proposed by Lee and Ready (1991a) (LR rule). Their analysis of trade and quote data of 150 firms listed on NYSE in 1988 indicate that quotes are recorded ahead of the trade that triggered them. A proposed quote delay of 5 seconds mitigates this issue and increases the reliability of the LR rule. We summarize the results of the most important empirical studies that use actual classification data in table 4-2.

Most studies evaluate their algorithms on trading data from NASDAQ and NYSE. The LR rule performs fairly well with 73% to 91% accuracy, depending on the study and market. Based on an analysis of the classification accuracy of trades within the spread on a NASDAQ data set, Ellis et al. (2000) propose a new classification algorithm (EMO) that improves accuracy on their data to 81.87%. Chakrabarty et al. (2007), in turn, propose another variation (CLNV), which outperforms the EMO algorithm on NASDAQ data by 0.72 percentage points (76.52%). In comparison to these two slightly more complex algorithms, the LR rule performs slightly better in two out of five studies, but is outperformed in the other three studies by a few percentage points. On the market where VPIN has been studied so far, the Chicago Mercantile Exchange (CME), the tick rule comes close to 90% accuracy (Andersen & Bondarenko, 2015; Easley et al., 2012a).

Given that all evidence collected in table 4-2 neither originates from the trading venue studied in the current study, nor is it very recent, at least for studies on equity data, we reached out to Deutsche Boerse to help us validate one of our key assumptions for this chapter – that trade-by-trade algorithms perform reasonably well even in recent market conditions. Deutsche Boerse provided us with a proprietary sample of trading data for 10 trading days from 12/3/2012 to 12/14/2012 for 10 stocks of our original sample[25]. This timeframe is outside of the sample period of the main empirical analysis in this chapter.

Table 4-2: Accuracy of traditional classification algorithms

This table lists the accuracy of trade classification algorithms per empirical studies. Column two and three denote the originating exchange and the first year of the considered data set. Column three to nine state the reported performance for each employed algorithm: tick rule (Tick), reverse tick rule (Re. Tick), quote rule (Quote), Lee-Ready algorithm (LR), algorithm by Ellis, Michaely, O'Hara (EMO), algorithm by Chakrabarty, Li, Nguyen, Van Ness (CLNV). Peterson and Sirri (2003) examined two periods during a NYSE tick change. Chakrabarty et al. (2012a) reported the accuracy of short sales (1) and long sales (2). The data of Aitken and Frino (1996) did not include trades occurring between the spread and quote changes preceding trades. Savickas and Wilson (2003) excluded midpoint trades in their sample. CME futures trades in Easley et al. (2012a) could only occur at quotes. The final row presents results from the current study.

Study	**Data**	**Year**	**Tick**	**Re. Tick**	**Quote**	**LR**	**EMO**	**CLNV**
Aitken and Frino (1996)	ASX	1992	74.4					
Lee and Radhakrishna (2000)	NYSE	1990				93.0		
Odders-White (2000)	NYSE	1990	78.6		75.0	85.0		
Tanggaard (2004)	NYSE	1990	81.5		73.4	84.7		
Finucane (2000)	NYSE	1990	83.3	72.1		84.4		
Ellis et al. (2000)	NASDAQ	1996	77.7		76.4	81.1	81.9	
Theissen (2001)	FWB	1996	72.2			72.8		
Peterson and Sirri (2003)	NYSE	1997(1)				91.0	90.8	
	NYSE	1997(2)				87.4	88.2	
Savickas and Wilson (2003)	NASDAQ/ NYSE	1995	60.8		79.0	79.0	74.4	
Chakrabarty et al. (2007)	NASDAQ	2005	75.4			74.4	75.8	76.5
Chakrabarty, Moulton, and Shkilko (2012a)	NASDAQ	2005(1)				78.6		
Chakrabarty et al. (2012b)	NASDAQ	2005(2)				78.3		
Easley et al. (2012a)	CME	2011	86.4					
Andersen and Bondarenko (2015)	CME	2006-2011	88.4					
The current study	XETRA	2012	82.0			89.6	90.4	90.5

[25] We thank Deutsche Boerse and in particular Dr. Miroslav Budimir, Dr. Alexandra Hachmeister and Thomas Katzschner for their advice and support.

The data contains all matching orders, limit or market orders, which compose each trade. The millisecond timestamp of each order then allows us to identify the order that came last and thereby led to the execution of the trade, i.e. the aggressor side. For about 10% of the trades in the data from Deutsche Boerse, the timestamps of the constituting orders are identical on the millisecond level, hence we cannot distinguish the true trade initiator for these trades and have to exclude them from the evaluation. Data from ThomsonReuters is acquired accordingly for the same sample and processed exactly in the same way as the trading data for the main empirical analysis (see section 4). Table 4-3 displays the evaluation sample and results. We are able to match each single trade from Deutsche Boerse to our trading data, although the timestamps between the two sources deviate in a range of milliseconds. We achieve this 100% match rate because the ordering of trades and their price and volume match to 100% for all trades where the Deutsche Boerse data allows to identify the true trade initiator.

Results of the trade classification are given in the last row of table 4-2 and in detail in the last four columns of table 4-3 for each of the four trade-by-trade classification algorithms per stock. The two most complex algorithms, EMO and CLNV, achieve the best results with above 90% accuracy overall. The performance of the simpler tick rule varies more between stocks and achieves a lower accuracy of 82%. So, is an accuracy of 90% “enough”? We think yes, it is sufficient for the task of serving as a benchmark to a heuristic classification procedure and for testing the robustness of VPIN.

4.2.4 Evaluation of bulk volume classification and VPIN

The procedure to classify trades not trade by trade, but instead heuristically by approximating the share of buys and sells from aggregated volume and returns, was first introduced by Easley et al. (2012a). In their comparison of traditional tick rule and new bulk classification, the tick rule identifies 86.43%, 67.18%, and 78.95% of the volume of E-mini S&P 500 Futures, WTI Crude Oil Futures, and Gold Futures correctly. If trades are aggregated into “bars” of several seconds up to one-minute length, performance of both the tick rule and BVC increases to ranges of 70% to 98% (tick) and 93% (BVC). Larger bars lead to higher accuracies as more classifications errors offset each other. Overall, however, the tick rule still outperforms BVC in terms of pure classification accuracy, a point analyzed in detail by Andersen and Bondarenko (2015). The major argument brought forward in favor of bulk classification is not raw buy-sell classification accuracy, though.

Table 4-3: Evaluation of trade classification algorithms with true trade initiator

This table presents results of an evaluation of the trade-by-trade classification algorithms based on a proprietary data sample provided by Deutsche Boerse specifically for this paper. The data covers a random sample of 10 stocks of our original sample and 10 trading days in December 2012 from 12/3/2012 to 12/14/2012. The second column indicates the number of trades in the data sourced from ThomsonReuters for the respective stocks and time period, where trading data for the main empirical part of this chapter comes from. Columns 3 and 4 list the trade count from Deutsche Boerse data. Column 5 counts the number of trades that can be matched between the two sources. Columns 6 to 9 present the share of trades correctly identified as buy or sell by the four trade-by-trade classification algorithms.

		Deutsche Boerse		**Matching**	**Classified correctly (buys+sells)**			
Security	**Thomson-Reuters**	All trades	with trade initiator	**Sample w/ trade initiator**	**Tick**	**LR**	**EMO**	**CLNV**
ADSGn	25,738	25,576	23,195	23,195	81.9%	90.2%	91.0%	91.0%
DB1Gn	25,918	25,660	23,415	23,415	82.3%	91.7%	91.6%	92.2%
DBKGn	100,559	98,488	85,443	85,443	85.0%	89.8%	91.1%	91.1%
FMEG	26,762	26,557	23,890	23,890	80.8%	89.9%	90.3%	90.4%
HNKG	30,231	30,077	27,347	27,347	80.7%	89.2%	89.4%	89.6%
IFXGn	38,530	38,426	33,349	33,349	83.0%	87.9%	89.5%	89.5%
LING	18,635	18,327	16,781	16,781	75.6%	87.9%	88.8%	88.8%
SAPG	47,472	46,703	42,186	42,186	78.8%	88.1%	89.0%	89.0%
SDFGn	34,510	34,500	31,166	31,166	84.0%	92.0%	92.6%	92.9%
TKAG	52,061	49,427	41,476	41,476	80.9%	89.0%	89.8%	89.9%
Total	400,416	393,741	348,248	348,248	82.0%	89.6%	90.4%	90.5%

Instead, it is the attempt to capture the behavior of informed traders, who supposedly do not trade aggressively on their information with market orders anymore but instead use a more subtle approach by creating price pressure through sliced limit orders and quick order cancellations (Easley et al., 2012a). By incorporating the price change directly, BVC is intended to also capture this aggressive, limit order-driven price pressure. Consequently, whether to prefer BVC over the tick rule depends on one's belief about informed trader's actual trading behavior. Easley et al. (2012a) argue in the direction favoring BVC, but only provide limited empirical evidence, least for any other markets than the ones studied in their paper. Hence, whether VPIN produces consistent results with either approach remains an important question.

Using signed NASDAQ OMX data from 2005 and 2011, Chakrabarty et al. (2012b) conduct a similar analysis of BVC, where, again, the overall performance of the tick rule is better than BVC on the aggregated bulk level, both for raw trade classification and order imbalances. In addition, VPIN estimates of the bulk tick classification are more strongly correlated to "actual" VPIN values based on the signed trading data. BVC is more successful in classifying large and more frequently traded stocks, whereas the tick rule suffers from an increased trading frequency in volume bars. Inconsistent with Easley et al. (2012a) is a result in Chakrabarty et al. (2012b) indicating the bulk tick rule to be a better indicator of order imbalance for all volume and time bars. The divergence of their conclusions could arise from the different methods used to estimate the accuracy of order flow imbalance. Easley et al. (2012a) use the correlation to the high-low spread, Chakrabarty et al. (2012b) use actual order imbalance. Overall, Chakrabarty et al. (2012b) focus on answering “who is right” and “who is faster” in classifying trades and less on the implications for calculating and applying VPIN, as we do in the current study. Further, Chakrabarty et al. (2012b) consider only the tick rule, but not more advanced trade classification algorithms as we do in this chapter.

Andersen and Bondarenko (2014b) challenge VPIN on its designated “home turf” data, that is, high frequency futures trading in the E-mini S&P 500 and find that the incremental predictive power of VPIN compared to existing measures is not existent or negligible, but definitely not significant. Further, they cannot confirm VPIN’s proclaimed detection of the “Flash Crash”, question the validity of VPIN’s link to theoretical microstructure work in Easley et al. (1996a) and also provide evidence in a separate study that transaction based classification is more accurate than BVC (Andersen & Bondarenko, 2015). Not surprisingly, these critical findings triggered a number of published replies and replies to replies, from the original authors as well as other scholars, with the tone of the argument steadily rising (Andersen & Bondarenko, 2013, 2014a; Easley et al., 2014; Wu et al., 2013b).

Further empirical evidence on VPIN originates from the computer science domain. Wu, Bethel, Leinweber, and Ruebel (2012) initially motivate their research as demonstration of state-of-the-art supercomputing and data management technologies applied to market surveillance. Their second publication still puts most emphasis on the technical challenges and computational cost involved in computing VPIN on high frequency trading data, but also analyzes VPIN’s robustness in a brute-force manner by computing results for 16,000 different parameter settings (Wu, Bethel, Gu, Leinweber, & Ruebel, 2013a). Apart from the discussion in Andersen and Bondarenko (2013) we find two points worth highlighting. Firstly, in the desirable approach to test VPIN’s precision, i.e. determining true positives and false positives, their criterion of a true positive simply requires another metric, the maximum intermediate return, to be larger than average after VPIN reaches its “alert” threshold. But simply requiring the ability to predict a “larger than average” event, i.e. something happening roughly 50% of the time, for whatever toxicity/volatility metric, is a very weak criterion, given VPIN’s aspiration to detect events like the flash crash, which, fortunately, happen to happen few times a year, if at all. Secondly, the authors note that already the choice of how to determine the price at the end of a bucket does affect VPIN results heavily: "it is somewhat surprising that computing prices differently can affect the VPIN events detected" (Wu et al., 2013a). Given the cited controversy around VPIN, this study intends to extend the existing evidence in several dimensions as noted earlier and to guide researchers in the application of VPIN for future research.

4.3 Research design

The following paragraphs describe how we will evaluate the different approaches to trade classification with respect to their application in computing VPIN. We start with briefly explaining the VPIN model and its origin and then describe the hypothesis and test procedures for analyzing VPIN's sensitivity to the choice of trade classification algorithm.

4.3.1 The Volume Synchronized Probability of Informed Trading (VPIN)

The VPIN model described in this section (Easley et al., 2012b) builds on a microstructure trading model initially developed by Easley and O'Hara (1987). The model's key parameters are the probability of an information event α and the arrival rates for two types of traders: μ for the informed traders, who are able to observe information events and their direction, and ε for the arrival of uninformed traders. The probability of informed trading (PIN), which was initially introduced in Easley et al. (1996a), is then defined as the ratio of informed trades relative to all trades:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s} \qquad (4\text{-}2)$$

These parameters can be estimated with a maximum likelihood estimation that requires only the number of buys and sells of each trading day as input. In recent high frequency market environments, however, the estimation of the maximum likelihood function becomes difficult or impossible. VPIN addresses several shortcomings of the PIN model to make it applicable to today's trading activities. Three key steps lead from PIN to VPIN. Firstly, VPIN uses a broader definition of information than the PIN model. Information events are assumed to occur frequently during a day and to differ in their relevance. The arrival of relevant information triggers the arrival of trades. Therefore, a burst in volume should correlate with the degree to which information is existent and relevant. These assumptions make the authors decide to switch from clock time to volume time, the major difference to the previous PIN model. By using volume time, they try to capture the amount of information that underlies bursts in trading volume. Sampling the input for the model's parameters, such as arrival rates for informed and uninformed traders, by volume time implies that times with high trading activity are considered in a "higher resolution" as times with less trading activity.

In a second step, the mathematical foundation for VPIN is laid in an approximation of the VPIN formula based on the dynamic version of the '96 PIN model that uses time varying arrival rates of trades (Easley et al., 2008). The authors show that in a fraction τ of the total trading volume, the product of the arrival rate of information times the arrival rate of informed traders, $\alpha\mu$, can be approximated as the expected absolute order imbalance $E(OI_\tau)$:

$$E(OI_\tau) = E[|V_\tau^B - V_\tau^S|] \approx \alpha\mu \tag{4-3}$$

Similarly, the joint arrival rate of all traders, $\alpha\mu + \varepsilon_b + \varepsilon_s$, where the uninformed trader arrival rate ε is split up for in ε_b for buyers and ε_s for sellers, can be approximated by the sum of the expected total number of trades $E[V_\tau^B + V_\tau^S]$:

$$E[V_\tau^B + V_\tau^S] \approx \alpha\mu + \varepsilon_b + \varepsilon_s \tag{4-4}$$

Joining these two approximations with the original definition of PIN in equation (4-2), the VPIN is defined as follows:

$$VPIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s} \approx \frac{E[|V_\tau^B - V_\tau^S|]}{E[V_\tau^B + V_\tau^S]}. \tag{4-5}$$

In effect, all trades are partitioned into buckets τ with a constant amount of volume V (but different time spans) in order to "mimic the arrival to the market of news of comparable relevance" (Easley et al., 2012b, p. 1465). The authors illustrate the computation of VPIN by setting the average daily number of buckets to 50. This corresponds to V equaling a fiftieth of the average daily volume. Then, with rolling window length $= 50$, VPIN can be written as:

$$VPIN = \frac{\sum_{\tau=1}^{n} |V_\tau^B - V_\tau^S|}{nV}. \tag{4-6}$$

The order imbalances $\sum_{\tau=1}^{n} |V_\tau^B - V_\tau^S|$ are determined by filling each bucket with the respective buy and sell volume. As long as the aggregated volume in a time or volume bar is insufficient to fill a bucket, the volume of the subsequent bar is used. The following bucket contains the excess volume. Since the very last bucket is always incomplete or empty, it is neglected in VPIN calculations. VPIN is computed in a rolling process, i.e. it is the moving average of the order imbalance over the last n preceding buckets. With the parameters chosen by Easley et al. (2012b), equation (4-6) corresponds to a daily VPIN[26]. Table 4-4 extends the example from table 4-1 to illustrate this process of volume bucketing. The remaining volume of time bar 1 is smaller than required to fill bucket 2, therefore 6435 units of time bar 2 are needed to complete the bucket. The excess of time bar 2 is then given to bucket 3, together with the whole volume of time bar 3. At last, the buy and sell volume is computed with the corresponding cumulative distribution function (CDF), similar to equation (4-1). This heuristic estimation of the buy and sell volume is the third key difference to the previous PIN'96 model. The CDF is calculated with the price changes and the standard deviation received from table 4-1. The advantage of using volume time is especially apparent in the process of updating VPIN, where the last bucket is dropped and the next bucket is added. This allows VPIN to be updated at a speed similar to the arrival of information, given that each bucket is likely to hold a homogenous amount of information.

Table 4-4: Volume bucketing of classified timebars

This table illustrates how the time bars resulting out of table 4-1 would fill three buckets, given a bucket size of V = 15000. Timebar 1 is split up between bucket 1 and 2, as only 1394 trades are required to fill bucket 1 up to the maximum volume of 15,000. A similar split of a timebar is made for the volume of bar 2 between buckets 2 and 3. Normally, more than two timebars will be required to fill a bucket.

Timebar i	Start	End	V_i^{B+S}	CDF	V_i^B	V_i^S	Bucket
		…more trades…					
1	09:02:00.001	09:03:00.000	1394	≈ 0	0	1394	1
			15000				
1	09:02:00.001	09:03:00.000	8565	≈ 0	0	8565	2
2	09:03:00.001	09:04:00.000	6435	0.1359	875	5561	2
			15000				
2	09:03:00.001	09:04:00.000	760	0.1359	103	657	3
3	09:04:00.001	09:05:00.000	14240	0.9606	13679	561	3
			15000				

4.3.2 Evaluation of the sensitivity of VPIN

The purpose of this study is to evaluate if, how and under what circumstances the choice of trade classification algorithm influences the computation and application of VPIN.

[26] Easley et al. (2012b) provide a more detailed description of the computation of the VPIN metric in the appendix to their paper.

Although the intention of both approaches is equivalent, they may not show identical results. Important is whether the differences in the classification influence higher-order, aggregate metrics that are used for financial research. The question is to what degree the differences detected in raw trade classification create biases in more sophisticated, aggregated measurements like VPIN and whether the differences will vanish in the aggregation of data. Our approach described in the following paragraphs follows the idea to run step-by-step, bottom-up from trade classification itself, to order imbalance, to actual VPIN calculations and finally to toxic periods, an application of VPIN.

4.3.2.1 Comparison of the classification of raw trading data

We let all algorithms classify the 34 million trades of our XETRA data for 2012 to prepare the comparison of trade-by-trade and bulk classification. The quote rule is omitted, since it leaves midpoint trades unclassified. The BVC and trade-by-trade algorithms cannot be directly compared given the different granularity of their results. We therefore sign the volume of a trade as buyer or seller initiated based on the result of the trade-by-trade classification and then sum the buyer and seller-signed volume over one minute time bars ($\text{Bulk}_{\text{Trad}}$ hereafter). Thereby we compute results on the same level of granularity as the BVC. A first simple step is then the comparison of the share of volume classified as buy (or sell). Second, we calculate the correlation of the buy classification rate between BVC and trade-by-trade algorithms to check the degree to which the results align or not.

4.3.2.2 Application 1: order flow imbalance

Order flow imbalance is the first step up the complexity hierarchy of measures that require trade-initiator signed data as input. Order imbalance has been shown to be correlated with or be a reliable predictor of informed trading (Easley et al., 2012a) and is the main input for the computation of the VPIN. We calculate two different metrics (see also Chakrabarty et al., 2012b) to compare the order imbalances based on trade-by-trade and bulk classification algorithms. Both metrics employ the common definition of order imbalance, i.e. the signed difference between buy and sell volume:

$$OI_{ji} = B_{ji} - S_{ji} \tag{4-7}$$

We index stocks with the letter j and time bars with the letter i. B_{ji} is the buy volume for stock j in time bar i and S_{ji} the corresponding sell volume. The volume-adjusted imbalance match over k time bars is then defined as:

$$VaOI_j = 1 - \frac{1}{2k}\sum_{i=1}^{k}\frac{\left|Trad(OI_{ji}) - BVC(OI_{ji})\right|}{V_{ji}} \tag{4-8}$$

where V_{ji} is the volume of bar i, $Trad(OI_{ji})$ is the order imbalance in bar i calculated using trade volume signed by trade-by-trade algorithms, and $BVC(OI_{ji})$ is the order imbalance resulting from bulk classification. We define imbalance direction match for stock *j* as:

$$DirOI_j = \frac{1}{k}\sum_{i=1}^{k}\frac{\left|sgn\left(Trad(OI_{ji})\right) + sgn\left(BVC(OI_{ji})\right)\right|}{2} \tag{4-9}$$

The direction match gives the percentage of bars where the direction of the order imbalance based on trade-by-trade algorithms matches the direction of order imbalance based on trade volume signed with bulk classification. While the definitions above are given for order imbalances based on time bars *i*, both metrics will also be calculated using order imbalances calculated over buckets τ. As buckets span several time bars, errors are more likely to offset each other and hence higher consensus between the two approaches can be expected.

4.3.2.3 Application 2: VPIN estimation

The third level up from classified trading data is the VPIN variable itself. VPIN is a specific aggregation of order imbalances over several buckets of volume. The bucket size needs to be set to a fraction of the average daily trading volume. This offers at least two options to calculate VPIN during the observation period, depending on whether we use a constant bucket size for the whole year or determine the bucket size for every month based on the monthly average trading volume. In both cases, the average number of daily buckets equals 50, as does the sample length of VPIN. We calculate VPIN with both a constant yearly bucket size and a varying monthly bucket size as a robustness check. The results of the VPIN computation are analyzed using cross sectional averages, volatility, correlation with trading volume and correlation between VPIN based on bulk volume classification, VPIN(BVC), and VPIN based on trade-by-trade volume classification, VPIN($\text{Bulk}_{\text{Trad}}$).

4.3.2.4 Application 3: toxic periods

The VPIN metric was developed to detect the degree of toxic order flow that results from asymmetric levels of information among traders. What certain value of VPIN should alert traders is inevitably an arbitrary choice. We follow the common definition used in recent publications regarding VPIN: "A toxic period begins when the empirical cumulative distribution function (CDF) of VPIN reaches or crosses the 0.9 percentile and ends when the CDF falls below the 0.8 percentile" (Chakrabarty et al., 2012b, p. 25). Accordingly, a toxic period starts with a bucket where the VPIN calculated with the current and the last 49 buckets crosses the 0.9 percentile. The 49 buckets before the current all contribute to the toxicity measured at the time of the current bucket number 50, but we still call only bucket 50 "toxic", as VPIN reaches its critical level only with the inclusion of this bucket. All consecutive buckets where the VPIN estimated from the rolling last 50 buckets does not fall below the 0.8 percentile form the "toxic period".

The toxic periods are the highest-order metric in terms of aggregation and complexity that we use in this chapter. We conduct two tests to compare toxic periods calculated on results from bulk classification to toxic periods based on trade-by-trade classification results. We identify all toxic periods with VPIN(BVC) and test how many toxic buckets within this period overlap with toxic buckets of toxic periods identified by VPIN($\text{Bulk}_{\text{Trad}}$). The first test, toxic *period* match, requires an overlap of only one toxic bucket to evaluate the two methods as equal. The second test, toxic *bucket* match, is the percentage of toxic buckets in a toxic period of VPIN(BVC) that are also regarded as toxic by VPIN (BulkTrad).

4.3.2.5 Determinants of VPIN differences

The computation of VPIN with two different strategies for trade classification may not yield the exact same results. To not only state a problem but also guide towards a solution, we want to give an indication in what circumstances either classification scheme is more applicable or when the bias induced by the choice of algorithm is especially large. Therefore, we design a regression model to test the influence of trading and stock characteristics on the difference in VPIN levels. We run the model in volume time. The dependent variable $\Delta VPIN_{j\tau}$ is the difference between every VPIN value calculated with BVC minus VPIN calculated with $\text{Bulk}_{\text{Trad}}$. The index j numbers companies and index τ numbers the VPIN estimations per company (or buckets, starting from bucket 50 for the first VPIN calculation onwards).

We hypothesise different levels of liquidity and return volatility to affect the differences because trade-by-trade algorithms can run into more errors in high frequency trading and the bulk classification is defined based on returns. In doing so, we want to distinguish a "size effect" in terms of both trading volume and returns from a "volatility effect". In other words, do the methods differ because of different levels of trading activity or do they differ in periods of abnormal trading activity - or both? We intend to measure the size effect with the variables $Trades_{j\tau}$ and $ReturnVola_{j\tau}$, representing the total number of trades and the volatility of the returns over the time bars within each VPIN. Whether current trading conditions are abnormal is measured by the variable $VolumeVola_{j\tau}$, representing the volatility of traded volume over time bars; the stock return, split into two variables $posReturn_{j\tau}$ and $negReturn_{j\tau}$ to capture rising and falling market conditions separately; and the variable $Timebars_{j\tau}$ which contains the number of time bars during each VPIN estimation as a standardized measure of abnormal volume (more time bars required to fill the VPIN buckets mean less than average trading volume during that period). In other words, the number of time bars indicates how far a VPIN estimation is spread across clock time. Variations on stock level, which are likely present, are taken into account with fixed effects per stock. Trading volume itself cannot be chosen as explanatory variable given that it is, by construction, constant per stock and VPIN estimation. This results in the following model:

$$\begin{aligned}\Delta VPIN_{j\tau} &= \beta_0 + \beta_1 Trades_{j\tau} + \beta_2 ReturnVola_{j\tau} + \beta_3 VolumeVola_{j\tau} + \\ &\beta_4 posReturn_{j\tau} + \beta_5 negReturn_{j\tau} + \beta_6 Timebars_{j\tau} + \alpha_j + u_{j\tau}\end{aligned} \tag{4-10}$$

The independent variables relate to the last 50 buckets over which the two VPINs for the dependent variable are calculated. Hence, on a day with roughly average trading volume, all variables span roughly one trading day. We run four different versions of the model. The first differentiation stems from using the both the absolute and the signed difference as dependent variable. While the absolute difference tells us which factors generally lead to divergent VPIN measures, the signed difference allows concluding which of those factors may tend to offset each other. The second differentiation is due to VPIN's calculation over a rolling window of 50 buckets. In one model, we use only every 50th VPIN observation to avoid incorporating autocorrelation by design. In a second approach, we use every VPIN observation but correct for autocorrelation with Newey-West standard errors for 49 lags.

4.3.2.6 Predictive power of VPIN

The analysis in the preceding sections will tell us whether VPIN is sensitive to the choice of trade classification or not and in what trading environments this sensitivity is especially pronounced. We also know from section 2.3 that the trade-by-trade classification algorithms perform very well in classifying single trades. At the end of the day, however, Easley et al. (2012b) argue that VPIN calculated with bulk volume classification captures something different than precise "theoretical" order imbalance. Especially the fast order entries and immediate cancellations pursued by high frequency traders, it is argued, render the identification of the actual aggressor side ambiguous. Bulk classification is claimed to be better at capturing *actual* informed trading or toxicity of order flow. To address this valid concern, we extend our analysis in two ways. First, in the spirit of the "flash crash" analyzed in the initial studies on VPIN, we evaluate VPIN's ability to detect a severe crash of a single security. Second, we test VPIN's predictive power for future volatility. Both Easley et al. (2012b) and Andersen and Bondarenko (2015) evaluate VPIN's ability to predict future volatility or future price movements. We choose an approach similar to Andersen and Bondarenko (2015) to test VPIN's *incremental* predictive power for one-day volatility with the following fixed effects regression model:

$$RV_{jt} \;\; = \beta_0 + \beta_1 RV_{jt-lag} + \beta_2 VPIN_{jt-lag}^{Algo} + \alpha_j + u_{jt} \tag{4-11}$$

Where index t denotes intervals of 10 minutes length; lag is 51 to lag observations by exactly one trading day (one trading day spans 8,5 hours, i.e. 51 10-minute intervals). RV_{jt} is the realized volatility for stock *j* at interval *t*, calculated as average of the absolute log returns over the last trading day's 10-minute-intervals, i.e. 34 lags. $VPIN_{jt-lag}^{Algo}$ is the latest complete VPIN calculation available at time t for stock *j*, based on either BVC or one of the trade-by-trade algorithms. We run regressions for 5 and 15 minutes intervals and respective lags of 34 and 102 as robustness check.

4.4 Data

This study uses trade and quote data from XETRA, the electronic trading system from Deutsche Boerse, by far the leading stock exchange in Germany[27]. XETRA offers an anonymous order-driven market model with an electronic limit order book. Orders are executed by price/time priority. Trading starts with an opening auction, followed by continuous trading, which is interrupted by an intraday auction at 1.00 pm and ends with a closing auction. We chose about a year of trading data for our analysis, that is, the 234 trading days from January 2, 2012 to November 27, 2012. We limit the analysis to those stocks that are listed in the index DAX, which is the leading stock index in Germany and

[27] See Deutsche Boerse's XETRA website: http://xetra.com/xetra/dispatch/en/kir/navigation/xetra/300_trading_clearing/100_trading_platforms/100_xetra.

includes the 30 largest and most traded stocks[28]. The raw data contains roughly 50 million quotes and 34 million trades, representing 29 billion shares traded.

Equity trading in the DAX on XETRA is on par, if not even more frequent than trading at the other major stock exchange in Europe, the London Stock Exchange (LSE). The average daily traded volume of the 30 stocks in the DAX in the first quarter of 2014 is 3.2 - 3.8 billion Euro, which is roughly 80% of the daily traded volume of the 100 stocks of the FTSE 100 traded on LSE[29]. The total turnover of the 30 DAX stocks alone in 2012 was 803 billion EUR, which is equal to 55% of the traded volume of all stocks on the LSE in the same year. A unique advantage of our dataset is its lower degree of fragmentation compared to other major marketplaces. Deutsche Boerse captures 70% of global trading in the 30 DAX stocks. In contrast, only 35% and 29% of trading in the stocks of the Dow Jones or S&P 500, respectively, runs via the leading exchange NYSE. Concentration for stocks on NASDAQ, the data used in Chakrabarty et al. (2012b), is also below 50% . Trading in the FTSE 100 is slightly less fragmented, but still lower, with 62% of traded volume being executed at the LSE[30]. Thus, the dataset employed in this chapter is both high frequency and comes closest to covering the full order stream in the observed security, hence providing perfect conditions to properly evaluate an order flow toxicity metric like VPIN.

The data is sourced from ThomsonReuters TickHistory. Trades, quotes and auctions are reported with millisecond timestamps. We exclude all trades and quotes resulting from opening, closing and intraday auctions. In a paper published recently, Holden and Jacobsen (2013) demonstrate how a lack of precision and diligence in the processing of intraday data can introduce a heavy bias in the results. We draw on the excellent list in Holden and Jacobsen (2013) of key checks to raw trading data, all of which we validate on our data: There are no negative trade prices or quotes. There are no crossed or blocked spreads anytime. The first column of table 4-5 shows the distribution of the trades based on their position relative to the spread. 69.3% are at the spread, 13.7% outside the spread, and 17% inside the spread. The relatively large share of trades inside the spread is surprising at first. A share of these trades is due to three types of special order execution in the XETRA trading model that allows trades to be cleared inside the spread, as our discussions with

[28] The stocks Lanxess and Continental are dropped from the sample as they joined the DAX only in late 2012.

[29] http://www.londonstockexchange.com/exchange/statistics/share-of-trading/lit-figures/FTSE100.html

[30] Fragmentation numbers from http://fragmentation.fidessa.com/

experts from Deutsche Boerse confirm: hidden orders, “Xetra Midpoint” and “Xetra BEST”. The validation of our data with proprietary data retrieved directly from Deutsche Boerse as described in section 2.3 shows that we can match 100% of trades from Deutsche Boerse to our sample. This is a very strong indication that a bias in results stemming from incorrectly processed data can be ruled out.

Table 4-5: Distribution of trades relative to spread before and after trade classification
This table reports the location of trades relative to the spread and the share of those trades classified as buy, for each of the traditional trade classification algorithms. The sample contains 34,335,259 trades in total.

Location	Share of trades	Classified as buy [%] Tick	Re. Tick	Quote	LR	EMO	CLNV
Above Ask	6.7	92.8	51.8	100	100	92.8	92.8
At Ask	34.8	79.1	48.6	100	100	100	100
Below Ask	6.4	65.4	47.0	100	100	65.4	76.0
Midpoint	4.3	50.0	49.7	0	50.0	50.0	50.0
Above Bid	6.3	34.2	53.0	0	0	34.2	23.7
At Bid	34.5	21.1	51.4	0	0	0	0
Below Bid	7.0	7.3	48.8	0	0	7.3	7.3
Total	100	50.0	50.0	47.9	50.1	50.1	50.1

4.5 Results

This section discusses empirical results by ascending in complexity and aggregation from raw trade classification up to toxic periods. The last subsections use results from regression analysis to identify market conditions where VPIN is most sensitive to the choice of classification algorithm and also evaluate VPIN’s predictive power towards future volatility.

4.5.1 Trade classification with trade-by-trade algorithms

The first level of our comparison is the pure classification of trades as buyer or seller initiated, after calibration of the trade-by-trade algorithms as described in the section on pre-processing. The buy classification rate of all trade-by-trade algorithms is shown in table 4-5, where results are split by the trade’s location relative to the spread. The buy classification rate deviates closely around 50% on an aggregated yearly level, which is the expected behavior. Differences between the algorithms become apparent when the classification rates are compared by the trade’s location relative to the quotes. The buy classification rate of the tick rule reaches its highest value above the ask and declines until below the bid. Around 20% of trades at the ask are classified as sells, which indicates a high number of quote movements contrary to the trade direction. The quote rule classifies all trades above the midpoint of the spread as buys and LR, EMO, and CLNV yield combinations of the results of tick and quote rule. A more detailed analysis of the classification results by single stocks does not show any differences or patterns by stock characteristics.

4.5.2 Comparison of trade-by-trade and bulk classification

Results for the bulk classification cannot be presented in a partition of trades relative to their location in the spread. Instead, granularity is increased on the stock level in table 4-6. We use only one-minute time bars, as in the original formulation. The buy classification rates of the trade-by-trade algorithms and BVC are all close to 50%. BVC yields a lower buy percentage with 49.55% and differs from $Bulk_{Tick}$ by 0.45 percentage points. The other three algorithms report similar results, ranging from 50.22% to 50.36%. A specific pattern regarding trades or volume is not observable.

The more sensitive part is the right-hand side of table 4-6. The correlations of the time bar series of the buy classification rates from BVC and trade-by-trade classification range from 44% to a maximum of 70% for Beiersdorf (BEIG). The correlation of trade-by-trade and bulk classification over all stocks reaches 64% for the Tick rule, 48% for the Lee-Ready-algorithm, 53% for the EMO-algorithm and 52% for the CLNV-algorithm. These levels are surprisingly low given that these algorithms achieve accuracies of up to 90% as shown in the literature review and the performance evaluation in this chapter. What does bulk classification compute that its results are so different, as indicated by the low levels of correlation? The trade-by-trade algorithm deemed least accurate among the four algorithms evaluated, the tick rule, has a substantially higher correlation with BVC than the other, more accurate algorithms EMO and CLNV. This is probably due to the intrinsic design of the tick rule. In rising markets, it is likely to classify more trades as buy because every trade going up in price is classified as a buy. This is not the case for EMO and CLNV algorithms, which also take into account quote data. To better benchmark the correlation values we also compute correlations for the results of the trade-by-trade algorithms among themselves; results are presented in table 4-7. The correlation between the sophisticated trade-by-trade algorithms LR, EMO and CLNV is quite high, with rates well above 90%. Consequently, the bulk classification evaluates trades substantially different.

In summary, the correlation between the results of bulk and trade-by-trade algorithms is very low. Results of the simple tick rule are closer to BVC than the more sophisticated trade-by-trade algorithms. Given that the purpose of these algorithms is the same, the low correlations of the time bar series support our concern that the heuristic approach of bulk classification is measuring something different than it is supposed to. The following sections analyze how these differences translate into different results for the calculation of order imbalance, VPIN and toxic periods.

Table 4-6: Trade classification results and correlations per stock

The first 5 columns of this table report the share of volume classified as buys for BVC and the four trade-by-trade classification algorithms $Bulk_{Tick}$, $Bulk_{LR}$, $Bulk_{EMO}$, and $Bulk_{CLNV}$. The right-hand four columns report the correlations of the classified buy volume time bar series between BVC and each of the four algorithms. Results are listed per stock. Row "Total" shows the volume-weighted average. The total values are averaged by applying Fisher's z transformation on the stock values with the number of time bars as sample lengths.

	Classified as buy volume [%]					**Correlation between classification of BVC and $Bulk_{Trad}$**			
Stock	BVC	$Bulk_{Tick}$	$Bulk_{LR}$	$Bulk_{EMO}$	$Bulk_{CLNV}$	$Bulk_{Tick}$	$Bulk_{LR}$	$Bulk_{EMO}$	$Bulk_{CLNV}$
Total	49.6	50.0	50.4	50.2	50.2	.643	.484	.530	.519
ADSGn	49.8	49.9	49.7	49.7	49.7	.672	.509	.559	.547
FMEG	49.7	49.9	49.8	49.9	49.9	.697	.538	.584	.574
ALVG	50.2	50.3	50.4	50.4	50.4	.604	.452	.506	.492
FREG	49.9	50.4	50.7	50.6	50.6	.681	.523	.576	.556
BASFn	49.8	50.1	50.3	50.2	50.2	.614	.460	.506	.499
HEIG	49.6	49.5	49.4	49.4	49.4	.661	.515	.566	.546
BMWG	49.9	50.2	50.2	50.3	50.2	.621	.461	.507	.500
HNKG	50.1	49.6	49.3	49.4	49.3	.680	.518	.564	.552
BAYGn	50.1	50.3	50.1	50.2	50.2	.632	.475	.523	.516
IFXGn	49.8	50.3	50.5	50.4	50.4	.640	.492	.543	.528
BEIG	50.5	50.1	50.2	50.3	50.3	.704	.519	.577	.560
SDFGn	49.3	50.1	50.1	50.0	50.1	.653	.501	.545	.532
CBKG	49.3	49.8	50.6	50.3	50.3	.608	.466	.499	.499
TKAG	50.0	50.4	50.2	50.2	50.2	.650	.470	.519	.513
EONGn	49.3	50.3	50.7	50.6	50.7	.642	.477	.510	.505
LING	49.4	49.9	50.5	50.5	50.5	.666	.506	.530	.527
VOWG	49.9	50.2	50.2	50.2	50.2	.636	.481	.512	.509
DAIGn	49.7	49.9	49.9	49.8	49.9	.603	.466	.513	.501
MRCG	50.2	50.2	50.6	50.7	50.6	.682	.508	.565	.543
DBKGn	49.9	49.5	49.4	49.3	49.4	.606	.478	.526	.517
MUVGn	50.3	50.7	50.8	50.8	50.9	.649	.498	.514	.509
DPWGn	50.0	50.1	50.1	50.1	50.1	.663	.488	.532	.526
RWEG	49.7	50.0	50.1	50.1	50.1	.637	.457	.510	.500
DB1Gn	49.5	50.1	49.8	49.7	49.7	.649	.466	.534	.510
SAPG	50.4	50.5	50.4	50.5	50.5	.621	.475	.517	.510
LHAG	50.1	50.4	50.5	50.4	50.4	.657	.501	.549	.534
SIEGn	49.7	50.4	49.0	49.1	49.1	.608	.453	.498	.493
DTEGn	49.5	50.3	50.4	50.4	50.4	.604	.437	.490	.480

Table 4-7: Correlation of trade classification results

This table reports the correlations of the time bar series of the share of volume classified as buy between BVC and the four trade-by-trade algorithms. The values are averaged over all stocks by applying Fisher's z transformation on the stock values using the number of time bars as sample lengths.

	BVC	$Bulk_{Tick}$	$Bulk_{LR}$	$Bulk_{EMO}$	$Bulk_{CLNV}$
BVC	1	.643	.484	.530	.519
$Bulk_{Tick}$		1	.635	.720	.697
$Bulk_{LR}$			1	.931	.953
$Bulk_{EMO}$				1	.985
$Bulk_{CLNV}$					1

4.5.3 Order flow imbalances in time bars and volume buckets

The next level up from raw trading data is the computation of order imbalance; results are shown in Table 4-8. The volume-adjusted order imbalance match between trade-by-trade algorithms and BVC is around 65% for the small bars of one-minute length and increases to close to 90% for the larger buckets. A higher value for the buckets is expected given that they aggregate more volume. The difference between using monthly data or yearly data to determine the buckets is small and there is only a negligible difference between the trade-by-trade algorithms. The direction of the imbalance is equal in around 70% of the cases, except for the tick rule, were bulk and trade-by-trade agree in 75% of the cases.

These two more complex metrics for comparing order imbalance again show that the tick rule is slightly closer to the bulk classification than the other algorithms. The difference is small regarding the volume imbalance and larger regarding direction imbalance. Overall, however, the absolute values are in ranges that one cannot conclude that trade-by-trade and bulk classification measure the exact same phenomena. On the other hand, the higher volume match in buckets and the much higher level of agreement compared to raw trade classification indicates that differences may diminish once we continue to aggregate further to higher-order metrics.

Table 4-8: Match of order imbalance
This table compares order imbalances calculated from BVC results with order imbalances based on classification results of the trade-by-trade algorithms. Two metrics are used for this comparison. Columns two to four report the volume-adjusted imbalance match *VaOI*. The right-hand side of the table reports the direction match *DirOI*. For both metrics, results are presented based on time bars and buckets, whereas buckets are based on monthly or yearly average trading volume. All values are cross-sectional averages of the stock values.

	Volume-adjusted imbalance match			**Direction match**		
	% Time bars	**% Buckets**		**% Time bars**	**% Buckets**	
		Yearly	Monthly		Yearly	Monthly
$Bulk_{Tick}$	.647	.893	.828	.771	.742	.743
$Bulk_{LR}$	.649	.872	.827	.698	.673	.672
$Bulk_{EMO}$	.652	.879	.829	.718	.692	.692
$Bulk_{CLNV}$	.651	.878	.828	.712	.689	.690

4.5.4 VPIN estimation

This section discusses cross-sectional averages and the time-series of VPINs to check correlations between the VPINs computed with either approach of trade classification algorithm. The yearly averages of the VPIN per stock and the cross-sectional averages are reported in table 4-9 for the five different algorithms[31]. The highest cross sectional average value is measured for VPIN($Bulk_{Tick}$) with 26.09%. VPIN(BVC) is slightly lower at 25.76%. Calculating VPIN with the output from the EMO algorithm, $Bulk_{EMO}$, results in the lowest VPIN value of 24.81%. The values are in the ranges of those reported by Abad and Yagüe (2012), but exceed the average VPIN of futures by Easley et al. (2012b) by roughly four percentage points.

A first approach to compare VPINs from different algorithms is the scatter plot in figure 4-2. The VPIN(BVC) is plotted on the horizontal axis and VPIN($Bulk_{Tick}$) as representative of VPIN($Bulk_{Trad}$) on the vertical axis. Three observations can be made. First, VPIN(BVC) and VPIN($Bulk_{Tick}$) do not consistently yield similar results. Second, the deviation does not look random. Instead, for smaller than average values of VPIN(BVC) the difference to VPIN($Bulk_{Trad}$) is negative, and for values larger than the average VPIN(BVC), the difference to VPIN($Bulk_{Trad}$) is positive. In other words, the slope in the graph is a lot steeper than the 45-degree line that one would expect if both approaches to trade classification algorithms would yield roughly the same results. The results of VPIN($Bulk_{Tick}$) spread across an interval that is twice as large as the min-max spread for

[31] We performed the VPIN computation separately for each month and for the whole year. As the results and conclusions are very similar, we discuss the yearly computations only in this section.

values of VPIN(BVC). For example, the VPIN values for Munich Re (MUV) differ by 7.69 percentage points between VPIN($Bulk_{Tick}$) and VPIN(BVC). The standard deviation for VPIN(BVC) is $\sigma = 1.64$, and for VPIN($Bulk_{Trad}$) it is $\sigma = 4.2$. Traditional trade classification algorithms seem to induce a higher cross-sectional variance in VPIN estimates.

Table 4-9: VPIN results per stock

This table reports results of the VPIN computation based on input from BVC and each of the four trade-by-trade algorithms. VPIN values are averaged per single stock. The row "Total" contains the cross sectional average. Yearly data is used for the results in this table.

Stock	BVC	$Bulk_{Tick}$	$Bulk_{LR}$	$Bulk_{EMO}$	$Bulk_{CLNV}$
Total	25.8	26.1	25.2	24.8	24.9
ADSGn	25.7	26.9	26.8	26.2	26.2
ALVG	23.5	21.1	20.1	20.5	20.4
BASFn	23.9	21.9	20.7	20.9	20.9
BAYGn	24.1	23.4	22.1	22.2	22.2
BEIG	28.0	33.2	33.0	32.2	32.4
BMWG	24.1	19.8	18.9	18.5	18.6
CBKG	28.0	27.0	24.8	24.6	24.6
DAIGn	23.1	18.3	18.0	18.0	18.0
DB1Gn	24.5	25.9	26.2	25.4	25.7
DBKGn	24.1	19.6	19.3	19.1	19.2
DPWGn	27.1	33.8	31.0	30.7	30.8
DTEGn	24.4	23.4	22.7	22.4	22.4
EONGn	25.1	24.6	22.1	22.0	22.0
FMEG	27.0	30.7	30.8	30.0	30.2
FREG	26.1	29.0	30.1	29.3	29.5
HEIG	24.3	24.2	25.0	24.3	24.5
HNKG	27.0	28.5	28.5	27.8	28.0
IFXGn	27.6	25.3	25.7	24.9	25.1
LHAG	27.9	29.1	28.2	27.6	27.7
LING	27.5	33.3	29.8	29.7	29.7
MRCG	28.1	31.1	31.8	30.5	30.9
MUVGn	26.8	34.5	30.5	30.5	30.5
RWEG	25.3	23.2	21.5	21.4	21.4
SAPG	25.6	23.8	23.5	23.0	23.1
SDFGn	26.2	26.7	27.1	26.2	26.3
SIEGn	23.2	20.6	19.6	19.6	19.6
TKAG	27.7	26.7	25.4	24.9	24.9
VOWG	25.6	25.1	22.8	22.5	22.5

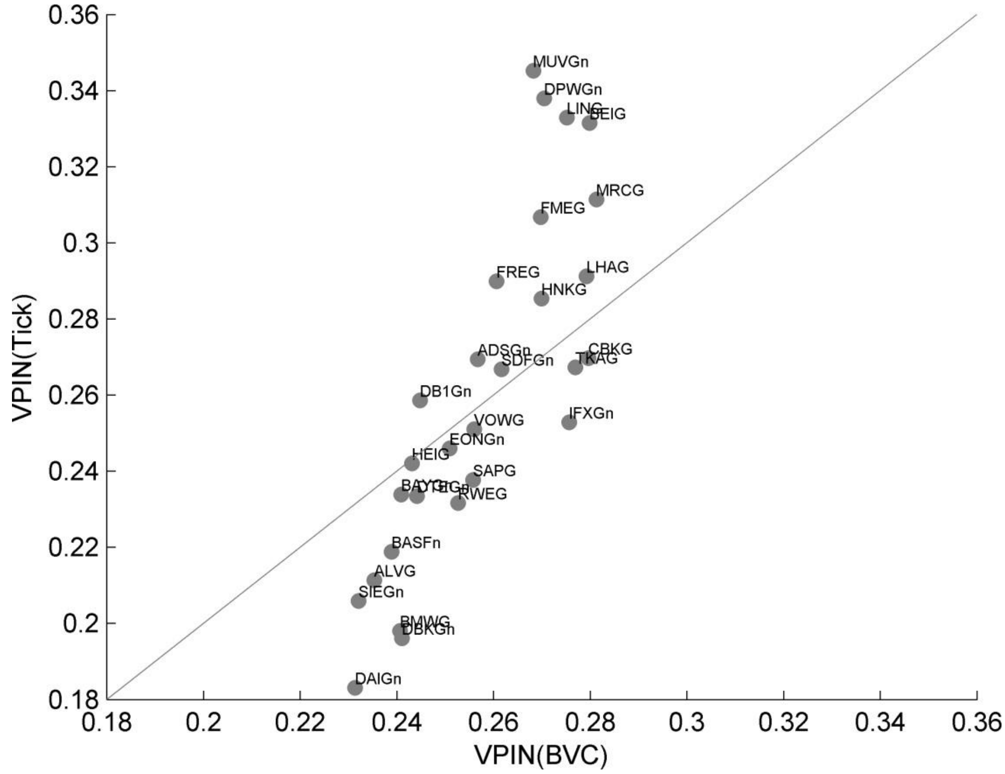

Figure 4-2: Scatter plot of VPIN(BVC) and VPIN(BulkTick)

This figure plots the volatility of VPIN(BVC) on the horizontal axis against the volatility of VPIN(BulkTick) on the vertical axis, per single stock. Average volatilities are σ_BVC= 1.642 and σ_Tick= 4.535.

This results changes if we look at the time-series of VPINs throughout the year. Figure 4-3 illustrates the series of VPIN($\text{Bulk}_{\text{Trad}}$) and VPIN(BVC) of three stocks in separate graphs. We first choose Daimler (DAI) and Munich Re (MUV) because Daimler is the stock with the lowest VPIN average (18.02%) and Munich Re the one with the highest (34.52%). The differences between the two series per stock are clearly visible. VPIN($\text{Bulk}_{\text{CLNV}}$) of Daimler is generally on a lower level and does not exhibit the high VPIN peaks as BVC does. In contrast, VPIN($\text{Bulk}_{\text{Tick}}$) of Munich Re is on a higher level and only a poor match of BVC. Other stocks show similar characteristics of visibly diverging graphs for VPIN calculations based on BVC and trade-by-trade algorithms.

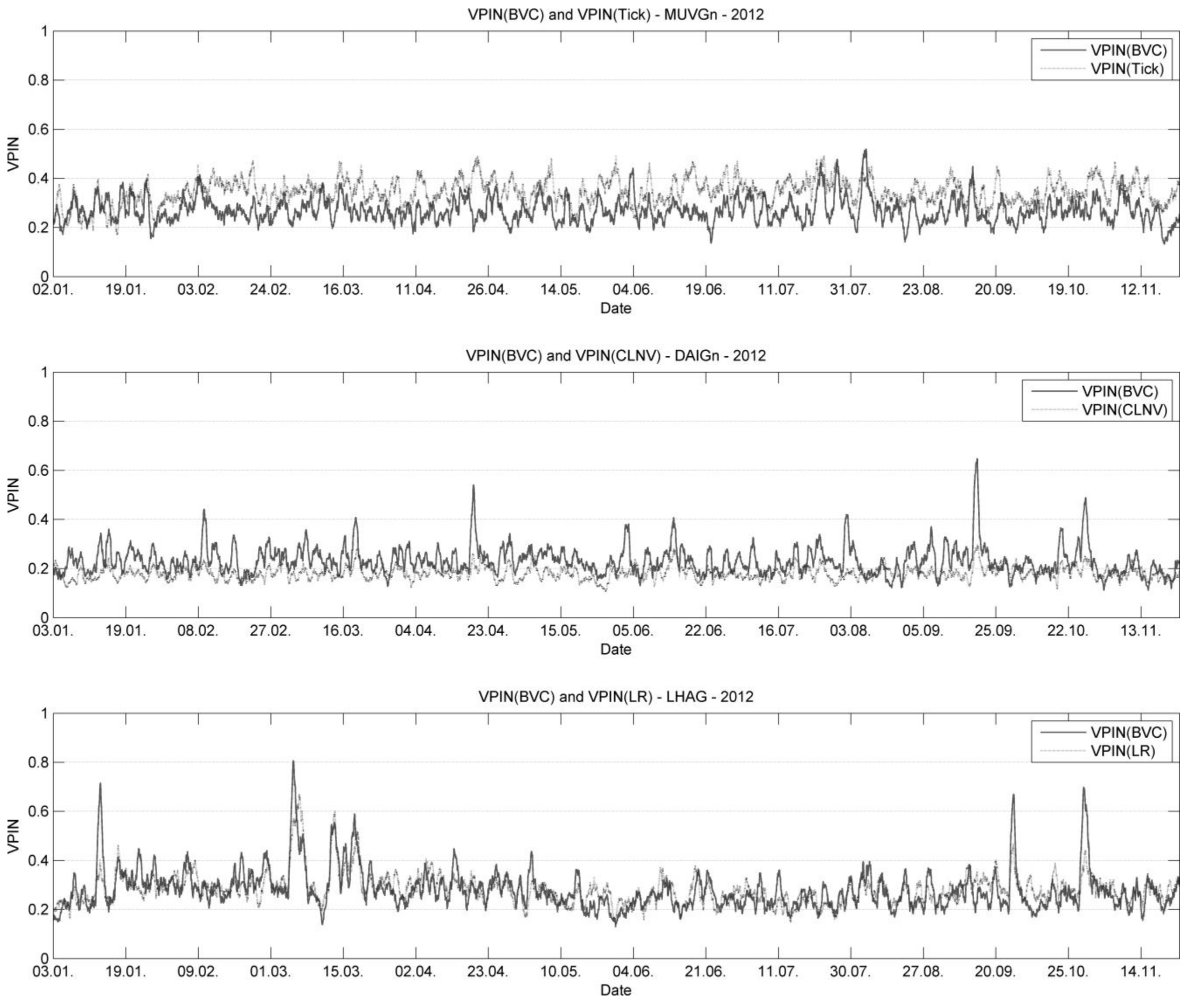

Figure 4-3: Time-series of VPIN for selected stocks, in volume time

This figure illustrates the VPIN time-series based on both trade-by-trade algorithms and BVC. The first graph exhibits the stock with the highest VPIN average (Munich Re - MUVGn), the second one the stock with the lowest VPIN average (Daimler - DAIGn), and the third one the stock with the highest series correlation between VPIN based on BVC and VPIN based on trade-by-trade classification (Lufthansa - LHAG). The VPIN(BVC) time-series are in blue color. The horizontal axis is in volume time scale.

The results in table 4-10 generalize the observations from the line charts. The time-series of VPIN($Bulk_{Trad}$) and VPIN(BVC) are not the same. The correlation between the two is only 40.7% for Munich Re and 51.7% for Daimler. The highest correlation between VPIN($Bulk_{Trad}$) and VPIN(BVC) is measured for Lufthansa, with 73.4%, illustrated in the bottom chart of figure 4-3. Overall, the average correlations are only slightly better than that of Daimler. VPIN($Bulk_{Tick}$) achieves the highest correlation with VPIN(BVC) of 55.3% and VPIN($Bulk_{EMO}$) the lowest match of 53.1%. Compared to the metrics of the previous sections, the correlations of VPINs computed with different trade classification algorithms are on a similarly low level as the correlations of the raw classification results in table 4-6 and table 4-7. The low variance between VPIN($Bulk_{Trad}$) and VPIN(BVC) for the calculation of the order imbalance reported in table 4-8 does not diminish further when calculating VPIN. Instead, the variance rises again.

Table 4-10: Correlations of VPIN results per stock

This table reports the correlations between VPIN(BVC) and VPIN(BulkTrad) per single stock. The row "Total" contains the cross sectional average, calculated by applying Fisher's z transformation with the number of VPIN observations as sample lengths. Yearly data is used for the results in this table.

	$\text{Bulk}_{\text{Tick}}$	Bulk_{LR}	Bulk_{EMO}	$\text{Bulk}_{\text{CLNV}}$
Total	.553	.542	.531	.532
ADSGn	.518	.550	.517	.520
ALVG	.342	.275	.332	.306
BASFn	.516	.409	.433	.432
BAYGn	.471	.447	.445	.437
BEIG	.713	.668	.673	.683
BMWG	.469	.536	.494	.506
CBKG	.520	.586	.570	.564
DAIGn	.448	.540	.504	.517
DB1Gn	.431	.317	.342	.314
DBKGn	.418	.438	.412	.411
DPWGn	.526	.542	.535	.535
DTEGn	.590	.561	.536	.553
EONGn	.585	.552	.538	.532
FMEG	.654	.635	.659	.655
FREG	.546	.516	.520	.520
HEIG	.597	.557	.541	.542
HNKG	.573	.491	.502	.496
IFXGn	.645	.669	.638	.660
LHAG	.721	.734	.702	.714
LING	.633	.600	.588	.592
MRCG	.648	.641	.607	.618
MUVGn	.407	.324	.356	.347
RWEG	.511	.508	.475	.484
SAPG	.629	.692	.656	.658
SDFGn	.637	.622	.614	.609
SIEGn	.457	.414	.412	.398
TKAG	.547	.600	.571	.569
VOWG	.533	.494	.505	.495

We also computed the correlations among the trade-by-trade algorithms themselves to serve as a kind of benchmark to what levels of correlation one would reasonably expect. Results are presented in table 4-11. The VPIN($\text{Bulk}_{\text{Trad}}$) series among themselves correlate to a much higher degree with rates from 79% to 99%. Similar to the correlations of the time bar series, the sophisticated trade-by-trade algorithms show correlations among one another way above 90%. Once again our results indicate that the bulk classification is measuring something different than it is supposed to.

Table 4-11: Correlation of VPIN results

This table reports the correlations of the time bar series between VPIN(BVC) and the VPIN results based on trading data classified using the four trade-by-trade algorithms. The total values are averaged by applying Fisher's z transformation with the number of VPIN observations as sample lengths. Yearly VPIN data is used for this table.

	VPIN(BVC)	**VPIN($Bulk_{Tick}$)**	**VPIN($Bulk_{LR}$)**	**VPIN($Bulk_{EMO}$)**	**VPIN ($Bulk_{CLNV}$)**
VPIN(BVC)	1	.553	.542	.531	.532
VPIN($Bulk_{Tick}$)		1	.790	.830	.820
VPIN($Bulk_{LR}$)			1	.957	.966
VPIN($Bulk_{EMO}$)				1	.994
VPIN($Bulk_{CLNV}$)					1

Table 4-12: Correlations between trading frequency and VPIN

This table reports the correlation between trading frequency and VPIN results. Yearly VPIN data is used for this table.

VPIN(BVC)	**VPIN($Bulk_{Tick}$)**	**VPIN($Bulk_{LR}$)**	**VPIN($Bulk_{EMO}$)**	**VPIN ($Bulk_{CLNV}$)**
-.737	-.882	-.911	-.905	-.907

Consistent with other studies that use PIN or VPIN, the VPIN values per stock are negatively correlated with trading frequency, as results show in table 4-12 (Abad & Yagüe, 2012; Easley et al., 1996a). According to Easley et al. (1996a), frequently traded stocks are less prone to the arrival of informed trading since uninformed traders tend to focus on these stocks, thereby offsetting informed traders and lowering the risk or probability of informed trading. The bar diagram in figure 4-4 orders the stocks in ascending order by their traded volume. Highly liquid stocks, such as Daimler (DAI) and Deutsche Bank (DBK), yield low VPIN values, whereas less frequently traded stocks, such as Beiersdorf (BEI) and Merck (MRC), yield high probabilities of informed trading. The relation is more pronounced for VPIN($Bulk_{Trad}$) than for VPIN(BVC), which is what can also be seen in the scatter plot discussed earlier. The bar diagram again visualizes the large differences in the average VPIN values based on the choice of classification algorithm.

The averaged correlations provide a rough indicator of the consensus between the VPIN series, given the high variation on stock level. However, differences in VPIN computations such as the reported 7.69 percentage points for Munich Re and single stock correlations around 30% are a cause for concern for the application of the VPIN methods. The next section tests how the large differences affect the detection of toxic periods, one of the proposed applications of VPIN.

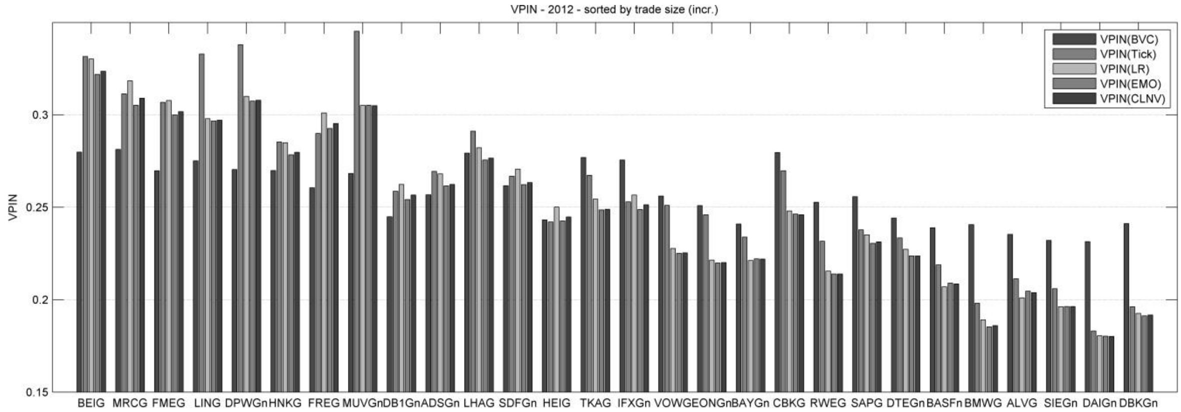

Figure 4-4: VPIN sorted by trading frequency

This figure shows the relation of trading frequency and VPIN values. The stocks are sorted on the horizontal axis from left to right by increasing trading frequency. The vertical axis denotes the average VPIN values for the different algorithms. Deutsche Bank (DBKGn) is the most actively trading stock, whereas Beiersdorf (BEIG) is the least.

4.5.5 Toxic periods

Results for toxic periods are depicted in table 4-13 and exemplified in figure 4-5 and figure 4-6. The number of toxic periods across the stocks in our sample varies from a maximum of 47 (Deutsche Boerse) to a minimum of 10 toxic periods within a year (SAP). In figure 4-5 the VPIN(BVC) series of these stocks for the whole year are illustrated together with the closest matching VPIN(Bulk$_{Trad}$) series. The graphs show that a high number of toxic periods is not equivalent with high volatility in VPIN. Instead, the VPIN series of SAP displays only a fifth of the number of toxic periods but the peaks are higher and last longer than the 47 peaks of Deutsche Boerse in the top chart.

Of 784 toxic periods identified by VPIN based on BVC, Bulk$_{Tick}$, Bulk$_{EMO}$, and Bulk$_{CLNV}$ identify around 55% of those periods as well. On single stock level, the share of consistent alerts varies across stocks and trade-by-trade algorithm. The highest overlap with BVC of 81.8% equally detected periods is achieved with Bulk$_{LR}$ and Bulk$_{CLNV}$ for Adidas (ADS). The lowest overlap with BVC of 32.5% is obtained with Bulk$_{Tick}$ for Bayer (BAY). Note, however, that we count a toxic period as identified by both if at least one toxic bucket overlaps, but not necessarily the whole toxic period. Therefore, we demand the absolute minimum to count a toxic period as a “match” between the two approaches and still, the numbers are quite low.

The right-hand side of table 4-13 reports results for the second comparison where we measure the percentage of toxic buckets that overlap, instead of toxic periods of several buckets length. $Bulk_{LR}$, which scores lowest in detecting toxic periods, achieves the highest score in detecting single toxic buckets with 48.2% overlap with the 35,553 buckets detected by BVC. A high accuracy on the period level does not seem to imply a high accuracy on the single bucket level. The other three algorithms achieve matches to the bulk classification VPIN of 47.5%. Again, the results differ a lot between stocks and algorithms. The highest score per stock is observable for Beiersdorf (BEI) in combination with $Bulk_{LR}$ (77.5%), the lowest for Deutsche Boerse and $Bulk_{LR}$ (22.5%).

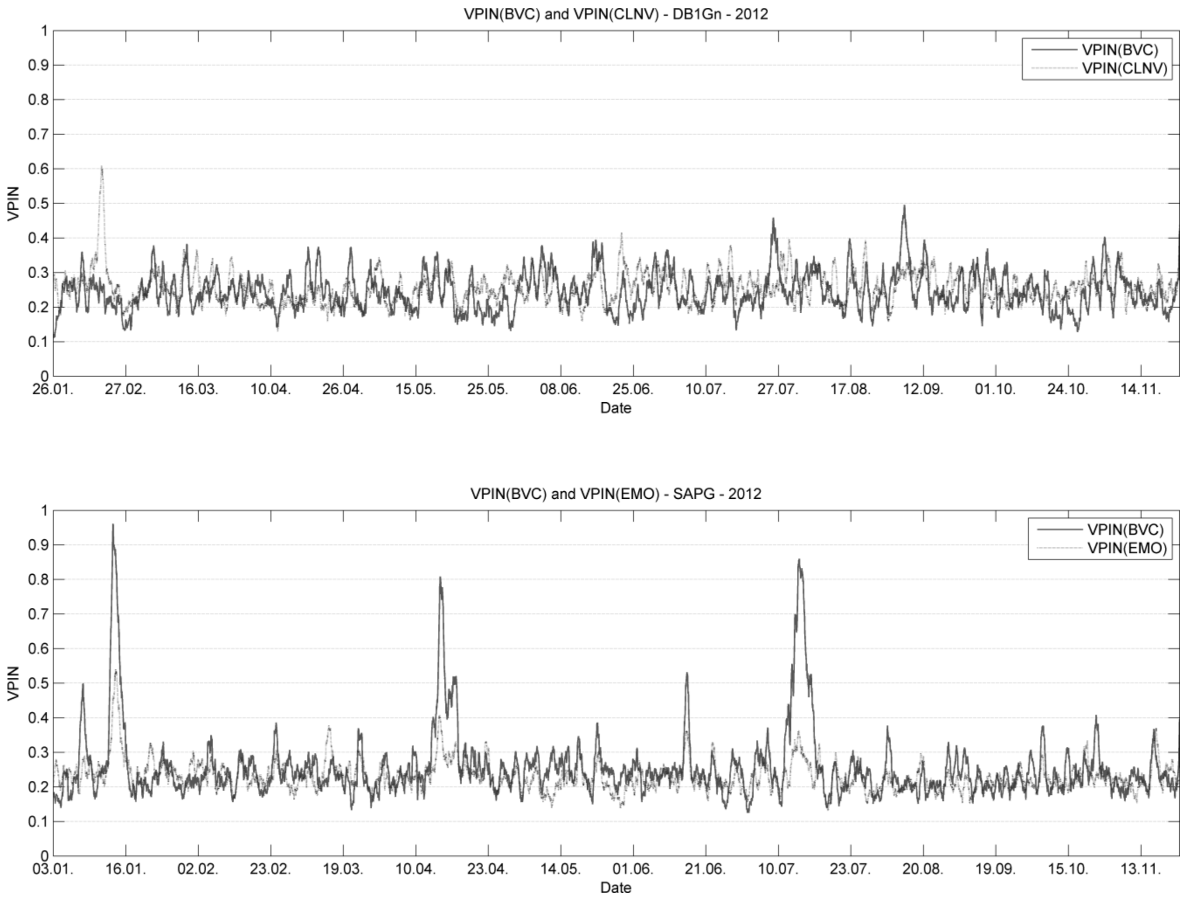

Figure 4-5: Toxic periods examples Deutsche Boerse and SAP

This figure illustrates the VPIN(BVC) series of the stock with the highest number of toxic periods (Deutsche Boerse – DB1Gn) and the stock with the lowest number of toxic periods (SAP - SAPG). Each chart also displays the corresponding VPIN(BulkTrad) series of the algorithm with the highest toxic period match. The horizontal axis is in volume time scale.

Table 4-13: Toxic periods

This table reports the number of toxic periods and toxic buckets per stock. The numbers in columns two, seven and eight are shown for calculations based on BVC. Columns three to six report the share of toxic periods identified by both VPIN(BVC) and VPIN(BVC_{Trad}), named toxic period match. The last four columns report the share of toxic buckets equally identified by VBPIN(BVC) and VPIN(BVC_{Trad}). The row "total" shows the cross-sectional averages. Yearly VPIN data is used for this table.

Stock	**# Toxic periods**	**Toxic *period* match**				**# Toxic VPINs**	**# VPINs**	**Toxic *bucket* match**			
		$Bulk_{Tick}$	$Bulk_{LR}$	$Bulk_{EMO}$	$Bulk_{CLNV}$			$Bulk_{Tick}$	$Bulk_{LR}$	$Bulk_{EMO}$	$Bulk_{CLNV}$
Total	784	.55	.53	.55	.56	35,553	326,228	.48	.48	.47	.48
ADSGn	22	.77	.82	.73	.82	1,216	11,651	.51	.55	.52	.55
ALVG	40	.40	.35	.48	.45	1,473	11,651	.26	.26	.30	.27
BASFn	47	.55	.38	.43	.45	1,787	11,651	.37	.31	.32	.33
BAYGn	40	.33	.48	.43	.48	1,353	11,651	.29	.37	.36	.34
BEIG	16	.75	.81	.81	.81	1,171	11,651	.65	.78	.72	.75
BMWG	37	.51	.49	.54	.51	1,405	11,651	.45	.47	.43	.46
CBKG	27	.56	.63	.56	.59	1,233	11,651	.35	.46	.42	.43
DAIGn	32	.41	.47	.41	.44	1,101	11,651	.43	.42	.44	.44
DB1Gn	47	.40	.45	.47	.49	1,673	11,651	.31	.23	.25	.25
DBKGn	26	.50	.50	.42	.42	1,162	11,651	.37	.37	.31	.31
DPWGn	22	.64	.46	.46	.50	1,207	11,651	.51	.43	.46	.46
DTEGn	27	.67	.67	.70	.70	1,362	11,651	.55	.51	.52	.52
EONGn	14	.64	.64	.57	.57	1,133	11,651	.64	.68	.71	.70
FMEG	26	.58	.62	.73	.69	1,196	11,651	.56	.60	.63	.64
FREG	22	.59	.55	.64	.64	1,039	11,651	.57	.56	.58	.59
HEIG	34	.65	.65	.68	.71	1,389	11,651	.47	.49	.48	.48
HNKG	28	.64	.50	.50	.46	1,339	11,651	.53	.41	.43	.41
IFXGn	16	.69	.81	.81	.81	1,057	11,651	.61	.77	.72	.74
LHAG	22	.64	.55	.59	.59	1,143	11,651	.65	.59	.50	.52
LING	24	.58	.38	.42	.42	1,095	11,651	.60	.50	.50	.50
MRCG	24	.63	.75	.75	.75	1,452	11,651	.56	.64	.59	.58
MUVGn	46	.46	.41	.44	.44	1,523	11,651	.36	.32	.34	.34
RWEG	21	,76	,57	,52	,62	1.235	11.651	,54	,49	,46	,47
SAPG	10	,70	,70	,80	,70	866	11.651	,67	,73	,69	,67
SDFGn	26	,73	,58	,62	,65	1.118	11.651	,62	,61	,60	,62
SIEGn	29	,41	,45	,45	,41	1.329	11.651	,37	,40	,41	,39
TKAG	22	,50	,68	,73	,68	1.211	11.651	,39	,50	,50	,49
VOWG	37	,51	,46	,51	,51	1.285	11.651	,49	,44	,42	,42

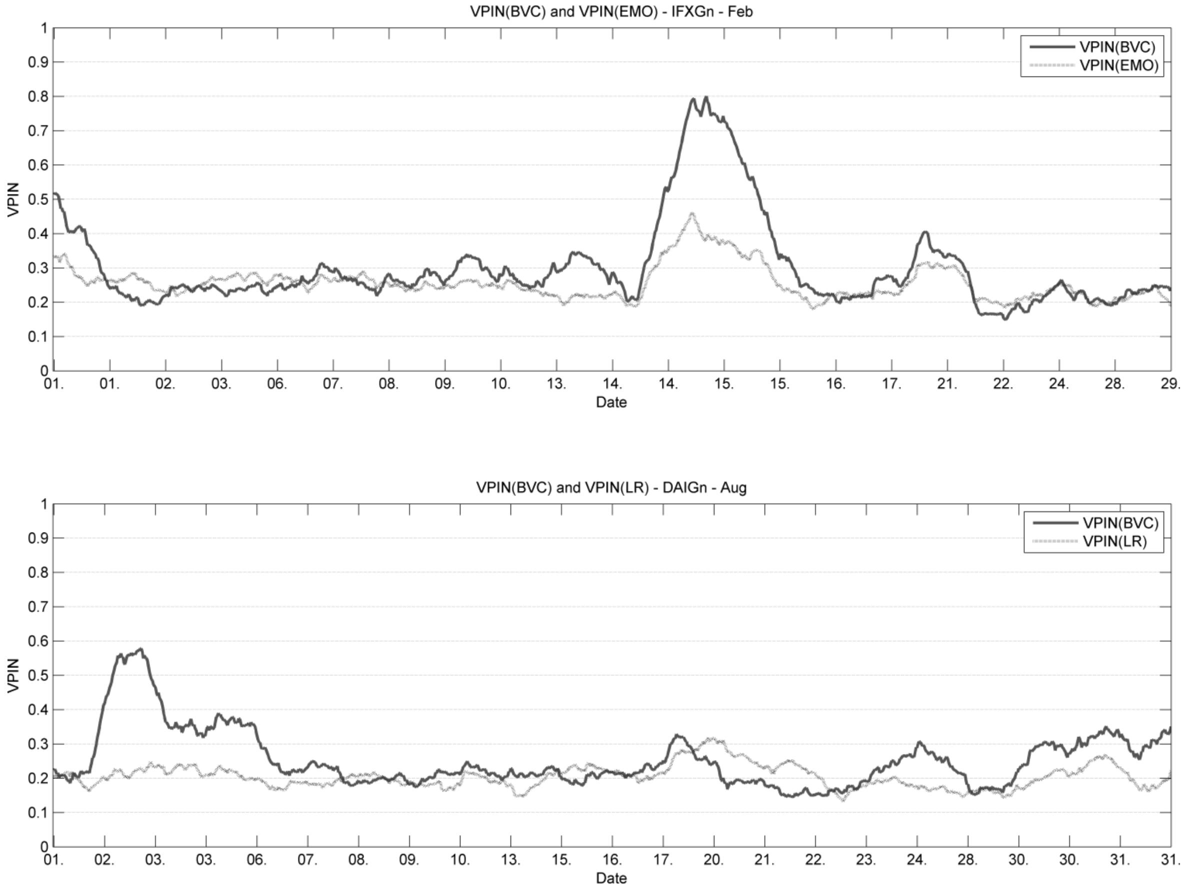

Figure 4-6: Extreme toxic period match examples: 100% match for Infineon, 0% for Daimler

This figure illustrates monthly VPIN series of a 100% toxic period and bucket match – Infineon (IFXGn) with VPIN(BulkEMO), and a 0% match – Daimler (DAIGn) with VPIN(BulkLR). For Infineon, two out of two toxic periods and 110 of 110 toxic buckets are detected equally. For Daimler, the numbers are zero of two toxic periods and zero out of 137 toxic buckets. The horizontal axis is in volume time scale.

How do the results of the toxic period overlap relate to our previous results of correlation between trade classifications? Figure 4-6 presents two VPIN time-series for one month, one where 100% of toxic and buckets periods match and a second one where 0% match. In the first graph, VPIN of Infineon (IFC) in February 2013, all toxic periods within one month are identified equally by VPIN based on bulk classification and the trade-by-trade algorithms. The second graph shows the VPIN results for Daimler (DAI) in August 2013. In this case, no toxic periods and no toxic buckets match between bulk classification and trade-by-trade classification results. The corresponding correlations of the monthly VPIN time-series are 87.3% for Infineon and 22.79% for Daimler. High correlation between the two approaches to calculate VPINs seems to imply correlation on the next higher-order level of toxic periods. Table 4-14 supports this for the whole sample. The correlation of the

VPIN time-series based on different classification algorithms correlates at about 85% with the equal detection of toxic buckets.

Table 4-14: Correlation between VPIN robustness and toxic period match
This table reports the correlation between the robustness of VPIN to calculations with different algorithms, i.e. Corr(VPIN(BVC),VPIN(BulkTrad)) from table 10, and the numbers from the toxic period and toxic bucket match, respectively, from table 13. Yearly VPIN data is used for this table.

	$\text{Bulk}_{\text{Tick}}$	Bulk_{LR}	Bulk_{EMO}	$\text{Bulk}_{\text{CLNV}}$
Toxic period match	.699	.683	.704	.676
Toxic bucket match	.858	.873	.829	.842

One of the intended practical usages of VPIN is to operate as an early warning system for toxic order flow. This application motivates the following final analyses. For all those toxic periods where both approaches overlap and thereby “agree” that there actually is a toxic period we check which approach is first in detecting the rise of toxicity in the order flow. Table 4-15 presents results per single stock. On average over all stocks, the result is close to a random draw. In 48% to 51.4% of the cases, depending on which traditional trade classification algorithm one chooses, the VPIN calculated with BVC rises first. On a single stock level this share deviates around the 50-50 split without a clear tendency towards one approach.

The detection of toxic order flow from VPIN($\text{Bulk}_{\text{Trad}}$) and VPIN(BVC) is not consistent. The choice of classification algorithm does influence the detection of toxic periods. Neither on the level of periods identified nor on the level of toxic buckets do the different methods achieve an agreement more often than in 60% of the cases in our sample. The choice of a certain trade-by-trade algorithm does not influence these results. It rather seems to be a systemic difference between bulk classification and trade-by-trade classification. In our view, the substantial difference between $\text{Bulk}_{\text{Trad}}$ and BVC in the detection of toxic periods is worrying for the application of these methods in empirical research. Towards the intended application of VPIN, the BVC approach is also not superior in detecting toxic periods earlier than VPIN calculated with the traditional approach. To clarify where this error comes from and to provide guidance about parameters that mitigate or amplify the inconsistencies we look at determinants of VPIN differences in the next section.

Table 4-15: First detection of toxic periods

This table extends the analysis of the toxic periods match of table 3-13 to check which VPIN calculation picks up toxic periods first. Only toxic periods where BVC and $Bulk_{Trad}$ equally identify a toxic period are included. For each trade-by-trade algorithm, the share of periods where the rise to a toxic level of VPIN is detected first is reported in comparison to BVC. For each stock and trade-by-trade algorithm, the remaining difference to 100% is made up from the share of simultaneous starts of toxic periods. The totals for 1714 equally identified toxic periods are 50% first detection for BVC, 46.4% for $Bulk_{Trad}$ and 3.6% simultaneously detected toxic periods.

	Tick Rule			**Lee-Ready**			**EMO**			**CLNV**		
	#over-lap tox. per.	**% First detected**		**#over-lap tox. per.**	**% First detected**		**#over-lap tox. per.**	**% First detected**		**#over-lap tox. per.**	**% First detected**	
		Bulk Tick	BVC		Bulk LR	BVC		Bulk EMO	BVC		Bulk CLNV	BVC
Total	431	.478	.480	417	.453	.504	430	.460	.502	436	.461	.514
ADSGn	17	.294	.706	18	.333	.611	16	.188	.750	18	.278	.611
ALVG	16	.500	.500	14	.357	.571	19	.474	.474	18	.444	.556
BASFn	26	.538	.423	18	.278	.556	20	.300	.650	21	.286	.714
BAYGn	13	.538	.308	19	.474	.526	17	.353	.471	19	.316	.632
BEIG	12	.500	.417	13	.692	.308	13	.692	.308	13	.538	.462
BMWG	19	.474	.421	18	.278	.667	20	.300	.700	19	.316	.684
CBKG	15	.533	.467	17	.647	.294	15	.533	.400	16	.563	.375
DAIGn	13	.462	.538	15	.333	.600	13	.308	.615	14	.429	.571
DB1Gn	19	.368	.632	21	.524	.429	22	.545	.455	23	.522	.435
DBKGn	13	.308	.538	13	.385	.538	11	.545	.455	11	.545	.455
DPWGn	14	.286	.643	10	.300	.700	10	.400	.600	11	.364	.636
DTEGn	18	.222	.722	18	.222	.778	19	.368	.632	19	.368	.579
EONGn	9	.444	.556	9	.333	.667	8	.500	.500	8	.500	.500
FMEG	15	.533	.467	16	.563	.375	19	.632	.368	18	.611	.389
FREG	13	.385	.615	12	.583	.417	14	.429	.500	14	.500	.500
HEIG	22	.455	.455	22	.500	.455	23	.478	.391	24	.458	.542
HNKG	18	.444	.444	14	.143	.857	14	.143	.786	13	.154	.769
IFXGn	11	.636	.364	13	.615	.308	13	.615	.308	13	.692	.231
LHAG	14	.429	.571	12	.417	.583	13	.308	.692	13	.308	.692
LING	14	.786	.214	9	.778	.222	10	.700	.300	10	.700	.300
MRCG	15	.400	.533	18	.389	.556	18	.389	.611	18	.389	.611
MUVGn	21	.714	.286	19	.526	.474	20	.500	.500	20	.500	.500
RWEG	16	.375	.625	12	.500	.417	11	.636	.364	13	.538	.462
SAPG	7	.714	.286	7	.714	.286	8	.625	.375	7	.714	.286
SDFGn	19	.579	.316	15	.733	.267	16	.563	.438	17	.588	.353
SIEGn	12	.667	.333	13	.462	.538	13	.462	.538	12	.500	.500
TKAG	11	.273	.636	15	.267	.533	16	.438	.438	15	.400	.467
VOWG	19	.579	.421	17	.588	.412	19	.684	.316	19	.684	.316

4.5.6 Determinants of VPIN differences

This section uses regression analysis to investigate the reasons for the sensitivity of VPIN to the choice of trade classification algorithm. The dependent variable is the difference in VPINs computed with bulk classification minus VPIN calculated with trade-by-trade algorithms. Panel A and B in table 4-16 use the absolute difference in VPIN values, panel C and D the signed difference, i.e. VPIN(BVC) - VPIN($Bulk_{Trad}$). What first draws attention is that, apart from the variables *posReturn,* all coefficients are statistically significant in every regression, most of them at the 1% level. In combination with an R^2 (within) of around 20% in panel A and B and around 30% in panel C and D, respectively, this allows the simple but important conclusion that the hypothesized variables that capture intensity in trading and pricing actually do affect differences in VPIN computation to a high degree. Each statistically significant coefficient across each panel also points in the same direction, except for the variables *Trades* and *negReturn*. This confirms results from the previous sections that the structural difference between the trade-by-trade approach and the BVC drives the different VPIN results and not minor variations in the algorithms. The variables *Trades* and *negReturn* change the sign of their coefficients when the dependent variable is used without taking the absolute value. While the model in panel A and B might be intuitive (what drives the difference irrespective of the direction?), we think panel C and D provide the sounder basis for interpretation because the unsigned VPIN delta takes on both positive and negative values[32]. The absolute coefficient of *Trades* being larger in panel C and D and 50% higher R^2 confirm the superior validity. Nevertheless, the fact that most coefficients do not change signs underscores the robustness of the findings.

The coefficients' direction and relative influence provide further insight. The coefficient of *Trades* is significantly positive in panel C and D. Given that the total volume per stock and bucket is fixed, a higher number of trades per VPIN observation, which spans 50 buckets, mean that average trade size is smaller, if at least parts of the difference in trading activity between stocks are captured by the fixed effect components of the model. Smaller trade sizes may stem from the order slicing and baiting by high frequency traders, hint at uncertainty given that no one risks larger trades or indicate the arrival of (uninformed) retail traders. All three possibilities are relevant for the degree of toxicity of order flow, albeit in different directions.

---

[32] If there is a positive relationship with another variable, such as *Trades*, that takes on only positive values, taking the absolute value of the VPIN delta will distort this relation by mirroring the negative values in the positive space, imitating a zig-zag instead of the potential linear relationship.

Table 4-16: Determinants of differences in VPIN results

This table reports the standardized coefficients and absolute t-statistics in parentheses from the fixed effects regression model of equation (4-10). Dependent variable is the difference between VPINs calculated with BVC and $Bulk_{Trad}$. The absolute difference is taken for panel A and B, the signed difference for panel C and D. Panel A and C use only every 50th VPIN observation, panel B and D use all observations and identify statistical significance with Newey-West standard errors with 49 lags. The explanatory variables relate to the timeframe of each included VPIN calculation, i.e. 50 buckets. *Trades* is the total number of trades and *ReturnVola* the volatility of all timebar-to-timebar returns within each VPIN estimation. *VolumeVola* represents the volatility of traded volume, *posReturn* and *negReturn* the stock return split into two variables to capture rising and falling market conditions separately. The variable *Timebars*contains the number of time bars during each VPIN estimation as a standardized (inverted) measure of abnormal volume. Yearly data is used for the VPIN computation. Asterisks ** indicate a 1% significance level and * a 5% significance level.

	Tick			**LR**			**EMO**			**CLNV**		
Panel A: absolute VPIN delta, spaced regression; n = 6552												
Trades	-.478	(10.5)	**	-.279	(6,0)	**	-,282	(6,2)	**	-,281	(6,1)	**
ReturnVola	.369	(17.0)	**	.365	(16,3)	**	,424	(19,3)	**	,411	(18,7)	**
VolumeVola	.233	(14.8)	**	.239	(14,8)	**	,243	(15,3)	**	,245	(15,4)	**
posReturn	-.033	(2.4)	*	-.020	(1,4)		-,028	(2,0)	*	-,026	(1,9)	
negReturn	.051	(3.7)	**	.036	(2,5)	*	,049	(3,5)	**	,047	(3,3)	**
Timebars	-.167	(12.5)	**	-.204	(14,9)	**	-,227	(16,8)	**	-,220	(16,3)	**
R^2 (within)	.190			.184			.219			.211		
Panel B: absolute VPIN delta, overlapping regression; n = 326228												
Trades	-.478	(10.8)	**	-.283	(6,7)	**	-,288	(6,8)	**	-,287	(6,8)	**
ReturnVola	.335	(7.5)	**	.329	(8,0)	**	,376	(8,2)	**	,367	(8,2)	**
VolumeVola	.205	(2.5)	*	.213	(2,5)	*	,215	(2,5)	*	,217	(2,5)	*
posReturn	-.031	(2.8)	**	-.009	(0,8)		-,013	(1,2)		-,012	(1,1)	
negReturn	.050	(4.7)	**	.030	(2,9)	**	,038	(3,7)	**	,037	(3,6)	**
Timebars	-.184	(10.6)	**	-.222	(12,4)	**	-,250	(13,7)	**	-,243	(13,3)	**
R^2 (within)	.185			.182			.219			.211		
Panel C: signed VPIN delta, spaced regression; n = 6552												
Trades	.439	(11.5)	**	.311	(7,9)	**	,284	(7,2)	**	,282	(7,2)	**
ReturnVola	.515	(28.2)	**	.456	(24,2)	**	,486	(25,7)	**	,476	(25,3)	**
VolumeVola	.171	(12.9)	**	.147	(10,8)	**	,153	(11,2)	**	,153	(11,2)	**
posReturn	-.006	(0.6)		.006	(0,5)		,000	(0,0)		,002	(0,1)	
negReturn	-.010	(0.9)		-.030	(2,5)	*	-,019	(1,5)		-,021	(1,7)	
Timebars	-.314	(28.0)	**	-.315	(27,3)	**	-,326	(28,1)	**	-,324	(28,0)	**
R^2 (within)	.318			.288			.305			.301		
Panel D: signed VPIN delta, overlapping regression; n = 326228												
Trades	.436	(11.4)	**	.298	(8,1)	**	,273	(7,3)	**	,269	(7,3)	**
ReturnVola	.458	(8.8)	**	.404	(8,3)	**	,431	(8,4)	**	,423	(8,4)	**
VolumeVola	.149	(2.3)	*	.129	(2,1)	*	,133	(2,1)	*	,132	(2,1)	*
posReturn	.011	(1.1)		.018	(1,7)		,011	(1,0)		,012	(1,2)	
negReturn	-.024	(2.5)	*	-.041	(4,3)	**	-,030	(3,1)	**	-,032	(3,4)	**
Timebars	-.328	(20.2)	**	-.326	(21,2)	**	-,339	(21,3)	**	-,337	(21,3)	**
R^2 (within)	.311			.281			.298			.295		

The evidence for the effect of high returns, positive or negative, is inconsistent. The variable capturing positive returns is mostly insignificant. Negative returns do have an effect, albeit in opposite directions and negligibly small compared the other variables' standardized coefficients. The analysis with signed delta indicates large negative returns to increase the difference.

Return volatility and volume volatility do significantly and to a higher degree increase the differences, as the significant, positive and large standardized coefficients of *VolumeVola* and especially *ReturnVola* indicate consistently in all panels. The fact that return and volume volatility positively influence the difference between VPIN(BVC) and VPIN($Bulk_{Trad}$) is worrisome, since it indicates that they react differently to periods of high volatility. Similarly, Andersen and Bondarenko (2015) find VPIN's correlation with future volatility to change its sign depending on the trade classification scheme[33]. However, especially in high volatile and potentially toxic periods should VPIN yield reliable results and not depend on the underlying algorithm. The variable *Timebars* confirms the observations from *ReturnVola* and *VolumeVola*. The more time bars a single VPIN computation spans, the less intense trading was at the time, as it took more time bars to fill the bucket with the required amount of volume. Hence, the large and negative coefficient indicates that VPIN estimations diverge less and are more robust in times of abnormally low volume, although rather the opposite market conditions require a robust estimation of VPIN.

At the end of the day an increasing positive difference means that VPIN(BVC) raises faster or exhibits more volatility than VPIN(Trad), assuming that both react roughly in the same direction, which is supported from the correlation values. This may be a welcome characteristic of VPIN, but only if a spike in VPIN(BVC) is no false positive alert. The next two sub sections provide evidence in this direction.

4.5.7 Predictive power of different VPIN implementations

The first five models in table 4-17 give a slight indication that VPIN calculated with BVC might actually capture something more than pure trade classification. The regression on realized volatility with lagged VPIN(BVC) is the only one from the first five models where

[33] "In contrast, the VPIN metric based on the actual order imbalance is negatively correlated with future volatility." (Andersen & Bondarenko, 2015)

VPIN's coefficient is significant and the R^2 is not zero, although 0.005 is very close to zero.

We used the within-R^2 from the fixed regression estimation, which basically de-trends the involved time series to correct for explanatory power stemming solely from the fixed effect dummy variables. Once we add lagged realized volatility to the regression the previous evidence is turned upside down. The relation of VPIN(BVC) and future volatility turns negative, as also identified by Andersen and Bondarenko (2015). Further, VPIN(BVC) does not add explanatory power, as the R^2 of VPIN(BVC) in regression (6) is even lower than the R^2 of model (11), which has only lagged realized volatility as single explanatory variable. The regressions with VPIN(Trad) as explanatory variables besides lagged realized volatility do not fair significantly better.

Table 4-17: Forecast regression

This table presents coefficients and t-statistics in parentheses of fixed effects panel regressions. The dependent variable is realized volatility, measured in 10-minute intervals as squared 10-minute returns, averaged over 34 intervals, i.e. the timespan of one trading day. The independent variables, listed in columns, are the latest VPIN observations available at each 10-minute interval, based on the different trade classification algorithms and realized volatility (rv). All independent variables are lagged by 51 intervals, i.e. the length of one trading day. The last column display the within R2. Significance is based on Newey-West standard errors with 51 lags. Asterisks ** indicate a 1% significance level and * a 5% significance level.

	Lagged independent variables							
Reg	**BVC**	**Tick**	**LR**	**EMO**	**CLNV**	**rv**	**Const**	**R^2**
(1)	0.00183 *** (11.7)						0.0015 *** (26.0)	.005
(2)		0.00052 * (2.5)					0.0019 *** (25.1)	.000
(3)			0.00021 (1.0)				0.0019 *** (27.4)	.000
(4)				0.00013 (.6)			0.0020 *** (27.6)	.000
(5)					0.00010 (.5)		0.0020 *** (27.7)	.000
(6)	-0.00044 ** (-2.9)					0.323 *** (17.5)	0.0015 *** (26.3)	.030
(7)		-0.00054 ** (-2.7)				0.313 *** (18.8)	0.0015 *** (21.5)	.030
(8)			-0.00091 *** (-4.9)			0.316 *** (19.1)	0.0016 *** (23.6)	.031
(9)				-0.00099 *** (-5.1)		0.316 *** (19.1)	0.0016 *** (23.8)	.031
(10)					-0.00102 *** (-5.3)	0.317 *** (19.1)	0.0016 *** (23.9)	.031
(11)						0.312 *** (19.1)	0.0014 *** (26.2)	.031

It is the sensitivity and robustness of VPIN to different choices of trade classification that form the core motivation and analysis in this chapter, and not VPIN's explanatory power. The preceding analysis, however, supports concerns around VPIN's predictive power also for equity markets, which so far have only been established on futures markets (Andersen & Bondarenko, 2014b, 2015) .

4.5.8 The crash of K+S

On July 30th 2013, Germany had its rare instance of a significant crash, roughly similar in magnitude to the "flash crash" analyzed in Easley et al. (2012b), when the stock of K+S, one of the 30 members of the leading German stock index DAX, crashed from 26,54 EUR closing price on July 29th to a closing price of 20,24 EUR on July 30th - a loss of 24% within less than a day. The better part of the crash happened in the first minutes and hours of trading. Figure 4-7 illustrates the two days surrounding the crash in clock time. The stock price starts to drop slowly during the afternoon on July 29, then crashes during the first 10 to 15 minutes of the trading the next morning. Another smaller crash happens right after noon on July 30. From just looking at the solid red and green lines, displaying VPIN calculated with BVC and tick-rule, respectively, both shoot up steeply as soon as the price drops on July 30th but do not seem to show a warning reaction before the crash. This changes when one looks at the CDFs. The CDF(VPIN) calculated with trade-by-trade classification (green dotted line) starts to rise quickly already in the late afternoon of July 29. From around 3:30pm, the CDF of VPIN(tick) climbs from around 0.35 upwards, passes the "critical level" of 0.9 (Easley et al., 2011) and turns "toxic" at 4:58pm. Traders could have been warned already the day before the actual crash. The CDF of VPIN(BVC), the red dotted line, instead even decreases during the last hours of trading on July 29. Hence, a VPIN calculated using traditional trade-by-trade algorithms would have signaled the crash, where the VPIN calculated with bulk volume classification would not. This is the exact opposite result to Easley et al. (2012b) where the analysis of the S&P 500 flash crash shows that VPIN computed with BVC signals the crash, whereas VPIN calculated with trade-by-trade algorithms does not. Admittedly, this is only one single case, but this is the by far most severe crash of a German blue chip in 2012 and 2013, and hence this evidence does not help to dispel doubts on the robustness of the VPIN metric.

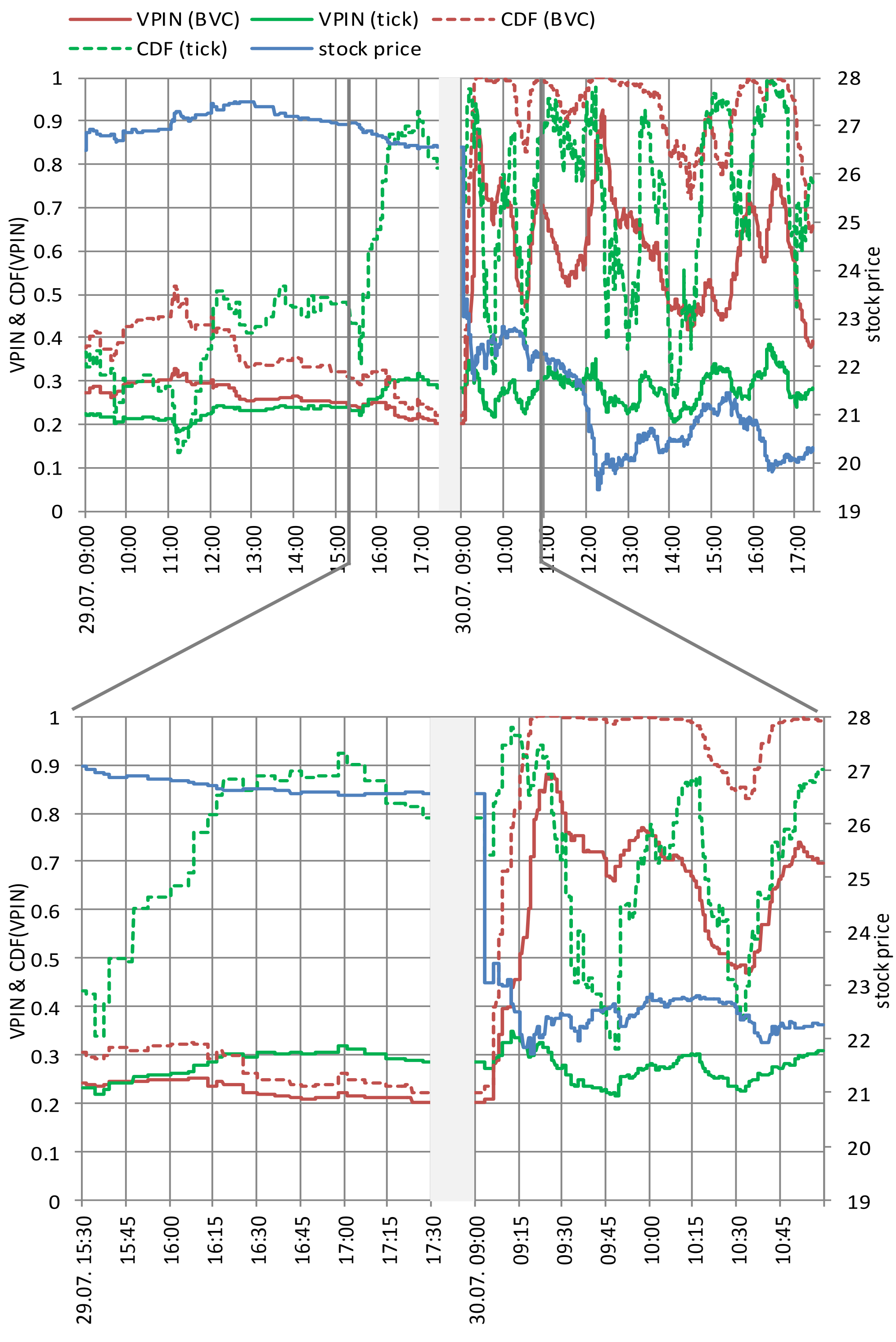

Figure 4-7: K+S crash July 29th-30th, 2013

This figure shows the crash of the stock price of K+S on July 29 and July 30 2013. The first plot gives an overview of the full two trading days surrounding the crash, the second plot zooms into the 2,5 last and first trading hours were most of the trading took place. The blue lines show the stock price, the red lines the VPIN calculated with BVC and its CDF, the green line the VPIN based on tick-rule classification and its CDF.

4.6 Conclusion

This study seeks to understand the sensitivity of the VPIN to the choice of trade classification algorithm. In light of high frequency market environments, Easley et al. (2012b) propose a heuristic approach to trade classification for the calculation of VPIN. Traditional trade-by-trade classification algorithms, however, have been evaluated on almost any financial market of interest to financial researchers and literature review and a new evaluation on proprietary signed trading data shows that they perform reasonably well, with accuracies up to 90%. We first examine whether a simple heuristic can beat a 90% accurate algorithmic approach. Second, if both approaches yield results with similar performance, we ask whether VPIN calculations provide similar results and are robust to the choice of (proper) trade classification. To address these questions we analyze comparisons of deterministic and heuristics approaches in the application of the detection of toxic order flow with VPIN.

What we find is that VPIN is not robust to the choice of classification algorithm and that the bias is increasing with the volatility of trading. On every level ascending in complexity and aggregation from raw trade classification, to order imbalance, to VPINs and toxic periods, the choice of trade classification algorithm induces substantial differences in the results. The gap is actually widening for the higher aggregate metrics instead of the generally more likely result of diminishing differences once one computes aggregates. The tick rule produces VPIN results closest to BVC despite being the least accurate trade-by-trade algorithm. In the detection of toxic periods, the major proposed application of VPIN, both approaches do not give consistent results more often than in 60% of the cases in our sample. Further, neither of the approaches is consistently faster or earlier in detecting toxic periods. On the joint sample of consistent toxic periods, VPIN based on BVC rises first in 48% to 51.4% of the cases.

Regression analysis provides hints in what trading environments VPIN should be used with particular caution. These are times of high return and volume volatility, which are especially those conditions where the application of VPIN is intended to be most useful. One promising way to extend this analysis is to evaluate the detection of true positives and especially false positives with the use of company specific or market wide information events. We have given an example in this direction by analyzing one of the most extreme crashes of a blue chip stock that occurred in Germany within the last two years– the crash of K+S on July 30th by 24%. In this case, VPIN calculated with the tick rule clearly signals the crash, whereas VPIN(BVC) does not.

Finally, our evidence for equity markets is consistent to finding of previous literature on futures markets. VPIN’s wide range of possible parameter settings, from trade classification algorithms, to trade intervals, bucket sizes or length of the rolling estimation window need to be tailored for every market design and trade volume, with an evaluation that is based, amongst others, on a real-world, true positive sample of toxic events. Otherwise, our results suggest that VPIN is not superior and reliable enough for its proposed application.

5 Information or Noise: Does Twitter Facilitate Information Dissemination?

5.1 Introduction

The social media platform and inventor of the concept of microblogging - Twitter - attracts exponentially growing volume of users and messages since its creation in 2006. The broad public is aware of Twitter not only due to its popularity as PR-platform for celebrities. Key political events in the last years such as the Arab spring and other uprisings relied substantially on Twitter for communication and organization of millions of people. Articles released by traditional news media reference how users on Twitter reflect, comment and think on the respective topic. The tabloid press is especially prone to produce headlines and articles that consist solely of consolidated discussions on Twitter. In Germany, in particular, the country of origin of this study, an "outcry" on Twitter has led to the resignation of the leader of one of Germany's five main political parties. Therefore, Twitter is certainly influential. Its constant stream of messages certainly contains information. But is it actually relevant for financial markets? Or *"is all that talk just noise?"* as Antweiler and Frank (2004) ask in their seminal study on a predecessor of today's Twitter ten years ago – stock message boards.

Whenever a new channel for information dissemination emerges, investors, companies and regulators alike are interested in determining its relevance to decide whether to incorporate it into their decision making, public relations strategy or supervision, respectively. The current study assesses the relevance and influence of Twitter in a setting where money matters – equity markets. In financial and computational research, a few initial studies have produced evidence of a predictive power of the sentiment of Twitter feeds for future market movements (Bollen, Mao, & Zeng, 2011; Nofer & Hinz, 2014b; Sprenger, Tumasjan, Sandner, & Welpe, 2013; Zhang, Fuehres, & Gloor, 2011). This study is distinguished by its approach in several dimensions: First, we couple Twitter data and stock market data on the single stock level. All other studies available analyze market- or industry-wide indices. Thereby they merely uncover that the general public mood level is somewhat correlated with the overall market movement. We would like to know whether this relation exists on a more granular, firm-specific level and hence investment decisions for single stocks should consider current activity on Twitter. Second, our study uses all data from Twitter for the covered companies and not a random subsample[34]. Third, we include all Twitter feeds that can be connected to a company and do neither filter tweets by applying sentiment analysis nor follow unofficial conventions of $- or €-tags attached to messages. The latter two choices aim at capturing the full information content on Twitter and not an arbitrary subset of it. The first choice aims at precision and practical application – is Twitter a relevant information source for the trading of single stocks, not just market wide movements? Fourth, while the mentioned studies rely on simple metrics retrieved from daily trading data, we use intraday trading data to compute advanced metrics of trading (e.g. effective spreads and intraday volatility) and, in particular, employ the daily version of the microstructure variable probability of informed trading from chapter 3 to measure Twitters impact on trading and information dissemination.

Results on daily granularity show that Twitter, from the point of financial markets, is a very active *post*-event processing or dissemination platform of new information. In a complementary interpretation, activity on Twitter can serve as yet another indicator of investor attention, but it is not superior to simply watching the market and known metrics themselves, as they react to the arrival of new information. Activity on Twitter highly significantly correlates with metrics of trading, like volume, volatility, spread and also with PIN and its parameters as one can expect from a news-triggered variable (positive with volume, volatility, spread; negative with PIN). Twitter is not correlated at all with the variable *alpha*, the probability of news arrival in the PIN model. Univariate comparisons based on quantile segmentation of the sample produce even stronger results than correlations. Lagged Twitter volume does correlate significantly with most of the mentioned variables, which may prematurely be interpreted as an indication for predictive power, but in similar or even stronger fashion do lagged variables of trading correlate with future Twitter activity.

In multivariate settings of fixed-effects panel regressions, the explanatory power of (lagged) Twitter activity diminishes (completely), but the contemporaneous relationships hold. An event study around earnings and ad-hoc announcements shows how abnormal activity on Twitter lasts a few days longer than the other trade metrics. There is absolutely no evidence, however, of increasing Twitter activity in anticipation of such announcements. A final intraday analysis confirms these findings. Twitter activity jumps

[34] The public application programming interface (API) of Twitter allows to retrieve at most 1% of the total message stream on Twitter. If a query is specific enough such that it matches less than 1% of the total message volume, the full result set is returned.

heavily within the first two minutes *after* the announcement, with no sign of an increase even just minutes prior to the release of an announcement.

The remainder of this chapter is organized as follows. Section 2 reviews related literature. Section 3 introduces our approach, methodology and variables. Section 4 describes the data and especially the collection and processing of the Twitter feeds. Section 5 gives a descriptive overview of Twitter activity in our sample. Section 6 presents empirical results and section 7 concludes.

5.2 Related literature and hypothesis

The relation between stock prices and alternative information channels different from traditional newspapers has been evaluated using internet message boards, Google Search Volume or Wikipedia page edits. Twitter, instead, has not been a topic of research in any notable finance journal yet[35]. Before social media platforms like Twitter became popular, designated internet discussion forums provided a platform to exchange opinions on a stock's value, albeit in a less open and less conveniently accessible way. Associated literature provides guidance on what to expect of the informational value of Twitter. A pioneering study by Antweiler and Frank (2004) extracts sentiment from messages posted in the discussion forums of ragingbull.com and Yahoo! finance by introducing unfamiliar methods like naïve bayes and support vector machines to the finance domain. In their intraday and daily analysis, message volume can predict small, economically insignificant negative returns and also stock volatility. Despite using the Wall Street Journal as benchmark, the authors conclude that the talk is not just noise – instead, "there is financially relevant information present" (Antweiler & Frank, 2004).

A generation forward from the simple message forums are websites like seekingalpha.com. These platforms add social media features known from facebook and the like. Following the Web 2.0 paradigm, the creation of actual content is left to the website's users, who post opinions in newspaper-imitating recommendation articles, which in turn initiate the actual discussion among users. Chen, De, Hu, and Hwang (2013b) demonstrate on a large sample covering about 8 years how sentiment analysis of content and comments published on seekingalpha.com is able to predict future stock returns and also earnings surprises. Similar

[35] To name a few, an article involving Twitter data has, at the time of writing, never been published in Journal of Finance, Review of Financial Studies, Journal of Financial Economics, Review of Finance, Journal of Financial Markets, Journal of Banking & Finance, Journal of Financial Intermediation or American Economic Review.

exercises have been pursued with similar results for other websites and market places, e.g. Germany (Nofer & Hinz, 2014a).

The described results are supportive of the “wisdom of crowds”-hypothesis that may be applicable for Twitter as well. The crucial advantage of stock message boards over Twitter, however, is the identification of companies, which is inherent in their structure. Each message post is unambiguously assigned to one or several companies. In contrast, Twitter messages need to be assigned to a company for company-level analysis. This is one reason why the existing studies on Twitter use only market wide indices to circumvent this problem. Second, stock message boards are designed for the purpose of discussing information relevant to stock pricing. The advantage of Twitter, in turn, is its global reach and a much lower “barrier of entry”, both of which result in a larger audience and a larger message volume than what is found on stock message boards. Research from the computer science domain identifies evidence that Twitter is in fact rather a platform to spread information than a social network (Kwak, Lee, Park, & Moon, 2010): The majority of the follower relationships are one-way, not two-way as a friend relationship on facebook, for instance, and 85% of the trending topics originate from headline news.

If Twitter can be regarded as a relevant information channel, the well-studied influence of traditional news channels and press coverage on financial markets provides further inspiration for testable hypothesis. Press coverage has been shown to influence investor’s attention for single stocks and hence increase or lead to an overreaction in trading volume and volatility and also predictably affect (abnormal) returns, at least in the short and medium run (Chan, 2003; Chen, Pantzalis, & Park, 2013a; Fang & Peress, 2009; Fang, Peress, & Zheng, 2014; Hillert, Jacobs, & Müller, 2014; Peress, 2014). As a general pattern among these studies, the effects of media coverage are most pronounced for small and medium-sized stocks (Chan, 2003; Fang & Peress, 2009; Peress, 2014). Peress (2014) exploits a unique sample of newspaper strikes in several countries to analyze the absence of media coverage on financial markets. His results suggest that a focus on single stocks is important, as the absence of one news channel does affect single stocks more than the market as a whole. Chan (2003) and Hillert et al. (2014), in contrast, focus on relatively long horizon drift, momentum and reversal patterns with monthly data.

Chen et al. (2013a) attribute the deviation from fundamental value triggered by high press coverage not to a bias in the media but rather to investor sentiment. The focus on sentiment in contrast to “just” the attention a stock receives through the media, measured in article count, was first prominently introduced by Tetlock (2007). In this paper, the content of a column in the Wall Street Journal generates return predictability when its tone is extracted and condensed with principal component analysis. The effect lasts for just a day, after which prices return to their fundamental value. The extraction of sentiment is also at the heart of the existing evidence on Twitter and financial markets.

Evidence for some predictive power of sentiment extracted from Twitter for a market wide equity indices has been produced for the US (Bollen et al., 2011; Zhang et al., 2011) and also Germany's DAX (Nofer & Hinz, 2014b). We are skeptical towards some of the results by Bollen et al. (2011) as their mood scores are calculated with a rolling window that looks backward *and* forward, hence a predictive relationship is not surprising but incorporated by design. There are a number of other reasons why predictive power may have falsely been attributed to Twitter. Novy-Marx (2014) entertainingly demonstrates how one can estimate potential positive abnormal returns from almost any "news source" as long as the available choice of trading strategies is diverse enough and statistics are employed and interpreted with inappropriately loose criteria.

Closer towards the approach in the current chapter are studies that relate Twitter feeds to single stocks. Sprenger et al. (2013) closely follow the research design of Antweiler and Frank (2004), replacing stock message posts with tweets. Their sample relies on a convention of tagging tweets with stock tickers like "$APPL" for tweets concerning Apple Inc. Thereby tweets are pre-filtered and assigned to a stock. The disadvantage of this approach is that it captures only a small share of all tweets related to a company. Their 6-month sample for all S&P 100 companies contains 250,000 tweets. In contrast, our sample of tweets accumulated over a 6-month period assigned to 83 companies in less-tech-affine Germany comprises more than 12 million tweets. Their results also indicate message volume to predict trading volume, but at a closer look this relationship holds in the opposite direction as well, with even stronger statistical support. The association between returns is also documented contrary to expectations - abnormal returns predict bullishness in lagged regression.

We decide against assigning sentiment to tweets by classifying them as good or bad news or defining other mood states, for several reasons. First, Twitter messages are strictly limited to just 160 characters. Sentiment is way harder to identify and misclassifications much more likely on such a very short message in comparison to a longer forum post or the articles and comments on seekingalpha.com. Since sentiment classification approaches rely on a set of keywords, a large share of messages cannot be classified if there is no keyword present. Further, the analysis is mostly limited to one language. Hence, we decide to instead capture the full attention and the most comprehensively exhaustive set of messages possible and not apply filtering by sentiment.

Empirical research with Google search volume (GSV) shares some methodology, objectives and hypothesis with our research on Twitter. GSV, like unclassified Twitter feeds, does not inherently provide a direction. Sentiment analysis is not impossible but slightly more complicated. Instead, the focus of researchers usually lies on the frequency of Google's usage as a measure of investor attention. A key difference and disadvantage compared to Twitter, however, is the granularity with which data is available. For GSV, data is usually available and analyzed weekly or monthly, rarely daily, and never intraday.

A number of studies relate GSV to market wide indices on quarterly (Andrei & Hasler, 2015), weekly (Vozlyublennaia, 2014) and daily (Da, Engelberg, & Gao, 2015) search volume data. Although starting from slightly different objectives, all three studies are again able to confirm the major implications of investor sentiment and attention theory, that is, predictable short-term return (reversals) and increased volatility. Da et al. (2015) are the first to build a sentiment measure from Google based on search terms related to household uncertainty, while the other two studies rely purely on fluctuations in search activity. On the single stock level, Da et al. (2011) established GSV as a robust indicator of investor attention, with several studies following - for example to demonstrate the increasing investor attention before scheduled company announcements (Drake, Roulstone, & Thornock, 2012).

The discussed literature provides plenty of guidance on how variables such as trading volume and volatility could react with Twitter. The following section describes our variable selection in detail. But apart from the variables and metrics used commonly in the cited studies, we also employ a variant of the microstructure variable Probability of Informed Trading (PIN), introduced by Easley et al. (1996b) and extended in Easley et al. (1997b), which we modify to allow a daily estimation of PIN. The variable PIN and the components of this well-established and empirically proven microstructure model allow us to ask a number of more detailed and specific questions. Does volume on Twitter attract uninformed traders, rather than informed traders? Is PIN affected by activity on Twitter? How does the probability of news arrival, a key parameter of the PIN model, correlate with Twitter?

5.3 Research design

We are interested in analyzing the relation of Twitter and the market (microstructure) indicators over time. To do so, our analysis is structured in five major steps. First, we need to understand the general characteristics of the Twitter data and identify suitable metrics to aggregate Twitter activity into actual variables that can be used in econometric methods. The inferential statistics then start with univariate analysis to explore basic relations between Twitter and trading. Multivariate panel analysis follows next to account for interactions between variables and to also properly differentiate lead-lag relationships. Fourth, we design an event study around major corporate announcements to explicitly evaluate Twitter in vicinity to traditional news channels. These analysis are all run with daily granularity. The involved metrics are computed from intraday data. In the last section of our results, we run a descriptive event study on the intraday level to validate conclusions drawn in the daily analysis.

The univariate analysis employs two different approaches to measure strength and direction of a relationship. The first approach uses standard correlations, measured with both the Pearson-coefficient for potential linear relationships and the nonparametric Spearman rank-correlation coefficient. The second approach, also commonly employed in the literature, compares means or medians of explanatory variables after splitting the sample into top and bottom quantiles based on the dependent variable. We run both approaches first on the full dataset and in Fama-MacBeth fashion per single company to afterwards aggregate the results.

The data set for the multivariate analysis at the daily level is well-balanced panel data. The level of tweet activity, however, varies tremendously between companies. Even after introducing standardization or a measure of abnormal tweet activity, the impact of activity on Twitter on continuous trading of the underlying stock (or vice versa) is likely to differ between companies. We follow the excellent guidance in Petersen (2009) to model our panel analysis. The author identifies a large share of studies in top finance journals to design their panel regression analysis such that standard errors are biased downwards and hence the statistical significance of results may often be overstated. Peterson's recommended approach, which he also simulates on a dataset almost exactly the size of ours (we have 83 x 123 and Petersen uses 100 x 100), is to address one of time or firm effect parametrically, i.e. including dummies, and then estimate standard errors clustered in the other dimension. While we need to control for company fixed effects we find no reason to believe that we have to control for time-fixed effects. Accordingly, following this approach results in a fixed effect regression model

$$Y_{it} = \beta_{1i}X_{1it} + \cdots + \beta_{mi}X_{mit} + \alpha_i + u_{it} \tag{5-1}$$

where Y_{it} is the dependent variable for company i at trading day t, α_i is the unknown intercept for each company, u_{it} is the error term and X_{mit} are m independent variables with their coefficients β_{mi}.

The event study covers two major types of announcements that are known to contain price-sensitive information and hence are very likely to have a significant impact on the stock price and trading: ad-hoc announcements stipulated in German insider trading law and quarterly earnings announcements. In contrast to "common" event studies we are not interested in measuring abnormal stock returns but instead focus on how and when Twitter activity reacts and compare this reaction with the other market indicators. The combination of the two news sources provides us with both scheduled and unscheduled announcements, to which pre-event anticipation and investor attention may differ.

5.3.1 Metrics of trading activity

The following variables characterize trading activity in several relevant dimensions: return, volume, spread, volatility and trade size. Trading volume is an obvious indicator to assess trading activity and is measured in money terms, i.e. the amount of shares traded times the price in Euro. Taking the log corrects for the skewness, hence the according variable is named $lnVolume$.

The spread is one of the most widely employed trading metrics to measure not only transaction cost but in particular liquidity, information risk or the degree of information asymmetry. We use two measures of the spread. The percent quoted spread is defined for stock *i* at time *t* as the value of the absolute spread relative to the midpoint price:

$$qspread_{it} = 2 * \frac{ask_{it} - bid_{it}}{(ask_{it} + bid_{it})} \tag{5-2}$$

For the variable $qspread$ we weight each quote over the time it was effective and also report results for an equal-weighted quoted spread with variable $qspreadeqw$. The log of both variables is used for empirical analysis. In times of high-frequency trading with multiple order entries and cancellations happening within milliseconds, the quoted spread may not give the most realistic picture of what investors actually have to pay in transaction cost once they execute a trade. Hence, we also compute the realized or effective spread on actual trades. Following Holden and Jacobsen (2013) the effective spread for a trade k is defined as

$$effspread_k = \frac{2D_k(D_k - M_k)}{M_k} \tag{5-3}$$

where D_k is an indicator variable equaling +1 if the k^{th} trade is a buy and -1 if the trade is a sell and M_k is the midpoint price of the quote prevailing at the time of the k^{th} trade. Most studies use the midpoint prices in both quoted or realized/effective spread measures (e.g. Scholtus, van Dijk, & Frijns, 2014). The effective spread is volume-weighted and analyzed in its log transformation.

Volatility in prices is measured as the standard deviation of 15-minute log returns. The time intervals used in the literature for computing intraday volatility typically range from 5 minutes (Ederington & Lee, 1993) to 60 minutes (Mitchell & Mulherin, 1994). We also calculate results for in 5 and 10 minute intervals as a robustness check. In this case, using raw returns or log returns does not make any difference since the absolute numbers fall in a range very close to zero where the log transformation is almost linear. By using midpoint returns instead of returns based on actual trades we reduce noise in the volatility measure due to the bid-ask bounce and instead capture the bid-ask spread with the afore mentioned metrics. For stock i on day t, whose trading hours are sliced into B 15-minute intervals, the intraday volatility is calculated as:

$$15minVola_{it} = \sqrt{\frac{1}{B}\sum\nolimits_{b}^{B}(r_{itb} - \overline{r_{it}})^2} \tag{5-4}$$

An alternative and simpler metric for daily volatility is the difference between the high and low price of stock i on day t, divided by the high price (Chakrabarty et al., 2012a).

$$hlvola_{ti} = \ln\left(\frac{p_{tiH} - p_{tiL}}{p_{tiH}}\right) \tag{5-5}$$

A final set of variables intends to figure out what types of traders react to certain news. Even a simple distinction between retail and institutional investors cannot be inferred from the trading data, as (anonymous) identifiers are rarely available to researchers. Antweiler and Frank (2004) use trade size in levels of 100,000 and 1 million USD to distinguish small from large traders and thereby hope to also distinguish retail and institutional investors. We replicate this idea in two variants on our trading data. First, we categorize trades with similar thresholds of 100,000 EUR and 1 million EUR in small, medium and large trades and also use 50,000 EUR and 500,000 EUR as a second set of smaller thresholds (variables *slt100, slt1mio, slt1miop* and *slt50, sl500, slt500p,* respectively).

We further introduce a second slightly more sophisticated approach to account for potential differences in typical trade sizes within each stock of our sample by building a historical distribution of trades' size individually per stock. The trade size distribution allows to determine what a "large" or "small" trade is per stock based on quantiles. The variable $xt3share_{it}$ measures the share of trades relative to all trades on day t for stock i that are smaller than the tercile of the stock's historical trade size distribution. The variable $xt3lnratio_{it}$ is the log of the ratio of the number of trades in the lowest tercile relative to the number of trades in the top tercile for stock *i* on day *t*. Again, the terciles are calculated individually per stock over the whole sample period. The variables $xt5share_{it}$ and $xt5lnratio_{it}$ are computed in similar manner with quintiles instead of terciles.

5.3.2 The probability of informed trading

The probability of informed trading is a composite variable based on a microstructure trading model developed in a series of papers by David Easley, Nicholas Kiefer and Maureen O'Hara (Easley et al., 1997b; Easley et al., 1996a; Easley & O'Hara, 1987, 1992). We introduced a modification of the model's estimation procedure in chapter 3 to allow one estimation of the model's parameters per day instead of one estimation per 30 trading days.

Let us quickly re-cap the model's intuition and parameters. There are two types of traders, informed and uninformed, who arrive sequentially to trade a risky asset with a competitive and risk-neutral market maker. The assets value is determined by information events, which happen to happen with probability α and contain bad news with probability δ and good news with probability $1-\delta$. Only the informed traders can observe existence and direction of a signal and consequently only trade if there is a signal. The uninformed traders trade independently of the arrival of a signal purely for liquidity reasons. The probabilities μ and ε describe the ability of informed and uninformed investors, respectively, to actually trade, once they decide to trade. This trading structure is described in figure 2-1. The market maker knows the structure of the trading process and must update his believes of the realizations of underlying parameters after every trade or the absence of a willing trade partner to adjust his quotes.

Derived from Easley et al. (1996a), we have shown that the described parameters can be estimated from tick data with the only input being the number of buys and sells and the number of times no trade happened for a certain amount of time. The estimation of the four parameters then allows to calculate the probability of informed trading:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon(1 - \alpha\mu)} \tag{5-6}$$

Relevant for our study is not only PIN, but in particular the reaction of the single parameters α for the arrival of news and ε and μ for the arrival of the different types of traders.

5.4 Sample composition and descriptives

This study covers the top German equity stocks belonging to the major indices DAX (Top 30), MDAX (next top 50) and TecDAX (Top 30 technology stocks). The resulting portfolio of companies is well diversified across all industries and hence representative. All shares are traded continuously on XETRA, the electronic trading platform of Deutsche Boerse. A technical advantage of the German market is its much lower level of fragmentation compared to the commonly studied markets of the US, where recent studies find the share stocks traded in their home venues to be as low as 25% for NYSE and 30% for NASDAQ (Holden & Jacobsen, 2013).

The final Twitter sample comprises about 12 million tweets, the trading data contains 29 million trades and 143 million updates to the best bid and ask. We describe in the following paragraphs how we collect, clean and organize the corresponding data from Twitter, as well as how we process the trading data, classify trades and calculate daily estimates of PIN.

Additional data sources employed in this chapter that do not require further pre-processing are ad-hoc announcements and earnings announcements for all stocks in the sample. Both types of announcements are sourced from ThomsonReuters.

5.4.1 Twitter data

Twitter lets anyone with suitable programming skills and a server connected to the internet listen to their real-time message feed via dedicated APIs (application programming interface[36]). For the purpose of this study, we distinguish three ways to identify relevant tweets. First, we identify tweets by searching the message text for the company's name or

[36] See https://dev.Twitter.com/streaming/overview for details. We used the "Public streams".

associated stock tickers (referred to as TEXT-tweets). Second, if a company runs its own Twitter account, we can easily retrieve the tweets sent from this account to the public (referred to as FROM-tweets). The third type of tweets are those sent from other Twitter users to the company account (referred to as TO-tweets). For most part of the analysis, we will focus on the by far largest group, the TEXT tweets, identified by the company name. In fact, TO and FROM tweets may be identified as well as a TEXT tweet if they contain the company name.

The stream of Twitter data that is available to the public for free but can only be retrieved in real-time. In consequence, historical tweets cannot be retrieved and instead interested researchers must set up and configure a server to collect the tweets and wait for a time-series sample to accumulate over time. We started this process in early October 2013. Configuring and customizing the Twitter API, the database and defining the stock filter criteria took us a few days with some trial and error involved. We end up with a usable and complete sample starting on October 20th 2013. The end of the observation period is April 30th 2014, which gives us almost half a year of Twitter data.

5.4.1.1 Collecting tweets via full-text search (TEXT-tweets)

Correct identification and assignment of tweets to a company in our sample is crucial for the validity of this study. In contrast to the literature using messages posted in forums, where each thread or topic is explicitly assigned to a company, this task is rather tricky and ambiguous for tweets. The majority of the literature on Twitter and stocks avoids this issue by relating overall Twitter activity (or sentiment) to the development of broad stock indices where identification of a specific company is not required. We want to assign each tweet to one company (or a few).

Text search with the basic name of the company is very likely to lead to many false positives in the case of synonyms (e.g. “Continental”) or short acronyms appearing within other words (e.g. “RWE” in “Orwell”). For a number of companies where the synonym usage is very obvious and difficult to circumvent, we do not include them in the initial Twitter text search at all (although we do collect tweets from and to their corporate accounts, if existent, e.g. Bayer, SAP, MAN, Henkel, Linde).

After excluding the most obvious synonyms from the start, the collected data still contains a number of companies with names likely to be used ambiguously. We manually examined a random sample of about 200 tweets for every company in our sample with potentially ambiguous names. For six of those, adding the term " AG", the abbreviation for stock corporation in German, at the end of the search string helped to identify tweets of these companies properly. Some special cases, however, like the social network company "XING" proved impossible as even "XING AG" picked up way more tweets about "boxing again" than the company we are looking for. Another good example of unexpected false positive is the search for the company "Dürr", which picked up thousands of tweets about Adidas-sponsored skateboarder Dennis Durrant, but almost none relevant to the automotive supplier Dürr AG. For another three, namely BASF, BMW and K+S, the manual inspection of the sub sample of tweets showed very few if any false positives, hence we decided to leave them in our sample. In the case of the search string "BWM", which, of course, also identifies tweets not relating the car company, the brand seems so strong that the number of true positives by far outweighs the false positives. In total 16 companies are removed from the sample (or not included in the first place) due to heavy noise from ambiguous tweets where no suitable circumvention could be found. Due to an unnoticed technical error, we did not collect tweets for 10 companies in the MDAX, hence our MDAX sample consists of 40 instead of 50 stocks. Table 5-1 gives a summary of the described Twitter text search by index membership, table A.6 in the appendix lists all companies of the sample with the respective search strings.

Apart from the company name, we also include the Reuters Instrument Code (RIC) form our trading data in the text search for tweets associated with companies. RICs are usually 6 to 8 characters in length. The last three characters identify the exchange a security is traded on and the first 3 to 5 characters are an acronym based on the company's name. The RICs

Table 5-1: Twitter text search sample

This table describes the composition of the sample of companies whose tweets are obtained by filtering the global Twitter stream for the name of the companies. Due to ambiguous names used as synonyms not all companies are included. Column (1)-(4) list the result of a manual examination of the company names and a random sample of their matching tweets. For a few companies in the MDAX technical difficulties prevented us from including them (6).

		Check for ambiguity in company name					
Index	**No of. members**	(1) Unam-biguous	(2) Adjusted search criteria	(3) Excluded due to ambiguity	(4) No tweets in obs. period	**(5) Final Twitter TEXT sample**	(6) Not in Sample
DAX	30	21	3	6	-	**24**	
MDAX	50	29	1	8	2	**30**	10
TecDAX	30	27	2	1	-	**29**	
Total	110	77	6	15	2	**83**	10

are added to the search criteria in both the long version containing the ending of the exchange and the short version without the exchange identifier (e.g. “BMWG.DE” and “BMWG” for the company BMW AG, where “.DE” is the exchange identifier for XETRA).

Upon validation of the received hits, however, we deleted tweets identified only by the RIC ticker and not by the company name because the RIC alone matched more often with abbreviated, encrypted links to websites in the message text than with actual company news. Further, all true positive tweets containing a RIC ticker also contain the company name, hence no tweet is lost by removing the ticker-only-identified tweets.

A number of papers leverage an informal convention on Twitter where users flag messages relating to a company’s stock with a “$”-sign (or €-sign) and its ticker symbol, like $APPL for messages concerning Apple Inc., instead of the hashtag “#” for tagging a message to a certain topic (Sprenger et al., 2013). The mentioned authors hope to achieve a higher share of actually relevant feeds by using only the small share of tweets flagged with $+ticker symbol. We decide against using this approach, for two reasons. Firstly, we expect this convention to be much less popular for German stocks, irrespective of using a dollar or a euro sign. Second, we want to capture all information content for a specific company, not just those messages declared relevant for trading by a smaller subpopulation of Twitter users who stick to this informal convention.

Descriptive evidence from the collected tweets supports our approach to neglect $- and €-tags. For a small sample of three stocks, the stock’s ticker symbol is identical to the body of the RIC code. Hence, we can estimate the potential losses from ignoring $-tags on this subset. Table A.7 in the appendix shows the search strings used and the corresponding number of messages found for the three companies. While EADS has over 100.000 tweets identified using their full company name “EADS”, only 405 tweets are identified by “$EAD” or “€EAD”. And of those 405 tweets, 398 also mention the full company name. The same pattern is observed in smaller magnitude, but similar relations, also for the two other companies, “NORMA Group” and “United Internet”. For these two companies, no additional tweet is gained through messages identified by a dollar or euro sign plus the ticker symbol. We have no reason to believe that this relation does not hold for the rest of our sample, that is, the majority of relevant tweets can be identified by using just the company name.

We lose some tweets due to Twitter's treatment of links to other websites in its messages. Given Twitters restriction on message length, links to a website are usually converted into a cryptic shortened link that redirects to the actual longer web address. The Twitter API applies its filtering to the original long version of the link, while only the short, encrypted version of the link arrives in our database. Hence, we lose all tweets that were identified only through matching company names in the long link, as we cannot re-assign them to a company.

In summary, after carefully choosing the right search criteria and manually validating the collected sample, we are confident to have accumulated a sample of tweets with a very high true positive rate regarding the true association of a tweet with the assigned company.

5.4.1.2 Collecting tweets from and to a company's Twitter account (FROM- and TO-tweets)

We track the Twitter representation of all companies in our sample at the start and at the end of the observation period by searching for their accounts directly on Twitter and on Google. In case of doubt, we track more than one account per company. Job advertising channels ("e.g. suedzucker_jobs") and country-specific websites for global companies like Siemens are not considered. Four of the 100 observed companies changed their Twitter account representation during the observation period by activating a new Twitter account, closing the existing one or switching to a different account. Hence, we cannot use these companies in any analysis that uses tweets from or to a company but can still use them for analysis covering tweets about a company[37]. Table 5-2 summarizes the available Twitter accounts by index membership. Given that we are looking at blue chips stocks in the DAX and at presumably "Technology" stocks in the TecDAX, it is quite surprising that only 72 of the 100 companies in the initial sample run a corporate Twitter account. Connecting to customers and investors on all available media channels does not seem to be on the agenda yet for a fair share of companies in Germany.

[37] A special case is the joined corporate account for the companies Fresenius SE and Fresenius Medical Care, who both belong to the top index DAX. Fresenius SE holds the majority stake in Fresenius Medical Care and the Fresenius Group manages a central Twitter account, @fresenius, that covers both companies. As we cannot properly distinguish automatically for which of the two companies a tweet was intended for, we assign every tweet from the corporate account to both Fresenius SE and Fresenius Medical Care.

Table 5-2: Twitter account sample description

This table describes the composition of the sample of companies with respect to the existence of an account operated by the company on Twitter. Company accounts were searched at the at the start of the observation period, results are listed in column (1). The existence of the accounts is checked again at the end of the observation period, April 30th 2014, in column (2). Some companies switched or deleted their account (3), some were never active our observation period (4) and for a few companies technical difficulties prevented us from including them (5).

	Twitter account			
Index	(1) Active October 2013	**(2) Still active April 30th 2014**	(3) Switched/ deleted accounts	**(4) Not active on Twitter**
DAX	26	**24**	2	**4**
MDAX	25	**24**	1	**15**
TecDAX	21	**20**	1	**9**
Total	72	**68**	4	**28**

5.4.1.3 Data collection and storage

The raw data collected amounts to more than 50 million tweets, which translates to roughly 15 Gigabyte stored in a MySQL database just for the raw tweets, without auxiliary tables or indices. On a dedicated server with large main memory, an appropriate configuration of the database allowed the complex join and filter queries with millions of records to be executed within a reasonable amount time of a few minutes or seconds. Table 3 lists the resulting number of tweets on relevant stages of the sample composition process. After all cleaning, pre-processing and the join with the trading data of the final sample of companies, 12 million tweets are left. The relevant data items within each tweet are the actual message text, a timestamp to the second of the time of the publication of the tweet and an indicator for re-tweets. Other fields like the location, user names, hashtags and other status information are not used in this study.

Table 5-3: Twitter sample build-up

This table documents the number of tweets collected for this study across several processing stages up to the final sample composition. The totals in the second and last column differ, as each tweet may be assigned to more than one company. All numbers in thousands.

		Tweets assigned to companies			
Sample pre-processing step	**Unique Tweets**	**Text**	**To**	**From**	**Total**
Total tweets collected in database	57,069	28,409	937	141	29,487
...within observation period	54,748				
...after removal of "MAN" synonyms	27,996				
...with at least one company assignment	26,390	26,200	867	129	27,196
Final company sample (all weekdays)	17,961	17,519	756	126	18,401
Final company sample (trading days only)	12,718	12,351	578	100	13,030

in thousands

What we do not collect from Twitter is information on an account's followers. Users on Twitter can "follow" other users and thereby subscribe to their tweets, which will then appear in the following user's news stream. The current literature has conflicting views on the importance of followership. Some scholars produce "better" results with follower-weighed tweets (Nofer & Hinz, 2014b) and argue that more followership implies tweets of higher relevance or quality. On the other hand a number of article conclude that re-tweets, which we cover, are a valid indicator of the quality of information (Kwak et al., 2010; Sprenger et al., 2013), whereas the pure number of followers is not informative (Kwak et al., 2010; Zhang et al., 2011).

5.4.1.4 A first descriptive dive into Twitter data

This study is not the first to analyze Twitter data, but the first to do so on a comprehensively exhaustive single company level. Hence, the following descriptive statistics and plots are intended to provide a basic understanding and first insights on the amount of information related to listed companies that is created on Twitter and how it evolves over time and in the cross section.

The cumulative volume of tweets by company, sorted in descending order, is plotted in figure 5-1. It is striking to see the high skewness of the activity on Twitter. The search term "Adidas" alone identifies half of our whole sample based on text search. The addition of two worldwide popular German car brands, BMW and Volkswagen, covers almost 90% of our sample - 11 million tweets over the 6-month observation period. On the other hand, even this heavily skewed distribution still leaves a million tweets for the other 80 companies in the sample and hence enough activity to distinguish regular from unusual activity. This observed skewness is a characteristic of news coverage that has already been identified in a number of other studies on different news channels (e.g. Chen et al., 2013a; Fang & Peress, 2009; Hillert et al., 2014). The challenge arising from this distribution is how to properly account for this disparity in the statistical analysis. Some standardization is clearly necessary; otherwise all regression or correlation analysis would purely resemble a size effect. The skewness in coverage also seems to draw a divide between consumer-focused and purely business-oriented companies.

Let us turn from the cross section to the time dimension. Figure 5-2 plots the weekly activity on Twitter over the observation period. The holiday breaks around Christmas and Easter are visible and would diminish if one standardizes the number of tweets with the days of actual trading activity (not shown). Upon manual inspection, the few spikes are

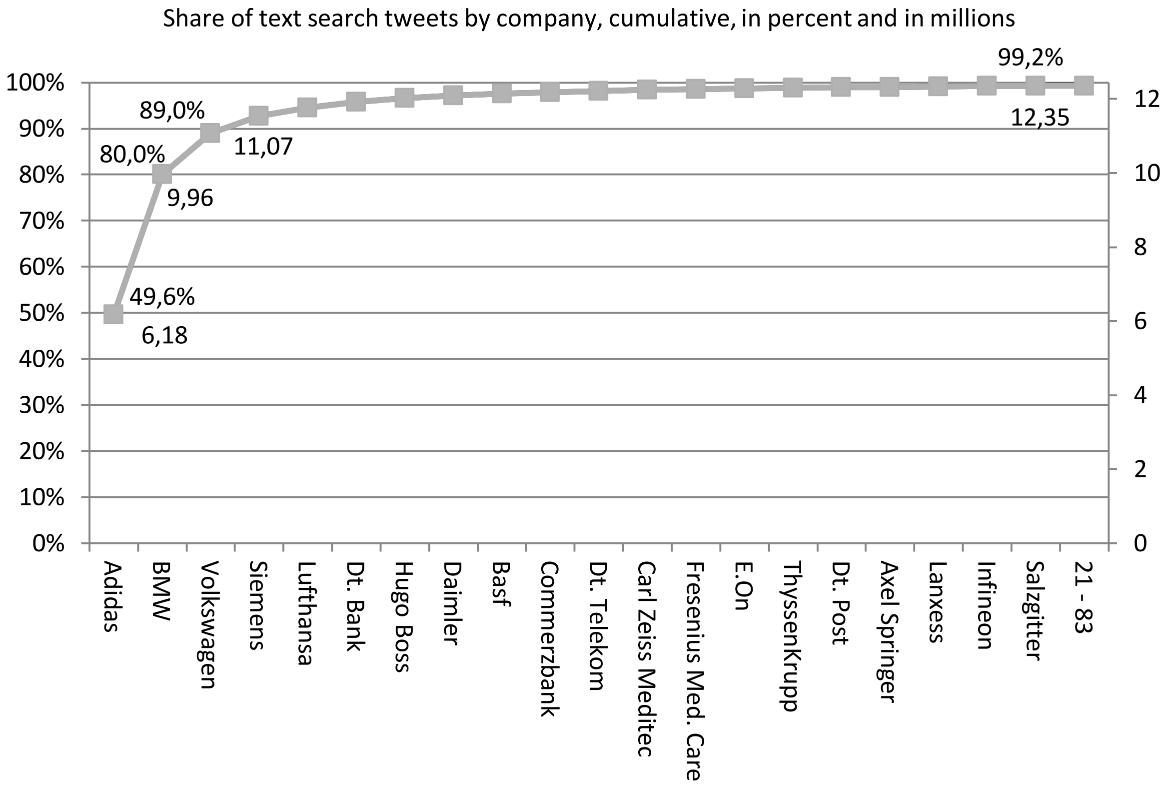

Figure 5-1: Cross-sectional distribution of tweet volume

This figure plots the cumulative share of the volume of tweets by company, sorted in descending order.

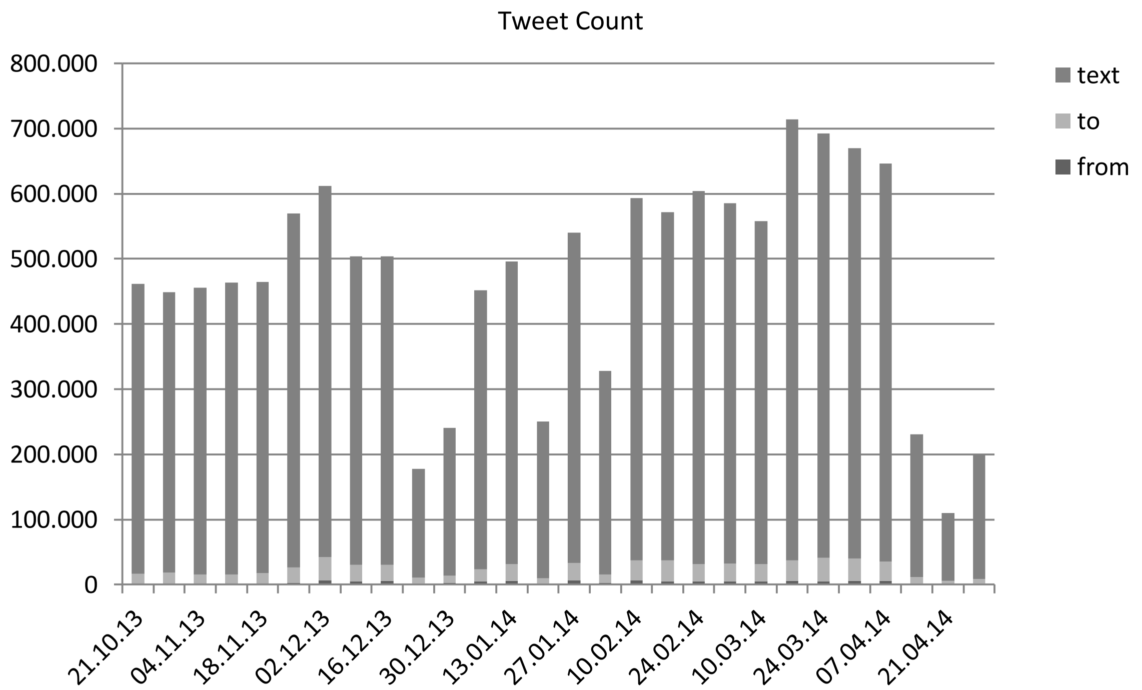

Figure 5-2: Weekly Twitter activity over observation period

This figure visualizes the tweet count per week over the observation period.

driven by highly unusual volume of some companies on certain, seemingly randomly distributed days. What is relevant for the statistical analysis is that there is no clear trend observable, which would need to be accounted for. This figure also indicates the relation of TEXT-tweets to those tweets sent from and to companies – they differ by several orders of magnitude and warrant but also enable different methodological approaches for statistical analysis.

Scholars have documented a day-of-the-week effect for various aspects of trading and the release of information. However, as figure 5-3 shows, the distribution of tweets is relatively uniform during weekdays from Monday to Friday. If anything, instead of the u-shaped pattern for information arrival, we find a small hump spanning from Monday to Friday. Activity on weekends does drop visibly, but not to level that can be regarded as negligible. However, we decide to forgo the complexity to map the weekend volume on Monday's tweet volume (other related studies ignore activity on weekends as well, e.g. Bollen et al., 2011; Das & Chen, 2007). Instead, we shift the Twitter volume occurring after the close of trading to the next trading day for every analysis based on daily data.

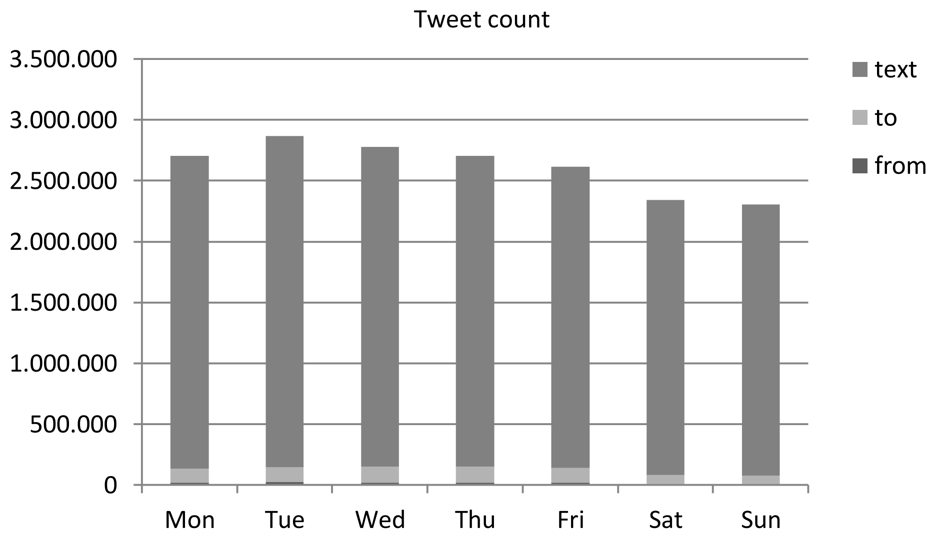

Figure 5-3: Tweets per weekday
This figure shows the average number of tweets per weekday.

In a final step down in granularity figure 5-4 plots the number of tweets per hour of a trading day. Tweet activity is slightly increasing during the trading hours and does not decline as much as one would expect overnight, which may also be attributable to the fact that we look at a worldwide sample and not a tweets originating in the German time zone only. Almost half of the activity in our sample (43%) occurs during trading hours from 9:00 to 17:30.

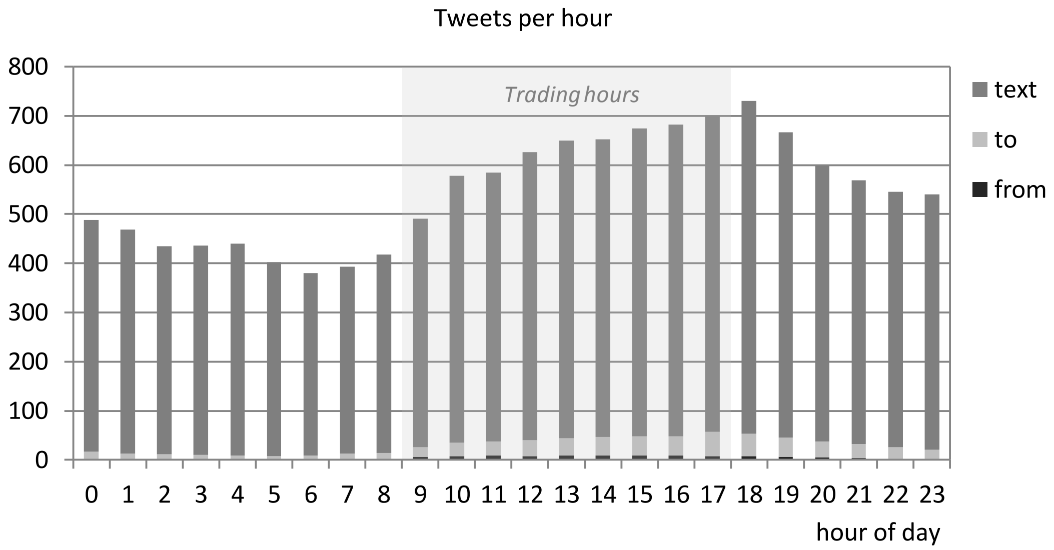

Figure 5-4: Tweets per hour of day

This figure shows the average number of tweets per hour of day. Trading hours on XETRA run from 8:30-17:30 and are shaded in light grey.

Re-tweeting is a core feature of Twitter that allows messages to spread exponentially through the network. Re-tweeting a message on Twitter means re-publishing someone else's tweet to one's own Twitter audience, i.e. his followers. Each re-tweet multiplies the number of potential message receivers. Hence, re-tweets qualify as a potential measure of relevance or quality of content. In our sample, about 8% of all tweets trigger a re-tweet. These re-tweets in turn account for about 36% of all tweets. In absolute numbers, 4.5 million of our 12.4 million tweets are re-tweets based on 640,000 tweets that users deemed worthwhile for redistribution. Usually little content is added to a re-tweet, the sole purpose is a magnified dispersion. One single tweet is re-tweeted more than 50.000 times in our sample (A promotional video contest by Adidas involving Justin Bieber), the average is 7.1 re-tweets and the median is just one re-tweet per tweet that is re-tweeted at least once. We have an implicit weighting by re-tweets in our sample as we do not filter out re-tweets but instead capture each re-tweet as another new tweet, just with a new timestamp. Thereby the increasing relevance of a topic is captured in exactly the same way as it spreads on Twitter.

5.4.1.5 Standardizing Twitter activity

The descriptive statistics of Twitter activity document a highly skewed coverage where some companies like Adidas regularly trigger thousands of tweets per day and other companies, like Heidelbergcement, stimulate only a handful of people worldwide to publish their thoughts online. Hence, developing a methodological consistent approach to measure abnormal Twitter activity is not trivial. Some studies of similarly skewed samples of news use the residuals of panel regressions (Chen et al., 2013a; Hillert et al., 2014). However, their models rely on monthly or yearly data that is not available or varies in a meaningful way on a daily basis, like company size and industry affiliation.

Instead, we use two different approaches to transform the tweet activity. First, we employ a normal standardization by subtracting the mean and dividing by the standard deviation – individually per company. Hence, for company i on day t with tweet count $tw_{i,t}$, sample mean $\overline{tw_i}$ and sample standard deviation $\widehat{\sigma_i}$ the variable $twbyz_{i,t}$ is calculated as

$$twbyz_{i,t} = \frac{tw_{i,t} - \overline{tw_i}}{\widehat{\sigma_i}} \tag{5-7}$$

Second, we normalize tweet activity by subtracting the (log) median number of tweets for company i on the same weekday k during the observation period from the (log) number of tweets on day t for company i. This second standardization has been introduced by Da et al. (2011) to capture abnormal activity of Google Search Volume and has subsequently been adopted by other scholars (e.g. Drake et al., 2012). The only difference is that, given our relatively short sample, we do not use a rolling standardization but instead the full sample:

$$atwwk_{i,t} = ln\left(\frac{1 + tw_{i,t}}{1 + median_{k\in\{k,k+7,\dots,n-mod(n,k)\}}(tw_{i,k})}\right) \tag{5-8}$$

The advantage of $atwwk$ over $twbyz$ is its provision for weekday effects. Further, the median is more robust than the average with respect to outliers.

Figure 5-5 exemplifies the process in histograms of the full panel set. The first histogram on top left shows the raw count of Twitter messages per day and company, for the whole sample. Skewed to the extreme, no one can expect reasonable results based on this data as every analysis would be driven solely by outliers. The second graph shows the natural logarithm of the tweet count. While the logarithm helps to transform variables such as trading volume closer to a normal distribution, it is not helpful in this case, as the high skewness and clustering at 0 would let regression results still exhibit a pure size effect with an inflated R^2. The graph on the bottom left is the result of standardizing the tweets per day and company to a mean of 0 and standard deviation of 1 ($twbyz$). While the graph still exhibits a slight skewness and fat tails, cross-sectional analysis is much less likely to be driven by single companies or outliers. The fourth graph on the bottom right shows the $atwwk$ measure which is the least skewed but slightly less continuous than $twbyz$. We will evaluate both variables $twbyz$ and $atwwk$ in comparison and as a complementary robustness check in our empirical analysis involving TEXT tweets.

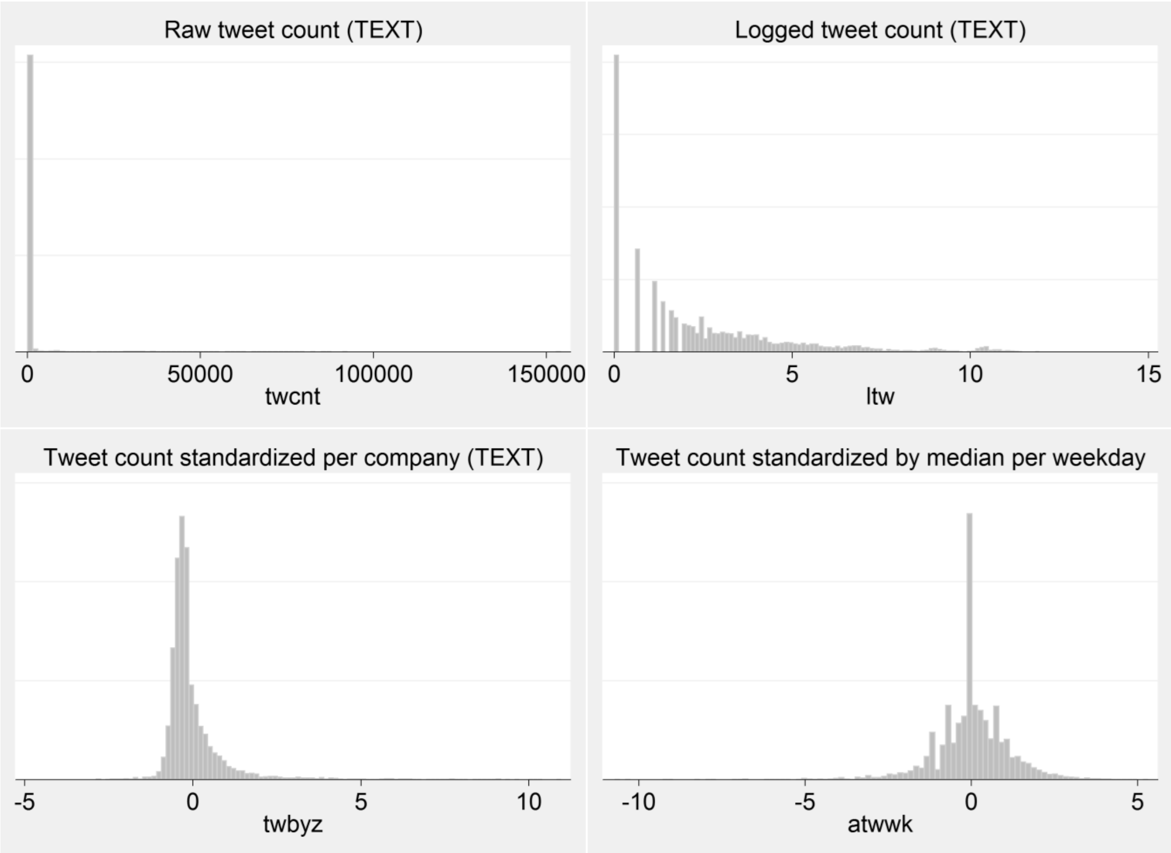

Figure 5-5: Standardization of Twitter activity

This figure presents four histograms for the daily Twitter activity of the full panel data set. The first histogram on top left shows the raw count of Twitter messages per day and company. The second graph shows the natural logarithm of the tweet count. The graph on the bottom left is the result of standardizing the tweets per day and company to a mean of 0 and standard deviation of 1 (twbyz). The fourth graph on the bottom right shows the atwwk measure.

5.4.2 *Trading data*

We obtain intraday trading data from ThomsonReuters Tick History[38] for our observation period from October 20th 2013 to April 30th 2014, which spans 131 trading days. The data contains all updates to the best bid and best ask as well as all trades for all covered stocks, both with their respective price and volume on offer or cleared, respectively.

[38] We thank the Capital Markets Corporative Research Centre for their support in the provision of access to trading data.

5.4.2.1 Preprocessing and cleaning

In a paper published recently, Holden and Jacobsen (2013) demonstrate how a lack of precision and diligence in the processing of intraday data can introduce a heavy bias in the results. As in their study, we also observe the trend of strongly increasing trading volume over the last decade when we compare the data used in the current study from 2014 to the data from 2005 used in the first empirical study in this dissertation. Table 5-4 breaks up descriptives on the trading activity by index membership and deciles based on traded volume. While the differences in trading activity are large even in a sample of stocks composed from the leading indices, the stocks in the lowest decile trade sufficiently frequent to warrant daily or intraday analysis.

Table 5-4: Trading activity on XETRA and default choice of PIN estimation parameters

This table shows the aggregated trading activity of all stocks in the sample, aggregated by index membership in panel A and deciles by trade volume in panel B. Columns two to four show the totals of the number of trades, the traded value and traded volume aggregated over the observation period. The first column of the sections daily traded value and daily trade count is the mean of the daily trade value or trade count. "Max" is the maximum of all daily values. The following two columns approximate the mean time in seconds between two trades by dividing the mean and maximum number of trades per day by the number of seconds per trading day (30600 seconds). The last two columns show the default assignment of the two parameters required for estimating PIN, the length of the no-trade interval in seconds and the length of an intraday bucket in minutes.

	Totals			**Daily traded value per stock** (mio. €)		**Daily trade count per stock**		**Mean time between trades** (sec.)		**Default PIN est. param.**	
	Trade count (mio.)	Value (mio. €)	Volume (mio.)	Mean	Max	Mean	Max	based on mean	based on max	no-trade	bucket length
Panel A: by Index											
DAX	18.99	464,801	12,526	107	924	4,365	28,299	7	1	-	-
MDAX	6.65	54,129	1,902	9	238	1,147	17,598	27	2	-	-
TecDAX	3.20	19,928	1,326	5	75	736	11,293	42	3	-	-
Panel B: by Decile (trade volume)											
1	9.21	271,431	5,182	187	924	6,348	28,299	5	1	5	8
2	5.78	130,785	5,875	90	452	3,985	26,728	8	1	5	8
3	4.16	63,092	1,565	44	313	2,866	16,745	11	2	5	8
4	2.59	26,005	1,138	18	230	1,788	17,598	17	2	10	8
5	1.90	15,839	491	11	238	1,311	7,244	23	4	10	8
6	1.81	12,252	382	8	46	1,245	6,035	25	5	10	8
7	1.47	9,337	637	6	34	1,012	6,562	30	5	20	12
8	1.01	5,449	261	4	31	697	3,310	44	9	20	12
9	0.64	3,642	146	3	35	441	4,111	69	7	30	15
10	0.29	1,026	77	1	9	197	2,417	155	13	30	15
Total	28.84	538,859	15,754	37	924	1,989	28,299	15	1	-	-

We draw on the excellent list in Holden and Jacobsen (2013) of key checks to raw trading data, all of which we validate on our data: There are no negative trade prices or quotes. There are no crossed spreads anytime. We do observe a very negligible number of blocked spreads, i.e. best bid equals best ask. The numbers, however, are so small that we decide not to alter any data (0 out of 112 Mio. quotes in the DAX, 8 out of 39 Mio. quotes in the MDAX and 30 out of 24 Mio. quotes in the TecDAX). On Nasdaq and NYSE, accounting for withdrawn and cancelled quotes combined with using millisecond timestamps leads to a reduction of trades falsely located outside the national best bid and offer (NBBO) by more than 15 percentage points (Holden & Jacobsen, 2013). In our sample, cancellations and corrections have already been applied to the data according to ThomsonReuters[39] and timestamps in our data are accurate to the millisecond, too. A very small share of timestamps found in our data are not unique per stock and day, hence we create an additional variable right after downloading the raw data and before any sorting or data manipulation is done to retain the initial order of quotes and trades. The share of trades we observe outside the spread for DAX and SDAX data (12.2 and 10.4%, respectively) is in the same range as it is in the US (11.2%). Only the numbers for MDAX and TecDAX are a few percentage points higher. The quoted percentage spread lies in reasonable ranges for all indices, with the observed maximum during the observation period being well below 3% for DAX and MDAX, 8% for TecDAX. The mean (equal-weighted) quoted spread is 0.045%, 0.140% and 0.282%, respectively. We further validated our retrieved data by comparing daily aggregates on prices and volume to publicly available data from Yahoo finance and summary statistics provided on Deutsche Boerse’s website.

As we report results also by index membership, changes in index membership during the observation period need to be considered. The DAX did not change its composition. In MDAX, however, GSW Immobilien dropped out on November 27th 2013 after its merger with Deutsche Wohnen, and was subsequently replaced with SGL Carbon, moving up from the SDAX. For our analysis, we treat SGL Carbon as member of the MDAX already from the start of the observation period and delete all data from GSW Immobilien. The TecDAX saw only on change at the very end of our observation period, which we ignore.

[39] The download manager of ThomsonReuters Tick History allows to tick a check box to “apply corrections/cancellations” before downloading the data.

5.4.2.2 Auction volume

Continuous trading on XETRA starts with, ends with and is interrupted midday or in highly volatile conditions by an auction. Based on our data the volume settled within auctions accounts for 21%, 17% and 13% of shares traded daily for DAX, MDAX and TecDAX, respectively. We delete all quotes originating from auctions but keep the volume settled in the auction in our data as a single trade, hence no volume is discarded. While 20% of trading is certainly not negligible, the majority of trading still takes place in continuous trading.

The existence of the auction has some advantages for our research design. The auction at the start of trading is the first chance for investors and traders to incorporate all the information that accumulated over night into prices. Hence, it should diminish the u-shaped pattern observed in intraday trading and make the assumption of equally distributed arrival of information throughout the day, which our PIN model requires, more realistic. For new information created during the day, traders will want to trade immediately and not wait for an auction, hence continuous trading will still account for the large majority of information processing. The midday and end of day auctions are therefore less likely to process a similar amount of information as the opening auction. The microstructure model underlying the PIN variables assumes a market maker who is represented in continuous trading by the collectivity of all limit order traders. During the auction there is no market maker, hence PIN cannot be applied, irrespective of the technical problem to distinguish buys and sells in an auction.

5.4.2.3 Trade classification

Classification of each single trade into a buy or sell is a pre-requisite for the estimation of PIN. We first apply the three pre-classification heuristics as described in chapter 3 to the raw trading data to aggregate trades likely to originate from one single aggressive order that matches several standing limit orders into one trade. The start of trading is identified individually every day and stock by the first trade after the opening auction. If there is no such trade, we assume continuous trading to have started at 09:03:00 latest. This assumption is required to count the no-trade intervals as input for the estimation of PIN. Trading is assumed to run for 8:30 hours until 5:35pm latest.

Trade classification can be simple, theoretically, if three conditions are fulfilled: Access to not only trades but also quotes, all trades occur at prices equal to the prevailing best bid or ask and timestamps (or the order) of trades and quotes are recorded by a perfectly synchronized clock. Unfortunately, inaccuracies in timestamp recording and data transmission as well as execution schemes that allow trades to be executed inside the spread lead to a significant share of trades not located exactly at the best bid or best ask[40]. One way to overcome the former issue is the introduction of a quote delay, a remedy often pursued in empirical studies that deal with intraday data. We simulated reporting delays in steps from several milliseconds up to 5 seconds to evaluate how the share of trades at the spread and outside the spread changes. Figure 5-6 displays results by index membership.

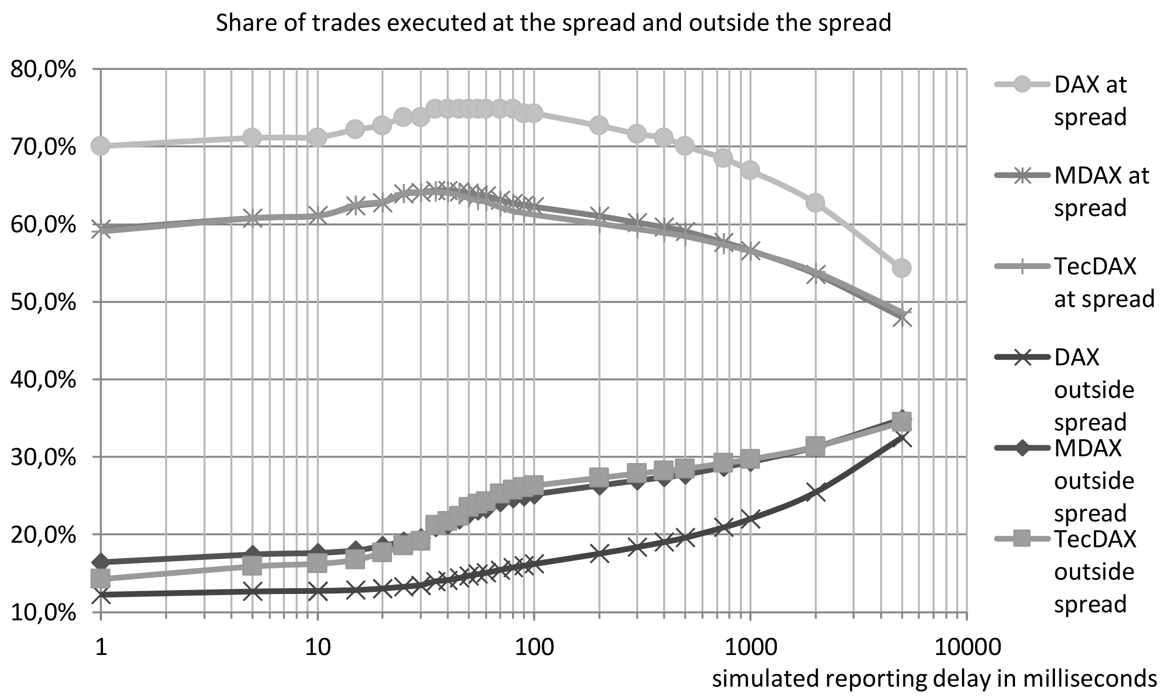

Figure 5-6: Simulated reporting delay

This figure presents results of a simulated reporting delay. The reporting delay is measured in milliseconds on the horizontal axis. The vertical axis shows the share of trades executed exactly at the spread (top three lines) or outside the spread (bottom three lines).

The top three lines in the chart indicate how a reporting delay of 30-60 milliseconds would improve the share of trades executed at the spread. This is an indication, but of course not proof, for the right choice of a reporting delay. On the other hand, introducing a delay also increases the share of trades outside the spread, which is easier to classify but less compelling evidence in favor of a delay as it seems to move trades from within the spread

[40] We discussed the observed trading delay and the share of trades within the spread with representatives of Deutsche Boerse. They provided some explanations in terms of trade execution schemes (e.g. Midpoint Execution) but attribute the larger share of irregularities to inconsistencies arising in the process of extracting, transmitting, transforming, and loading the data from Deutsche Boerse to the data the provider ThomsonReuters.

outside the spread. Given this conflicting evidence, we decide not to introduce a quote delay. Recent evidence recommends a very short 1-millisecond delay, which would not alter our results at all (Holden & Jacobsen, 2013).

5.4.2.4 PIN Estimation

We estimate one PIN per day and stock for different combinations of no-trade intervals (5, 10, 30 and 30 seconds) and the length of a bucket (8, 12 and 15 minutes), to prepare for robustness checks. A reasonable choice for the length of the no-trade interval can be deduced from the average time gap between trades. The last four columns of table 5-4 illustrate our reasoning. The average time between two trades is calculated as the time per trading day, which is 8.5 hours, divided by the average number of trades per day. The results range between 5 and 155 seconds if we split the sample in deciles. This number is obviously an upper bound for the choice of the no-trade interval as trades are not spread uniformly throughout the day. The more relevant lower bound can be proxied as the average time between trades on a day with maximum trading frequency. This number ranges from one second to 13 seconds. In line with these upper and lower bounds, we choose parameter settings for PIN that seem most reasonable given the average trading activity of the stocks per decile. We do of course produce results and validate robustness for all estimated combinations of PIN, but for the sake of brevity and readability will display and discuss results only for the most reasonable default choice as listed in the last two columns of table 5-4. We also run a PIN estimation with a reporting delay of 20 milliseconds. But as results differ by less than half a percentage point we do not investigate these differences any further.

The large number maximum likelihood estimations (one per stock per day per bucket/no-trade combination) require several days of computing time (232 hours in total). A dedicated computer with a powerful processor and, most importantly, 32 GB of RAM allows for fast in-memory calculations of the whole sample. The whole process from reading in raw data, pre-processing and cleaning the sample, classifying trades, aggregating into buckets and estimating the intraday PIN runs with automated scripts written and run in Stata 12.

Figure 5-7 depicts histograms for the four parameters of the model from which PIN is composed and a histogram of PIN itself. All parameters roughly resemble a normal to log-normal distribution without visibly clustering around extreme solutions. PIN is clearly skewed to the right and is lower in the 2013/2014 observation period than in 2005, as estimated in chapter 3. We omit a detailed discussion on PIN results at this point but can confirm the findings from the chapter 3: PIN and alpha increase from high to low volume stocks. The average of delta is very close to 50%. The arrival probability for the uninformed traders, epsilon, increases in line with the length of the no-trade interval. The arrival probability for the informed trades, mu, is also increasing but does so less strongly than epsilon, hence the decreasing PIN from high to low volume stocks.

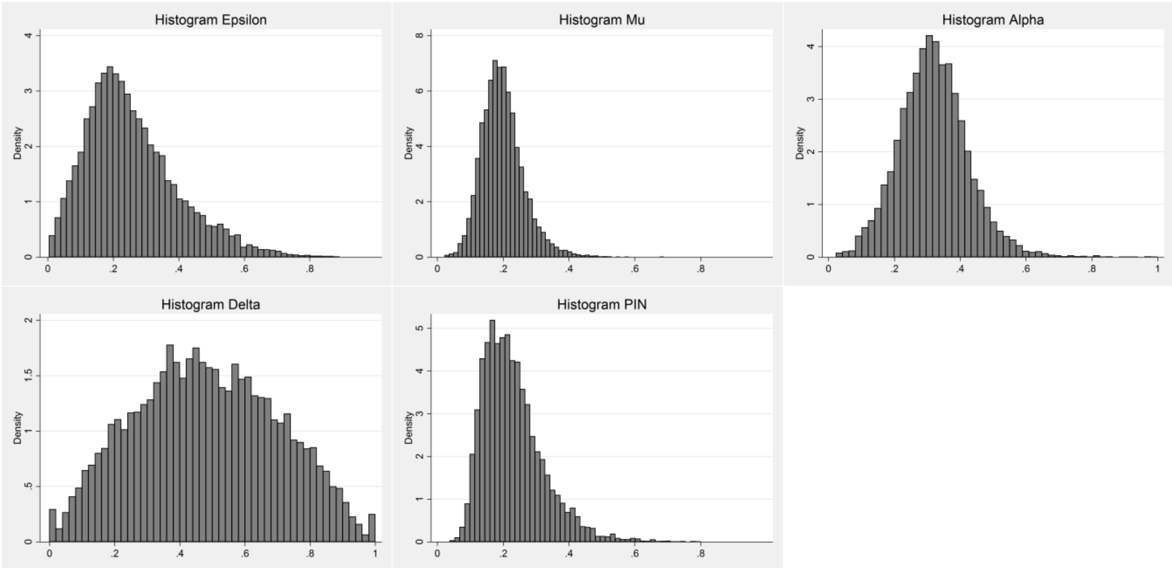

Figure 5-7: PIN histograms

This figure shows histograms of the four parameters of the PIN model and PIN itself. Each estimation is run per stock and trading day.

5.5 Results

5.5.1 Univariate analysis

We designed four different ways to analyze the univariate relationship between activity on Twitter and indicators of trading activity and information dissemination. The results are very consistent across the four approaches and hence we discuss them by topic or variable across tables. First, we focus on the full sample analysis in table 5-5 and table 5-7 and afterwards use table 5-6 and table 5-8, which contain the by-company analysis, to validate our interpretation. All tables examine the contemporaneous relationship and also look one day forward and one day backward to distinguish whether tweet activity leads trading or is triggered by abnormal trading activity.

Table 5-5 displays results for univariate correlations (pearson and spearman) of the full sample, for both Twitter variables *atwwk* and *twbyz*. Given that the full sample contains more than 10,000 observations and we are looking at 25 independent variables, achieving statistical significance at the 5% or even 10% level should not be the benchmark in this case. Instead, the table displays only those coefficients that are significant at the 1% level and indicates those with a star that are significant at the 0.1% level. Table 5-6 aggregates results for univariate correlations by single company by averaging the correlation coefficients and counting for each variable pair how many of those are statistically different from zero at the 10% and 1% level. Table 5-7 display results of the non-parametric Mann-Whitney test on differences in means and the Moods-Median-test on differences in medians between days with Twitter activity in the bottom tercile and days with Twitter activity in the top tercile. Table 5-8 presents results for the same methodology as in table 5-7, but results are calculated per single company and then aggregated, as in table 5-6.

Let us start with the contemporaneous relationships. A positive and statistically significant correlation of Twitter activity and trading volume is consistently present in every analysis. In both the quantile comparison and the spearman rank correlation both variables of Twitter activity react highly significant to volume at the 0.1% level and even lower (not displayed). Only the full sample linear correlation for the variable *atwwk* is not significantly correlated with trading volume. The average difference in trading volume on days with bottom tercile Twitter activity and days with top tercile activity is 63% (table 5-7). In summary, higher activity on Twitter certainly goes along with increased trading volume, a very weak but first indicator for information processing taking place through Twitter.

Results for volatility are the strongest of all evaluated variables in the full sample analysis (table 5-5 and table 5-7). Regardless which of the three measures for volatility one is considering, Twitter activity is positively correlated with higher volatility. The quantile analysis in table 5-7 suggests an average increase of 10-12% on days in the top tercile of Twitter activity.

Results for the spread are not as consistent and unambiguous. The full sample correlation with the effective spread is positive and significant. An elevated effective spread indicates a higher degree of information asymmetry or the presence of information being processed by market participants, which would support the notion that Twitter helps or at least is an indicator of information dissemination taking place. Surprisingly, however, the tercile split is insignificant, whereas all other variables show strongest results in this analysis.

Table 5-5: Pairwise correlations for daily Twitter activity and trading on XETRA (full sample)

This table shows pairwise (rank) correlations coefficients between Twitter activity and indicators of trading activity. Both metrics for Twitter activity atwwk and twbyz are shown. Only correlations that are significantly different from zero at the 99% confidence level are reported. Correlations significant at the 99.9% level are denoted with a star (*).

n = 10481		Same day				Lag 1 day				Future 1 day			
		corr		spearman		corr		spearman		corr		spearman	
Group	**Variable**	atwwk	twbyz	atwwk	twbyz	atwwk	twbyz	atwwk	twbyz	atwwk	twbyz	atwwk	twbyz
Volume	lnVolume		.071 *	.035 *	.080 *		.035 *		.055 *		.029		.062 *
Volatility	hlVola	.068 *	.127 *	.071 *	.054 *		.036 *			.049 *	.052 *	.061 *	.036 *
	15minVola	.073 *	.127 *	.066 *	.050 *		.038 *			.048 *	.052 *	.055 *	.029
	15minVolaMax	.079 *	.122 *	.065 *	.054 *		.030			.053 *	.052 *	.056 *	.036 *
Spread	effspread	.078 *	.059 *	.069 *		.054 *		.047 *		.074 *	.040 *	.086 *	
	qspread				-.028		-.028		-.034 *	.030		.033 *	
	qspreadeqw	.049 *		.030	-.031	.034 *			-.043 *	.050 *		.060 *	-.031
Small-Large Trades	xt3share	-.088 *	-.096 *	-.105 *	-.102 *	-.065 *	-.052 *	-.081 *	-.078 *	-.054 *	-.045 *	-.080 *	-.072 *
	xt5share	-.075 *	-.085 *	-.091 *	-.090 *	-.052 *	-.046 *	-.073 *	-.073 *	-.045 *	-.042 *	-.073 *	-.064 *
	xt3lnratio	-.095 *	-.105 *	-.115 *	-.110 *	-.074 *	-.063 *	-.091 *	-.086 *	-.063 *	-.050 *	-.089 *	-.080 *
	xt5lnratio	-.098 *	-.111 *	-.113 *	-.108 *	-.078 *	-.067 *	-.092 *	-.089 *	-.064 *	-.048 *	-.089 *	-.079 *
	slt50		.082 *	.035 *	.084 *		.039 *		.054 *		.035 *		.064 *
	slt500		.073 *	.045 *	.081 *		.036 *		.054 *		.033 *		.060 *
	slt500p		.071 *		.046 *								.021
	slt100		.082 *	.035 *	.084 *		.039 *		.054 *		.035 *		.064 *
	slt1mio		.073 *	.045 *	.081 *		.031		.047 *		.031		.060 *
	slt1miop		.044 *		.030							-.030	
Return	retln	.036 *		.025	.027					.031			
	retNeg	-.030	-.083 *										
	retPos	.089 *	.115 *	.040 *	.037 *					.070 *	.049 *	.027	
PIN	PIN		-.053 *		-.059 *		-.025		-.047 *				-.036 *
	mu	.067 *	.105 *	.064 *	.070 *	.041 *	.054 *	.037 *	.035 *	.054 *	.062 *	.056 *	.053 *
	epsilon	.030	.105 *	.054 *	.090 *		.060 *	.030	.061 *		.048 *		.064 *
	delta	-.038 *	-.032	-.036 *	-.040 *		-.033 *			-.029		-.030	-.034 *
	alpha												

Table 5-6: Pairwise correlations per single company for daily Twitter activity and trading on XETRA (atwwk)

This table shows the number of correlations calculated per company that are significant on a 5% and 1% level. Correlations are calculated between standardized daily Twitter activtiy and indicators of trading activity on XETRA. The total number of companies and hence the maximum possible count is 83. Results for both Pearson correlation and the Spearman rank-correlation are displayed. The third column of each section displays the average of the 83 correlation coefficients.

n = 83 Companies		Same day						Lag 1 day						Future 1 day					
		Pearson			Spearman			Pearson			Spearman			Pearson			Spearman		
Variables		Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01
Volume	lnVolume	.169	54	35	.160	51	30	.083	38	16	.101	35	11	.115	35	19	.123	39	14
Volatility	highlowVola	.089	36	15	.082	21	12	.019	22	6	.028	16	6	.064	26	10	.068	20	6
	15minVola	.102	37	18	.076	24	11	.013	20	9	.020	15	8	.062	26	10	.061	21	8
	15minVolaMax	.100	26	14	.080	22	8	.022	16	2	.021	21	3	.066	20	9	.063	20	6
Spread	effspread	.092	26	7	.095	29	10	.040	16	4	.044	18	5	.081	28	7	.083	27	9
	qspread	-.013	15	3	-.010	12	3	-.029	13	3	-.018	10	1	-.005	15	4	-.009	13	3
	qspreadeqw	.025	32	10	-.008	28	8	-.057	30	7	-.047	24	8	.011	30	10	-.006	29	9
Small-Large Trades	xt3share	-.097	32	18	-.103	35	17	-.075	25	11	-.080	28	10	-.061	21	11	-.079	25	13
	xt5share	-.079	25	16	-.086	30	17	-.058	21	8	-.068	29	10	-.048	20	8	-.068	24	10
	xt3lnratio	-.110	32	20	-.115	28	21	-.090	31	14	-.094	36	13	-.074	28	10	-.090	32	14
	xt5lnratio	-.111	32	18	-.113	28	19	-.092	30	14	-.093	31	12	-.072	29	9	-.088	32	15
	slt50	.141	50	33	.132	41	21	.056	34	16	.078	28	8	.097	32	15	.102	34	11
	slt500	.160	54	26	.158	48	21	.080	25	9	.094	25	7	.106	32	11	.116	34	13
	slt500p	.077	18	7	.065	14	3	.031	10	2	.033	13	2	.037	12	4	.034	12	2
	sl100k	.142	49	34	.133	43	21	.057	34	15	.079	27	8	.098	32	15	.103	34	12
	sl1mio	.142	43	18	.137	40	18	.065	20	7	.073	20	3	.096	26	8	.102	29	9
	sl1mioplus	.051	12	3	.043	13	1	.026	8	1	.026	13	2	.017	8	2	.012	6	2
Return	retln	.044	20	4	.029	15	3	-.002	8	0	-.020	9	0	.038	13	2	.020	14	1
	retNeg	-.022	23	9	.003	11	3	-.007	10	3	-.025	8	3	-.011	14	1	.000	11	2
	retPos	.096	31	8	.045	14	1	.003	8	0	-.015	6	0	.074	21	3	.029	14	2
PIN	PIN	-.055	19	4	-.044	17	4	-.031	16	2	-.030	17	3	-.029	15	5	-.024	21	5
	mu	.074	26	8	.067	19	5	.043	14	2	.042	14	2	.059	23	5	.059	19	8
	epsilon	.106	43	26	.101	34	16	.049	32	8	.066	23	6	.064	33	9	.074	23	7
	delta	-.038	15	1	-.038	16	2	-.014	12	4	-.016	16	4	-.029	11	0	-.031	10	4
	alpha	-.004	7	1	.013	8	2	.001	7	1	.007	8	0	-.005	9	1	.011	9	1

Table 5-7: Differences in trading indicators - top vs. bottom tercile of Twitter activity

This table display compares indicators of trading activity on days in the top tercile of Twitter activity to days in the bottom tercile of Twitter activity. Terciles are calculated per single stock over the whole observation period based on the variable *atwwk*. The first three columns compare trading and Twitter activity on the same day. The following columns base the comparison on Twitter activity of the previous day (lagged), the last columns are based on the following day's Twitter activity. In each section "MW" is the p-value of a Mann-Whitney mean comparison, "MM" the p-value of a Moods-Median comparison and "Delta Avg" is the difference of the averages in the bottom vs. the top tercile.

Terciles based on atwwk top vs. bottom		**Same day**			**Lag 1 day**			**Future 1 day**		
		sig. of difference			sig. of difference			sig. of difference		
Variable		MW	MM	Delta Avg	MW	MM	Delta Avg	MW	MM	Delta Avg
Volume	lnVolume	.	.	63%	.	.	49%	.	.	53%
Volatility	15minVola	.	.	12%	.045	.091	3%	.	.	10%
	15minVolaMax	.	.	10%	.005	.022	3%	.	.	8%
	highlowVola	.	.	11%	.005	.1	4%	.	.	10%
Spread	effspread	.806	.749	0%	.007	.235	-4%	.265	.438	-1%
	qspread	.	.	-11%	.	.	-12%	.	.	-9%
	qspreadeqw	.	.	-10%	.	.	-12%	.	.	-10%
Small-Large Trades	xt3lnratio	.	.	-9%	.	.	-7%	.	.	-7%
	xt3share	.	.	-4%	.	.	-3%	.	.	-3%
	xt5lnratio	.	.	-11%	.	.	-9%	.	.	-8%
	xt5share	.	.	-5%	.	.	-3%	.	.	-3%
	slt50	.	.	41%	.	.	31%	.	.	36%
	slt500	.	.	72%	.	.	53%	.	.	58%
	slt500p	.	.	8%	.	.	4%	.	.	5%
	sl100ks	.	.	42%	.	.	32%	.	.	36%
	sl1mio	.	.	52%	.	.	36%	.	.	41%
	sl1mioplus	.	.	3%	.004	.004	1%	.114	.122	1%
Return	retln	.004	.049	0%	.092	.115	0%	.076	.227	0%
	retNeg	.965	.	0%	.028	.	0%	.754	.	0%
	retPos	.	.02	0%	.222	.162	0%	.003	.13	0%
PIN	PIN	.	.	-8%	.	.	-7%	.	.	-6%
	mu	.	.	5%	.	.	3%	.	.	5%
	epsilon	.	.	16%	.	.	13%	.	.	14%
	delta	.001	.007	-3%	.158	.443	-2%	.022	.009	-2%
	alpha	.045	.021	1%	.061	.066	1%	.064	.108	1%

Table 5-8: Quantile comparison for daily Twitter activity and trading on XETRA (per Company)

This table shows the number of stocks for which the difference in the mean between the top and bottom tercile trading days is significant at the 10% or 1% level. The tercile split is based on each stock's individual Twitter activity. The first three columns compare trading and Twitter activity on the same day. The following columns base the comparison on Twitter activity of the previous day (lagged), the last columns are based on the following day's Twitter activity. Results for two measure of abnormal Twitter activity, *atwwk* and *twbyz*, are shown.

n = 83 Companies		**Same day**				**Lag 1 day**				**Future 1 day**			
		atwwk		twbyz		atwwk		twbyz		atwwk		twbyz	
Variables		10%	1%	10%	1%	10%	1%	10%	1%	10%	1%	10%	1%
Volume	lnVolume	49	30	49	27	30	9	32	7	34	13	36	17
Volatility	hlvola	25	11	26	9	17	7	16	8	21	6	21	8
	15minVola	25	10	24	11	18	8	17	7	20	8	24	10
	15minVolaMax	22	5	22	6	17	3	17	5	16	7	19	5
Spread	effspread	28	8	27	9	14	2	13	2	20	6	19	5
	qspread	13	3	11	2	11	2	12	1	13	3	10	0
	qspreadeqw	28	5	27	5	22	5	24	5	27	7	26	7
Small-Large Trades	xt3share	31	17	30	17	24	9	22	8	26	13	26	12
	xt5share	32	15	33	13	21	9	24	9	24	11	24	11
	xt3lnratio	38	19	32	16	28	12	23	14	28	13	29	10
	xt5lnratio	34	19	33	17	28	11	30	8	32	12	30	12
	slt50	44	21	40	23	23	6	27	6	32	11	31	10
	slt500	48	23	48	23	28	8	27	4	32	12	30	13
	slt500p	13	2	11	3	9	2	10	2	11	1	8	1
	slt100	44	21	41	23	23	6	27	6	32	11	32	11
	slt1mio	33	19	37	18	17	1	15	1	22	7	26	7
	slt1miop	10	1	13	2	5	1	6	2	4	1	4	1
Return	retln	12	2	11	4	10	0	9	1	14	1	12	1
	retNeg	12	4	11	3	11	1	11	3	11	1	12	1
	retPos	12	3	13	2	8	0	6	1	13	1	14	1
PIN	PIN	19	6	17	4	18	4	18	3	16	4	15	3
	mu	18	4	17	3	11	3	12	1	17	6	18	4
	epsilon	34	13	30	13	23	6	21	6	27	6	30	8
	delta	16	2	16	2	11	3	15	2	10	1	11	1
	alpha	5	1	5	1	8	0	6	0	6	0	5	0

This discrepancy may be driven by a size effect, as spreads depend on liquidity, which varies broadly in our sample and hence differences in spread related to Twitter activity show up in a correlation analysis but not in any quantile split of the sample. The time-weighted quoted spread in turn seems relatively resilient and unaffected in the full sample correlations and each of the analysis run per company. A significant and negative effect of the quoted spread on *twbyz* and *atwwk* in the quantile comparison of the full sample in table 5-7 is counter-intuitive and contrary to the rising effective spread. At the end of the day, however, we would give more weight to the results of the effective spread, as this is the one relevant for investors and captures actual trading, hence a rising spread indicates information dissemination to take place in higher Twitter activity.

We expect the distribution of trade size to differ depending on the level of tweet activity, based on results in Antweiler and Frank (2004). What we find in our results, however, is only partially supportive. The coefficients of the variables *xt3share*, *xt5share* as well as *xt3lnratio*, *xt5lnratio* are all negative and significant at the 0.1% level for their correlation with both Twitter variables. This trend is also confirmed in the quantile analysis. The share of *relatively* smaller trades is declining with rising Twitter volume. On the other hand, if we use the fixed cut-off irrespective of a stocks typical trade size distribution in the same manner as Antweiler and Frank (2004), namely the variables *slt100ks*, *slt1mio* and *slt1mioplus*, the increase in trades smaller than 100,000 Euro is twice as strong as the increase of trades larger than 1 million Euro for correlation with the variable *twbyz*. This trend is also confirmed in the quantile analysis where the differences and the delta in mean or median of the respective variables are highly significant for both *atwwk* and *twbyz*. At the end, the results are dependent on where one wants to draw the line between “retail” investors and institutional investors based on trade size. Considering known strategies of especially high frequency traders to slice their orders in ever smaller parts, trade size seems to be an indicator that is too much biased and ambiguous in today’s trading environments to allow a robust conclusion.

Studying results of the PIN variables and its parameters allows a different view on the types of traders involved. The arrival rates of informed and uninformed traders, *mu* and *epsilon*, in sum represent total trading volume in the underlying model; hence, the correlation is both positive and significant, given the positive correlation of Twitter and trading volume. Neither correlation coefficient is consistently larger for one variable. The quantile comparison, however, on the full sample and the company level, strongly indicates a much larger increase in *epsilon*, i.e. the uninformed traders. The parameter *alpha* reveals one of the most striking results at this stage. Twitter activity is not at all correlated with the probability of information arrival. This is especially remarkable given that every other variable is at least somewhat correlated to Twitter in the full sample.

Joining the findings so far, Twitter volume does not necessarily convey new information, but it does attract uninformed traders, who push up both volume and volatility. The effective spread rises as an indication of uncertainty. The probability of informed trading, *PIN*, decreases, by 8% on average between high and low Twitter activity, and has a negative and significant coefficient for *twbyz*, which further supports this interpretation.

Table 5-6 and table 5-8 break down the full sample analysis on the company level. Overall, no result or interpretation of the full sample analysis is contradicted and all coefficients point in the same direction. But it becomes clear that the observed relationships between Twitter and trading are not consistently valid for all companies across our sample. On the other hand, the averaged correlation coefficients over single company results are much larger than the full sample coefficients, indicating that the relationship, if it exists for a company, is larger and more economically significant than the relatively small coefficients in the full sample suggest.

The contemporaneous, univariate analysis does not allow to draw conclusions on causality. Lagging the Twitter variables by one day and vice versa correlating lagged trading variables with Twitter is our first step in establishing a causal link. The general trend across all four tables is that a predictive correlation is bidirectional and not causal. If lagged Twitter activity significantly correlates with future trading activity, this correlation also shows up significantly when trading leads Twitter. In fact, in some instances trading rather seems to precede Twitter activity than Twitter activity precedes trading. Current volatility is much stronger correlated with future tweet activity than lagged tweet activity with current volatility. Positive returns have some predictive power over higher tweet activity on the next day, but not vice versa. In any case, these results are not sufficient and hence we turn to a multivariate analysis with lagged variables in the following section.

5.5.2 Multivariate analysis

The results of the univariate analysis are strong enough to state that the information flow on Twitter is certainly related to trading activity. The explanatory power and the direction of the information flow, however, are not clear. Does Twitter guide or follow the market?

The panel dataset consists of the cross section of 83 companies and a time-series of 131 trading days. The data set is fully balanced and organized in long form. While we need to control for company fixed effects we find no reason to believe that we have to control for time-fixed effects. The observation period spans just half a year and by the end of 2013 Twitter was already an established social media and news platform. There was also no single significant event that heavily affected all companies like the start of the financial crisis. The degrees of freedom in the time dimension in our cross-sectional dataset shield our regression from an omitted variable bias. Further, a Hausmann test rejects the null hypothesis of random effects and confirms the intuition that a fixed effects model is the appropriate choice. Running a random effects model yields very similar results, though, in terms of direction and significance of coefficients.

Table 9 presents result for fixed-effect panel regressions with the key variables from every dimension of trading as dependent variable. Twitter activity is one of the independent variables, along with several control variables from the univariate analysis. For each dependent variable, one model looks purely at the contemporaneous relationship – the first step forward from the univariate analysis. The second model includes five lags of both the Twitter variable and the dependent variable to ascertain whether any explanatory power in the first model may be substituted by autocorrelation.

In the regression on volume, all control variables including *PIN* and its parameters react significantly in the expected direction, which underscores the general validity of our panel model and also the PIN variation used in this paper. *PIN* declines in increasing volume, the coefficients of both arrival probabilities are significantly positive, *alpha*'s coefficient is positive as the arrival of new information pulls investors to the market and *delta* is insignificant as both positive or negative news trigger volume. Volatility also rises strongly along with volume. These effects change only marginally when lagged values of Twitter and volume are added. Twitter's contemporaneous coefficient (*atwwk*) is significant and positive even when lagged volume is added to the regression. There is, however, none of the predictive power left that was seen in the correlation analysis. None of the lagged Twitter variables is even close to a significant influence, whereas the autocorrelation of volume is visible. This result is irrespective of the choice of Twitter variable (results for *twbyz* are shown in the appendix).

In the third and fourth regression, volatility goes along with increased volume, higher spreads, higher PIN and higher arrival rates of both informed and uninformed, all of which are known relations and hence supportive of our modeling. Surprisingly, we see a significant negative correlation of volatility and Twitter activity lagged by one day, although the standardized coefficients (not shown for brevity) are very small compared the control variables. For the alternative Twitter variable *twbyz*, the positive contemporaneous relation with volatility is significant. Further, the lagged negative influence of Twitter activity on volatility is even stronger than for *atwwk* and lasts for two lags (table appendix 5).

Table 5-9: Regression on trade metrics (atwwk)

This table display results of a fixed-effects panel regression with *lnVolume*, *15minVola*, *effspread* and *retln* as dependent variables. For each of the dependent variables the first regression is run on contemporaneous data without any lags. The second regression is run including 5 lags of the dependent variable and 5 lags of the Twitter variable. Fixed effects are per company and standard errors clustered by date. Significance is indicated by asterisks: * p<0.05, ** p<0.01, *** p<0.001

Group	Variable	lnVolume		15minVola		effspread		retln	
Twitter	atwwk	0.030***	0.031***	-0.001	0.008	0.022	0.016**	0.001	0.001**
	L.atwwk		-0.010		-0.017**		-0.008*		0.000
	L2.atwwk		0.007		-0.009		-0.007		-0.001
	L3.atwwk		0.006		0.003		-0.002		0.000
	L4.atwwk		0.000		-0.004		0.003		0.000
	L5.atwwk		0.005		-0.001		-0.004		0.000
Control	lnVolume			0.225***	0.239***	0.009	0.005	0.000	0.001
	15minVola	0.241***	0.244***			0.239***	0.205***	-0.002	-0.002*
	effspread	0.007	0.013	0.166***	0.155***			-0.003	-0.003
	retln	0.028	-0.017	-0.479	-0.813*	-1.421*	-1.662***		
PIN	PIN	-1.257***	-1.100***	0.270*	0.333**	0.330**	0.216*	-0.001	-0.002
	epsilon	2.521***	2.304***	1.021***	0.689***	-0.344*	-0.178*	-0.011	-0.016*
	mu	2.912***	2.777***	0.339**	0.326**	0.002	-0.107	0.008	0.010
	alpha	1.488***	1.365***	-0.032	-0.075	-0.251**	-0.252***	0.004	0.005
	delta	-0.023	-0.022	0.002	-0.001	-0.068**	-0.043**	-0.025***	-0.025***
Lags dep. var.	L1 dep. var.		0.062***		0.155***		0.254***		-0.056
	L2 dep. var.		0.039***		0.080***		0.168***		-0.047
	L3 dep. var.		0.012		0.038**		0.110***		-0.063*
	L4 dep. var.		0.023*		0.036**		0.092***		-0.030
	L5 dep. var.		0.018		0.015		0.100***		0.019
	Constant	15.97***	13.70***	-8.90***	-7.13***	-5.67***	-0.68*	-0.024	-0.032
Observations		10467	10065	10467	10065	10467	10004	10467	10065
R2-Adj		0.969	0.970	0.622	0.646	0.753	0.821	0.099	0.114
Clustered std. errors		date	date	date	date	date	date	date	date
Firm FE		yes	yes	yes	yes	yes	yes	yes	yes

That also the effective spread rises with Twitter volume despite all control variables and lagged spread being controlled for is the first indication in the multivariate analysis that there may be some additional information on Twitter that is still to be incorporated into prices. This effect is even stronger for the *twbyz* Twitter variable (see table A.10).

The log of each stock's daily raw return is of course the by far least explained dependent variable in our regression models, otherwise market efficiency would be seriously violated. Again, the coefficient of Twitter activity is significant only for the same trading day. The effect is slightly positive, indicating a higher chance of positive news being discussed or spread on Twitter. Interestingly, *delta*, the probability of bad news, has a negative and highly significant coefficient, just as one would expect from the model.

A second set of regressions employs *PIN* and its components as dependent variables in table 5-10. The results are striking. Apart from a few significant coefficients on the 5% level in higher lags, Twitter activity does not add any explanatory power for *PIN* or any of its components. Even the contemporaneous relationship, which was significant in the univariate analysis, vanishes completely in the multivariate setting. Purely their transitive relations with volume or other variables must have driven the previous univariate results between Twitter and PIN.

There is one exception: The contemporaneous coefficient of *alpha* in the regression with the alternative measure of Twitter activity, *twbyz*, as dependent variable (table A.11) is significant at the 0.1% level, but the coefficient is negative. This means that the probability of information arrival is negatively related to the arrival of tweets. On the other hand, the coefficients of the control variables once again confirm the general validity of the regression model and the underlying data. *PIN* decreases with volume, *alpha* increases. The arrival probabilities *epsilon* and *mu* are positively related to both volume and volatility while the coefficients for the effective spread point in the opposite direction - as expected for the arrival of informed and uninformed traders.

In summary, in multivariate settings of fixed-effects panel regressions, the explanatory power of lagged Twitter activity diminishes completely. Further, the contemporaneous relationships are weaker than expected from the univariate analysis. The relation holds for volume and the effective spread. An effect on volatility is supported in the regression with *twbyz* but not *atwwk*. Even stronger is the rejection of Twitter's relevance for information processing when the PIN variable and its parameters are taken as benchmark. Only *alpha* reacts, but does so in the opposite direction, hinting at the conclusion that a burst on Twitter is rather noise than information.

5.5.3 Event study

The sample of events for the event study consist of 110 ad-hoc announcements and 98 earnings announcements for all companies where trading and Twitter data are available. The requirement of an event window from day t-5 up to day t+5 being covered in our data and non-overlapping windows reduces the joined sample of ad-hocs and earnings announcements to 140 events. The analysis of PIN reduces the sample for those variables

Table 5-10: Regression on PIN (atwwk)

This table display results of fixed-effects panel regression with *PIN* and its components *epsilon, mu* and *alpha* as dependent variables. In this table *atwwk* measures Twitter activity. For each of the dependent variables the first regression is run on contemporaneous data without any lags. The second regression includes 5 lags of the dependent variable and 5 lags of the Twitter variable. Fixed effects are per company and standard errors clustered by date. Panel A displays regular coefficients, Panel B standardized coefficients. Significance is indicated by asterisks: * p<0.05, ** p<0.01, *** p<0.001

		PIN		**epsilon**		**mu**		**alpha**	
Panel A: regular coefficients									
Twitter	atwwk	-0.001	-0.001	-0.001	0.000	-0.001	0.000	-0.001	-0.001
	L.atwwk		-0.001		-0.001		0.001		-0.001
	L2.atwwk		0.001		-0.002*		0.000		0.001
	L3.atwwk		0.001		-0.002*		0.000		0.000
	L4.atwwk		-0.002*		0.001		0.000		-0.001
	L5.atwwk		0.000		-0.002*		-0.001		0.000
Control	lnVolume	-0.037***	-0.032***	0.119***	0.106***	0.052***	0.051***	0.017***	0.015***
	15minVola	-0.006	-0.006*	0.037***	0.029***	0.016***	0.017***	-0.005	-0.007
	effspread	0.006	0.004	-0.012***	-0.008**	0.004*	0.004*	-0.010*	-0.009*
	retln	0.094*	0.093*	-0.089	-0.131*	-0.008	0.008	0.102	0.092
Lags dep. var.	L1 dep. Var.		0.165***		0.222***		0.102***		0.113***
	L2 dep. var.		0.030*		0.025*		-0.004		-0.008
	L3 dep. var.		0.017		0.019		0.007		0.023
	L4 dep. var.		-0.003		0.010		0.004		-0.016
	L5 dep. var.		0.019		0.024**		0.026*		0.001
	Constant	0.823***	0.686***	-1.494***	-1.384***	-0.511***	-0.497***	-0.047	-0.068
Panel B: standardized coefficients									
Twitter	atwwk	-.010	-.008	-.010	.002	-.011	-.006	-.014	-.014
	L.atwwk		-.011		-.008		.010		-.009
	L2.atwwk		.012		-.015*		-.002		.011
	L3.atwwk		.012		-.017*		.000		.003
	L4.atwwk		-.020 *		.006		.000		-.012
	L5.atwwk		.004		-.013*		-.010		-.004
Control	lnVolume	-.659***	-.567***	1.403***	1.249***	1.359***	1.224***	.266***	.232***
	15minVola	-.028	-.032*	.120***	.102***	.114***	.124***	-.024	-.031
	effspread	.045	.034	-.058***	-.038**	.038*	.041*	-.066*	-.059*
	retln	.018*	.019*	-.011	-.018*	-.002	.002	.017	.016
Lags dep. var.	L1 dep. var.		.165***		.223***		.102***		.113***
	L2 dep. var.		.030*		.025*		-.004		-.008
	L3 dep. var.		.017		.020		.007		.023
	L4 dep. var.		-.003		.010		.004		-.016
	L5 dep. var.		.018		.024**		.027*		.001
Observations		10467	8708	10467	8708	10467	8708	10467	8708
R2-Adj		54.9%	54.5%	84.6%	85.8%	40.5%	41.5%	5.4%	5.5%
Clustered std. errors		date	date	date	date	date	date	date	date
Firm FE		yes	yes	yes	yes	yes	yes	yes	yes

further, as the maximum likelihood estimation does not converge on every trading day. Even though the conversion rates are above or close to 90% the requirement of valid PIN estimates on every day in the event window leaves us with 98 events with PIN results. The development of averages is illustrated in figure 5-8 while table 5-11 presents t-tests for the differences in mean between successive event windows and also a non-parametric Kruskal-Wallis test for a long -5 to +5 event window and a short -1 to +1 event window.

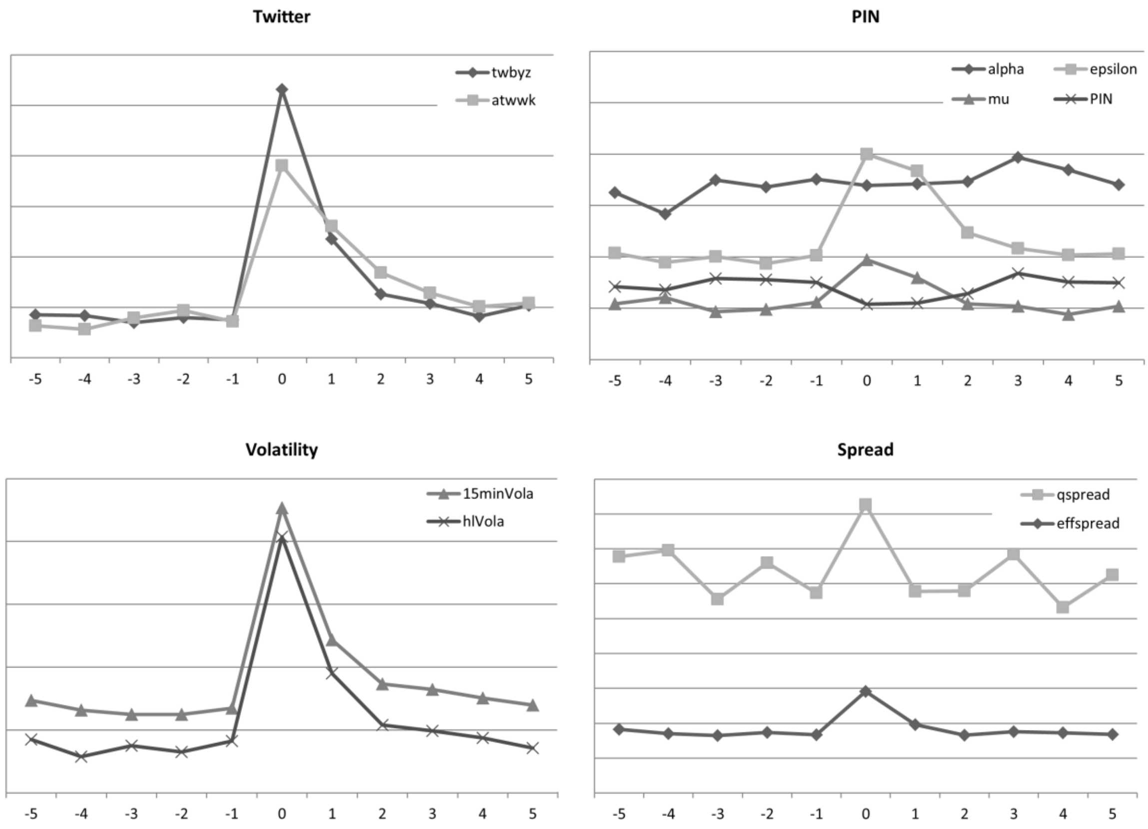

Figure 5-8: Event study

This figure depicts the evolution of relevant metrics in the event study around ad-hoc and earnings announcements. Clockwise from upper left, the first graph shows the two metrics to measure Twitter activity, atwwk and twbyz, the second graph shows PIN and its components epsilon, mu and alpha, the third graph shows two measures for volatility, 15minVola and hlvola, the final graph shows two measures for the spread, qspread and effspread.

The reaction on the announcement day is clearly visible for every variable. Twitter activity, volatility, spread, arrival probabilities (and volume, not shown in the graphs) all rise steeply on day zero. PIN declines due to the publication of previously private information. All those jumps on day zero are statistically significant at the 0.1% level, except for the quoted spread.

Table 5-11: Event study with ad-hoc and earnings announcements

This table displays results of an event study around ad-hoc and earnings announcements. The first eight columns display t-statistics to test difference in means on successive days in an event window from t-4 to t+4. The final four columns display the chi-square statistic and the resulting p-value from a Kruskal-Wallis test on difference across all days in an event window, once for a t-5 to t+ 5 window and also for a smaller t-1 to t+1 event window.

		t-test on successive event days								Kruskal-Wallis			
										(-5, 5)		(-1, 1)	
Variable		**(-4, -3)**	**(-3, -2)**	**(-2, -1)**	**(-1, 0)**	**(0, 1)**	**(1, 2)**	**(2, 3)**	**(3, 4)**	**chi2**	**p**	**chi2**	**p**
Trading and Twitter data for every day in the event window: n = 140													
Twitter	atwwk	-.83	-.57	.93	-10.56 ***	3.92 ***	3.46 ***	1.47	.92	210.3	.0	112.5	.0
	twbyz	.76	-.68	.27	-9.90 ***	5.89 ***	3.98 ***	.92	1.40	216.7	.0	117.7	.0
Volume	lnVolume	-.15	.36	-.40	-4.02 ***	1.80 *	1.31	.42	-.14	37.5	.0	18.5	.0
Volatility	hlVola	-.71	.42	-.75	-8.95 ***	5.13 ***	2.76 ***	.35	.45	139.5	.0	72.0	.0
	15minVola	.28	-.01	-.41	-8.77 ***	5.09 ***	2.30 **	.33	.52	123.6	.0	70.6	.0
	15minVolaMax	.39	.65	-.19	-7.72 ***	4.39 ***	1.98 **	.21	.38	94.4	.0	53.9	.0
Spread	effspread	1.56	-1.11	.92	-2.61 ***	2.54 **	-.02	-1.19	1.78 *	16.6	.084	9.8	.008
	qspread	.17	-.18	.15	-1.96 *	1.94 *	.06	-.04	.05	6.9	.736	6.0	.049
PIN estimation for every day in the event window: n = 98													
PIN	PIN	-1.10	.12	.28	2.46 **	-.11	-1.01	-1.75 *	.73	17.1	.072	12.3	.002
	mu	1.51	-.26	-.72	-3.73 ***	1.75 *	3.01 ***	.25	.89	55.1	.0	29.7	.0
	epsilon	-.31	.36	-.48	-4.91 ***	.72	3.06 ***	.90	.42	64.4	.0	30.3	.0
	delta	-1.90 *	-.57	1.42	1.45	.30	-1.58	-.11	-.51	19.9	.03	1.9	.384
	alpha	-2.32 **	.49	-.54	.42	-.11	-.17	-1.58	.85	16.8	.08	1.2	.559

Judging from the t-statistics[41], activity on Twitter reacts strongest on the announcement day. This indicates a strong link of Twitter to the processing of information which is not just chatter but also relevant for stock markets. However, two points are more interesting than a strong reaction on the announcement day. First, no variable, especially not Twitter activity, shows even the slightest reaction on the day before the announcement. There are, of course, ways to dig way deeper into Twitter data with filtering, sentiment and other methods, but looking at the full picture as we do, there is no indication of a pre-event discussion or rumors, least predictive power. Second, the reversal to a normal level of Twitter activity starts with two days of statistically significant jumps downwards and another two days of a slower decline until Twitter activity returns to its pre-event average. In contrast, volume and spread take just one day for their reversal. Volatility, like Twitter, also takes longer to calm down from the news shock to pre-event levels. The slower reversal in Twitter activity suggests that information diffusion is not as instantaneous as expected from the jump and drop in trading volume, but is spread over several days after an announcement. Overall, we find these results to be a strong indication that information processing on Twitter is related to trading and valuable information moving prices, but it is rather a post-processing platform for the dissemination and interpretation of news than a platform that generates news in the first place. When traders in the market place already incorporated the news into trading, users on Twitter seem to still discuss its proper interpretation.

5.5.4 Intraday event study with ad-hoc announcements

Fair enough and not too surprising, one could say, Twitter is an instant, real-time media, how can you expect to discover its true informational value when looking at data aggregated at the daily level? Hence, we descend one level deeper and analyze the intraday reaction of Twitter around ad-hoc announcements in ranges of several minutes and the accuracy of seconds. We do so in two steps. First, we look at the FROM tweets to understand how companies use Twitter as dissemination channel for their corporate news in comparison to the established ad-hoc channel. Second, we run a descriptive intraday event study around the ad-hoc announcements to see the immediate reaction of all tweet activity with reference to the publishing company (TEXT).

In this first descriptive analysis, 87 ad-hoc announcements belong to companies with a valid company Twitter account. Of those 87 announcements, only 39 are surrounded by at

[41] A non-parametric Mann-Whitney test produces almost identical results to the t-test.

least one tweet of the respective company within an hour around the time of the ad-hoc announcement. That is, for more than half of the companies, the ad-hoc news was not (re-) published on Twitter by the company at all within a reasonable timeframe. Upon manual inspection of those 39 ad-hoc announcements and all corresponding tweets released from the respective company within one hour before and after the release of the ad-hoc announcement, we can find tweets that match the content of the ad-hoc news for 27 announcements. Figure 5-9 maps the matched tweets on a timeline relative to the ad-hoc release time. No tweet is released before an announcement. This result is not too surprising, as companies are legally obliged to use the ad-hoc channel as the first channel for these kinds of announcements. However, all tweets replicating or referring to the ad-hoc announcement are released with a considerable delay of several minutes, on average 18 minutes; the shortest delay is 1:41 minutes. Given the known relevance and price impact of ad-hoc announcements (Muntermann & Guettler, 2007), relying on Twitter alone, notwithstanding the share of companies who do not use Twitter as a secondary release channel for ad-hoc announcements, is clearly not ideal. An investor watching for relevant news on a company's Twitter channel may not become aware at all, or, if it is published, is informed on average more than a quarter of an hour later.

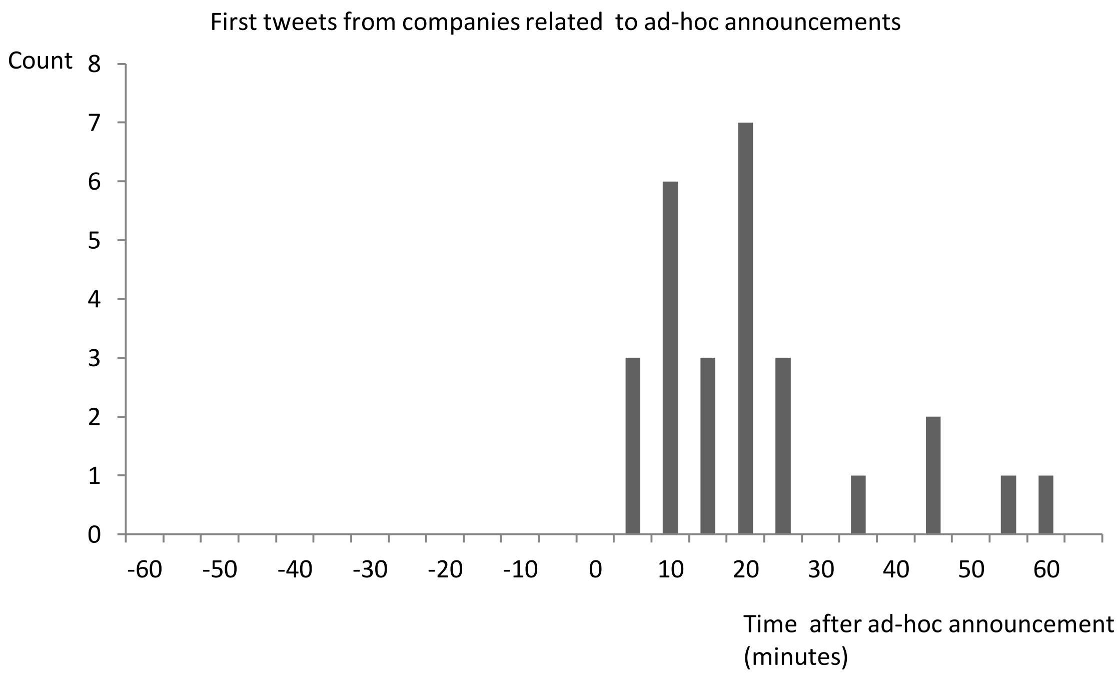

Figure 5-9: FROM-tweets around ad-hoc announcements

This histogram depicts the time it takes until the first tweet covering the content of an ad-hoc announcement is released from the company Twitter account. The sample covers 27 announcements.

It is positively to note, however, that we could also see how some companies not only release a tweet that replicates the ad-hoc release but also provide supplementary information in subsequent tweets (e.g. Munich Re on 7.11.2013 tweets key results from their press conference, which was kicked off by an ad-hoc statement).

In a second approach to intraday analysis, we return to the use of the TEXT tweets, which have been at the center of empirical analysis for most parts of this paper, to look at the overall tweet volume around ad-hoc announcements. In this case, the sample of ad-hoc announcements from the 83 companies with Twitter data comprises 116 events. For 99 of those announcements, at least one tweet occurs within an hour around the announcement time. The timestamps of ad-hoc announcements are accurate to the second, hence we are able to distinguish for each tweet whether it was published before or after the announcement. In similar fashion to the previous analysis, the simple descriptive results of a histogram alone provide already persuasive evidence. The variables *twbyz* and *atwwk* are not applicable for the intraday timeframe. To achieve an equal-weighted presentation in the histogram that accounts for the skewed distribution in tweet activity per company, we standardize the number of tweets per company by weighting each single tweet relative to a multiple of the maximum number of tweets per company. Thereby we can standardize the discrete events to display them in the desired histogram. Figure 5-10 presents results.

Twitter activity jumps significantly immediately within the first minute *after* the release of an ad-hoc announcement. The activity in the hour before the announcement is completely flat, noise, without the slightest indication of a pre-event run-up even in the very minute before the announcement release. Tweet activity stays elevated within the observation period of an hour. For us, this picture completes and validates the event study that is run with daily granularity. Twitter acts as post-processing and information dissemination platform and there is no indication of Twitter feeds being able to predict or forecast news events.

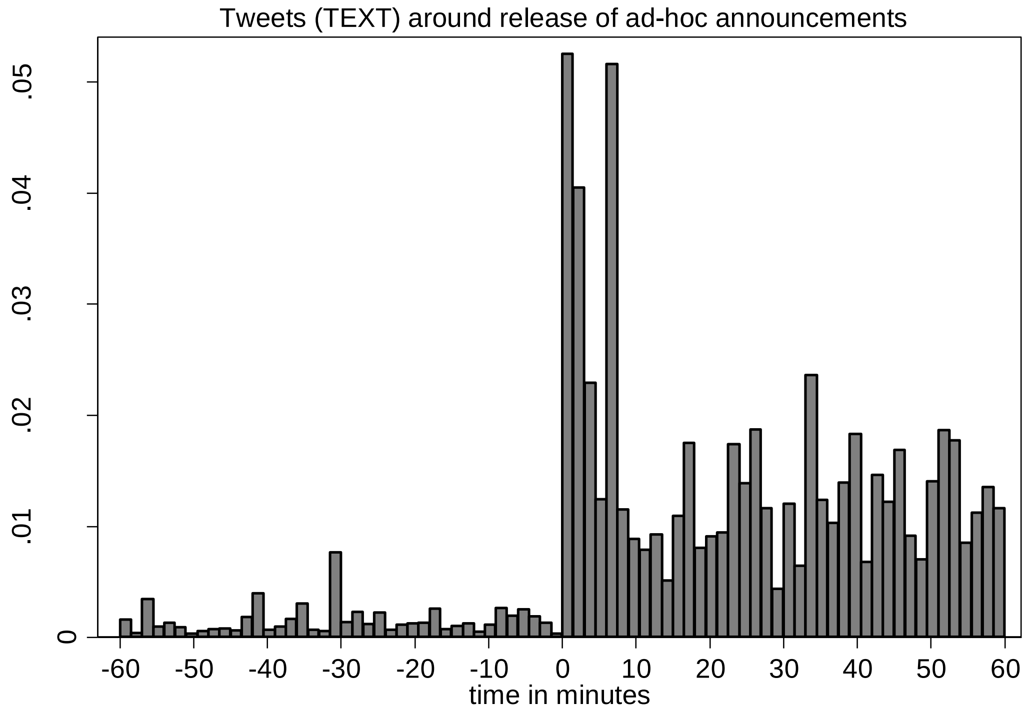

Figure 5-10: Overall tweet activity around ad-hoc announcements

This histogram depicts the tweets published within one hour before and after the release of an ad-hoc announcement by the respective company. The sample covers 99 announcements.

5.6 Conclusion

We started this chapter by expressing our fascination of how Twitter, a micro-blogging and social media platform, attracts tremendous attention from all around the world and, within few years after its foundation, emerged as an essential institution in the media landscape. An obvious test for Twitter's true informational value, in our view, and also of interest to investors, companies, regulators and researchers alike, is to analyze its interaction with financial markets.

This is the first study to relate a comprehensively exhaustive sample of tweets to a large and diverse sample of companies in a major equity market. We do not filter tweets by stock tickers or semantic classification, nor do we relate Twitter activity "just" to a market wide index. How we start out and also what we find is different compared to the few previously available studies, but relates to and echoes literature on stock message boards and short-term information processing.

Twitter can serve as an excellent indicator of news activity that is relevant for the stock market. However, it is not superior to simply watching the market and known metrics itself. Volume, volatility and spread rise contemporaneously with Twitter, all of which indicate that relevant information is processed or disseminated with the help of Twitter. The PIN model indicates that more uninformed than informed traders rush to the market along with rising Twitter activity. The probability of news arrival, *alpha*, is the only variable that is not correlated with Twitter. In a multivariate analysis with lagged variables, all initial indications of a forward-looking predictive power of Twitter vanish. An event study around major corporate announcement again confirms that Twitter activity is related to trading and valuable information moving prices, but also reveals that it is rather a post-processing platform for the dissemination and interpretation of news, than a platform that generates news in the first place. Intraday analysis confirms this finding. Twitter is post-processing or close to real-time, but it is not "ahead" of traditional news channels. The long reversal after a news spike further supports the conclusion that Twitter helps information dissemination and interpretation *after* the release of genuinely new information.

For whom and how can observing Twitter be useful? For investors, it can simply point their attention to the right stocks. That is, stocks where new information requires more attention by investors to guide investment decisions. One may not require some advanced algorithm or prediction machine to tell which stocks to buy or sell. Simply pointing to the stocks in the large universe that require attention for intelligent research and information gathering may be enough. Furthermore, given the link between investor attention, uncertainty and volatility documented by Andrei and Hasler (2015), a rise in Twitter activity along with volatility, as documented in the current study, may also indicate higher uncertainty. Regulators will find it reassuring that Twitter does not seem to change the pecking order of company reporting and evidence of informational advantages enabled by Twitter, on a large scale, cannot be identified. On the other hand, for those corporations that are not yet active on Twitter the presented results are yet another argument to embrace this channel to be able to exert control and influence on the information about their company and use it as an effective way to communicate with investors, as they clearly use Twitter for their information gathering.

From another viewpoint, given the absence of investor-aimed filtering or sentiment processing, our results are also surprising. We did not retrieve data from a platform aimed at investors, but the most generic and general platform available today. We did not filter just stock tickers but considered every tweet that just mentioned the name of a company.

We still see a strong link between financial markets and the company-linked activity on Twitter, but less broadly popular stock discussion forums might be the better place to data mine for hints and opinions as discussions are more continuous and also focused than the outbursts on Twitter.

A number of research questions remain for further research. We did not analyze the smaller subset of FROM and TO message in more detail, which can help to understand how companies and investors use Twitter as an additional communication channel between them. One could also slice the company sample based on industry sectors to distinguish companies like Adidas, who are consumer-focused, from other purely business-to-business oriented companies. There are, of course, way more news channels that could be evaluated in comparison to Twitter to understand how price-sensitive information travels between different media players and financial markets.

6 Overall Summary and Conclusion

This dissertation aims to enhance our understanding of information asymmetry and information dissemination in high-frequency capital markets. To do so, we employ methods and models from market microstructure theory to be able to measure actual changes in information asymmetry. Further, we not only employ existing models proven from the literature but also enhance these models to be applicable to our research questions and also enhance the "tool kit" available to like-minded researchers.

As outlined in the introduction, the research in this work is motivated and necessary due to the exponentially growing volume and speed of trading in capital markets together with likewise strongly increasing volume and speed of information generation and dissemination. The trading data processed in the three empirical papers together comprises more than 80 million trades, i.e. 80 million information items with a date, time, price, volume and a stock ticker symbol and about three times as many updates to the best bid and best ask from the top German stocks from the years 2005, 2012 and 2013. While the event study data with traditional news channels such as ad-hoc or earnings announcements easily fits into an Excel worksheet from 1995, the Twitter data employed in chapter 5 comprises more than 50 million tweets at the start of data processing - and this is just a very small snapshot of the available worldwide volume.

Chapter 2 lays the foundation for the following theoretical and empirical work. We introduce the market microstructure model by Easley et al. (1997b): a risk-neutral, competitive market maker trades an asset whose value depends on new information about the asset's underlying company. New information occurs randomly between trading hours. Only a fraction of traders, the informed traders, are aware of the new information and are able to deduct the new true value of the asset. The informed traders only trade when there is new information to trade on. The other share of traders, the uninformed traders, trade purely for liquidity reasons. The market maker now faces an adverse selection problem in setting his bid and ask quote. He always loses if he trades with an informed trader. That is why, apart from inventory cost, he sets a spread, that is, a bid price lower than the ask price. As uninformed investors, on average, buy and sell with the same probability, the market maker earns the spread and can re-coup the losses to informed traders over time. Of course, the market maker also has to adjust his quotes after every trade. Using Bayesian logic, he updates his believes about the existence of an information event, the direction of an event, and the share of informed traders in the market after every trade.

While the key assumptions in this model have been presented earlier by other scholars already, Easley et al. (1997b) are the first to deduct an estimation procedure that allows to estimate the parameters of this trade model purely from the number of buys and sells per trading day. With the right algorithms, these inputs can be computed from the intraday series of trades (and quotes). Hence, this estimation procedure makes the model's parameters and the composite variable "Probability of Informed Trading" (PIN) easily available to researchers. PIN measures the share of informed trading relative to all trading for a stock. With this variable, a sophisticated variable is introduced that can directly measure information asymmetry.

PIN has been employed for numerous research questions. However, PIN lacks applicability in today's market environments for two major reasons. First, the maximum likelihood estimation of PIN's parameters does not converge reliably in recent, high-frequency trading data. Second, the observation period required to compute one PIN observation spans 30 to 50 trading days, which is too long for many research questions.

The first empirical work in this dissertation therefore presents a new estimation procedure that allows computing one PIN per single trading day. The approach is simple, yet results are compelling. We slice the trading day into buckets of several minutes' length, assuming independent arrival of traders and news throughout the day instead of news arriving only overnight. What resembled a trading day in the previous specification is now a timespan of several minutes within one trading day. Results show that the intraday PIN and its single parameters exhibit the previously demonstrated behavior. PIN is lower for the most liquid stocks and higher for the least liquid stocks; information events are more frequent for higher liquid stocks, to name a few of the empirical properties. What is even more striking and important is that the necessary assumptions of independence across information events and arrival of traders prove to be more realistic than for the original approach. Further, the maximum likelihood estimation following the new specification reaches convergence rates of 95% even for the most liquid stocks of the top 100 German stocks. Hence, short event studies with PIN are now possible in today's market environments. We find that the valuable insights to be gained by employing the intraday PIN justify the considerable effort required to estimate the model parameters.

As a first example of the potential applications, we employ the new intraday PIN in an event study around official announcements stipulated in German insider trading legislation. PIN demonstrates a significant decrease of information asymmetry on the day of disclosure for high and medium liquid stocks. According to this evidence, Ad-hoc announcements fulfill their mission of disclosing price-sensitive information to all market participants simultaneously and lower information asymmetry.

In a logical next step forward, we examine an alternative to our intraday PIN in the next chapter: VPIN – the volume-synchronized probability of informed trading. The authors of VPIN share the exact same motivation and goal as we do with the intraday PIN: developing a metric of information asymmetry or "toxicity of order flow" that is applicable in today's high-frequency trading environment. VPIN is an evolution of PIN, as well. The difference can be outlined in three key steps. First, both nominator and denominator of the PIN formula are approximated such that a simple term remains which effectively resembles relative order imbalance, i.e. buy volume minus sell volume divided by total volume. Thereby, the tedious maximum likelihood estimation is circumvented. Second, for calculating VPIN, clock time is replaced with volume time. In other words, one VPIN estimate is calculated over a fixed amount of volume (e.g. 5000 shares as 1/50th of the average daily trading volume), not over a fixed amount of time (e.g. 5 minutes as in our PIN computation). Third, to calculate order imbalances, traditional trade-by-trade classification algorithms are replaced with a heuristic. VPIN is very controversial and has spawned a large amount of critical literature. Proponents say that VPIN can provide a real-time indicator of market toxicity that can predict extreme volatility. Opponents say it does no better than standard alternatives.

It is the last of the three mentioned innovations of VPIN over PIN where our research starts. We wonder why the change from presumably accurate deterministic algorithms to a heuristic is necessary and how this change affects VPIN. For people not familiar with market microstructure, it may sound trivial: Why is it so hard to identify whether a trade is an aggressive buy or an aggressive sell? Why do we care about whether a trade is called a "buy" or a "sell"? There are several reasons. First, most data providers do not include this (relatively simple) flag in the data they provide. Second, in many markets order flow explains contemporaneous price movements, predicts future price movements through liquidity and information effects and is highly auto-correlated. Our extensive literature review as well as our out-of-sample test of proprietary data from Deutsche Boerse indicates that trade-by-trade algorithms reach above 90% accuracy in classifying trades even in high-frequency trading data. Consequently, we evaluate how VPIN results differ when calculated with inputs from different trade classification schemes.

What we find is that VPIN is not robust to the choice of classification algorithm. On every level ascending in complexity and aggregation from raw trade classification, to order imbalance, to VPINs and toxic periods, the choice of trade classification algorithm induces substantial differences in the results. The gap is widening for the higher aggregate metrics. In the detection of so-called “toxic periods”, the major proposed application of VPIN to, for example, trigger a stop in continuous trading or warn investors to review their limit orders in light of a crash, both approaches reach consensus in only 60% of the cases. Further, neither of the approaches is consistently faster or earlier in detecting toxic periods. Regression analysis identifies high return and volume volatility as main contributors to the difference. These are exactly those conditions where the application of VPIN is intended to be most useful.

After we demonstrated our concerns with VPIN, we rely, among other variables, on the intraday PIN introduced in chapter 3 to analyze in chapter 5 how the information stream on Twitter affects information processing on XETRA. The volume of information released on Twitter is exploding and the attention it commands in the wider media is continuously increasing. This spurred our interest to figure out how much Twitter actually matters for financial markets. This particular field of research is relatively young. Previous research focused mainly on sentiment extraction and the relation of some variant of sentiment measure with a broad market index. In contrast, our approach relates Twitter messages to each single company. Further, we do not rely on sentiment extraction but consider all messages related to a certain stock without any filtering applied. Thereby we intend to measure the true attention for a stock on Twitter.

Both descriptive and sophisticated panel analysis support our key results: Twitter can serve as an excellent indicator of news activity that is relevant for the stock market. However, it is not superior to simply watching the market and known metrics itself. Volume, volatility and spread rise contemporaneously with Twitter, all of which indicate that relevant information is processed or disseminated with the help of Twitter. The PIN model indicates that more uninformed than informed traders rush to the market along with rising Twitter activity. Daily and intra-day event studies confirm that Twitter, on average, from the point of view of financial markets, is a fast *post*-processing platform of new information but it is not “ahead” of traditional news channels or generating market relevant news in the first place. Interestingly, activity on Twitter takes one to two days longer to reverse to normal levels than indicators of trading activity which further supports our hypothesis of Twitter being a post-dissemination and interpretation platform.

Let us review the questions raised in the introduction. We can confirm that information asymmetry in the market can be reduced for all market participants by the timed release of previously private information. Ad-hoc announcements do serve this purpose. Private investors and regulators will find this property of ad-hoc announcement reassuring. However, the meager frequency with which these announcements are released, compared to the constant flow of information on Twitter, clearly raises doubts about companies’ willingness to use this channel as desired. Twitter, on the other hand, does not serve this purpose in similar manner, but it seems to help to spread new information to a broader audience and help its interpretation.

Information travels very fast today. Within seconds or minutes after the release of an ad-hoc news does activity jump on Twitter. The strong correlation with volume and the event study analysis confirm the immediate impact on capital markets. Since all metrics of market activity and especially information asymmetry return to their levels mostly within a day, information is also quickly incorporated into prices.

We also learned that, according to our models, it is rather the uninformed who react to presumably new information on channels such as Twitter, whereas the informed do also react on information, but can hide in the stronger increasing flow of informed traders. This empirical finding was conjectured in early theoretical literature and confirmed in different market settings by other scholars as well. Nevertheless, regulators will find it reassuring that Twitter does not seem to change the pecking order of company reporting, as we find no evidence of any types of investors to trade ahead of the release of information. Our strong advice from these results is to take the hype around Twitter with a grain of salt when it comes to information that is relevant for markets. Of course, our results cannot rule out that there is relevant information hidden somewhere in Twitter or that some news might be released preliminary, as anecdotal evidence suggests. However, on average, traditional news channels, research and existing metrics that gauge market activity, volatility or toxicity may be more fruitful to guide investment decisions.

Regarding methodology, our conclusion is mixed. On the one hand, we propose a new way of measuring information asymmetry with an evolution of the estimation procedure for

PIN. This approach is of high importance for researchers but probably less so for practitioners on a day-to-day basis. On the other hand, we have doubts on a promising and fascinating new way of calculating an intraday measure of flow toxicity – VPIN. While the key ideas of switching from clock to volume time and also abolish the tedious maximum likelihood estimation in favor of a simple approximation do make sense intuitively, the approach does not yield consistent results, especially in times of high volume and volatility. Employing VPIN as early warning signal, from either an investor's or a regulator's perspective, seems premature at its current stage of development. Researchers, however, should continue to work on improving and validating VPIN and other metrics of order flow toxicity.

The empirical and methodological results in this thesis provide several avenues for future research beyond the scope of this work. The newly proposed intraday PIN could be further evaluated and applied to different kinds of news streams. A comparison to related, established and more advanced metrics of information asymmetry could evaluate its incremental explanatory power. A test on both a true-positive and true-negative sample of events where information asymmetry is known to rise and also known to fall could eliminate the claim of PIN being purely driven by rising volume or merely resembling order imbalance. We also discussed empirical literature where the new intraday PIN with its shorter time span may help to resolve previously contradictory results, such as the behavior of PIN around earnings announcements.

The debate around VPIN is very intense in the research community and hence the way forward in our view is to firstly agree on a standard set of parameter settings for computing VPIN or, if VPIN requires adjustment for every market, propose a standard, deterministic procedure to be able to compute "benchmark" VPINs for every relevant market. Afterwards, VPIN needs to demonstrate both its incremental predictive power over established metrics for the toxicity of order flow and also prove its accuracy on a hand-classified sample of toxic events.

Our work on Twitter opens up an even larger field of research questions, as we outlined in the previous section. The ignored subset of FROM and TO messages can, for example, help to understand how companies and investors can use Twitter as an additional communication channel for investor and customer relationship management. Incorporating additional explanatory factors such as industry sectors would distinguish companies like Adidas, who are consumer-focused, from other purely business-to-business oriented companies, where different market reactions and investor attention can be expected. Further news channels could be evaluated in comparison to Twitter to understand how price-sensitive information travels between different media players and financial markets. It would also be interesting to run analysis in similar fashion as demonstrated in this work on data from other platforms and channels of social media or user generated content where the participants are inherently more focused on actually discussing information relevant for stock markets to see whether a supposedly better informed audience is able to lead the market better than the very general platform Twitter does.

Overall, we think all three major goals formulated in the introduction are achieved with the presented work. Our results provide insight for the research community and likewise offer important implications for investors, regulators and listed companies.

References

Abad, D., & Yagüe, J. (2012). From PIN to VPIN: An introduction to order flow toxicity. *The Spanish Review of Financial Economics*.

Admati, A. R., & Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies, 1*(1), 3-40.

Aitken, M., & Frino, A. (1996). The accuracy of the tick test: Evidence from the Australian stock exchange. *Journal of Banking & Finance, 20*(10), 17151729.

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

Aktas, N., Debodt, E., Declerck, F., & Vanoppens, H. (2007). The PIN anomaly around M&A announcements. *Journal of Financial Markets, 10*(2), 169-191.

Allen, F., & Gale, D. (1992). Stock-price manipulation. *Review of Financial Studies, 5*(3), 503-529.

Andersen, T. G., & Bondarenko, O. (2013). Comments on "Testing VPIN on Big Data–response to reflecting on the VPIN dispute". *Available at SSRN 2331106*.

Andersen, T. G., & Bondarenko, O. (2014a). Reflecting on the VPIN dispute. *Journal of Financial Markets, 17*, 53-64.

Andersen, T. G., & Bondarenko, O. (2014b). VPIN and the flash crash. *Journal of Financial Markets, 17*, 1-46.

Andersen, T. G., & Bondarenko, O. (2015). Assessing Measures of Order Flow Toxicity and Early Warning Signals for Market Turbulence. *Review of Finance, 19*(1), 1-54.

Andrei, D., & Hasler, M. (2015). Investor Attention and Stock Market Volatility. *Review of Financial Studies, 28*(1), 33-72.

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

Ascioglu, A., Hegde, S. P., & McDermott, J. B. (2008). Information asymmetry and investment–cash flow sensitivity. *Journal of Banking & Finance, 32*, 1036-1048.

Barber, B. M., & 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.

Barclay, M. J., & Litzenberger, R. H. (1988). Announcement effects of new equity issues and the use of intraday price data. *Journal of Financial Economics, 21*, 71-99.

Benos, E., & Jochec, M. (2007). Testing the PIN variable. *Available at SSRN 933211*, 1-22.

Bhattacharya, U., & Daouk, H. (2002). The world price of insider trading. *The Journal of Finance, 57*(1), 75-108.

Boehmer, E., Grammig, J., & Theissen, E. (2007). Estimating the probability of informed trading—does trade misclassification matter? *Journal of Financial Markets, 10*, 26-47.

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

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

Busse, J., & Green, T. C. (2002). Market efficiency in real time. *Journal of Financial Economics, 65*(3), 415-437.

Chakrabarty, B., Li, B., Nguyen, V., & Van Ness, R. A. (2007). Trade classification algorithms for electronic communications network trades. *Journal of Banking & Finance, 31*(12), 3806-3821.

Chakrabarty, B., Moulton, P. C., & Shkilko, A. (2012a). Short sales, long sales, and the Lee–Ready trade classification algorithm revisited. *Journal of Financial Markets, 15*(4), 467-491.

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

Chan, W. S. (2003). Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Financial Economics, 70*(2), 223-260.

Chen, C. W., Pantzalis, C., & Park, J. C. (2013a). Press Coverage and Stock Price Deviation from Fundamental Value. *Journal of Financial Research, 36*(2), 175-214.

Chen, H., De, P., Hu, J., & Hwang, B.-H. (2013b). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*.

Chen, Q., Goldstein, I., & Jiang, W. (2006). Price Informativeness and Investment Sensitivity to Stock Price. *Review of Financial Studies, 20*(3), 619-650.

Collin-Dufresne, P., & Fos, V. (2012). Do prices reveal the presence of informed trading? *National Bureau of Economic Research Working Paper Series, No. 18452*.

Da, Z., Engelberg, J., & Gao, P. (2015). The Sum of All FEARS Investor Sentiment and Asset Prices. *Review of Financial Studies, 28*(1), 1-32.

Da, Z., Engelberg, J., & Gao, P. J. (2011). In Search of Attention. *The Journal of Finance, 66*(5), 1461-1499.

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

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012). Investor Information Demand: Evidence from Google Searches Around Earnings Announcements. *Journal of Accounting Research, 50*(4), 1001-1040.

Duarte, J., & Young, L. (2008). Why is PIN priced? *Journal of Financial Economics, 91*, 119-138.

Dufour, A., & Engle, R. F. (2000). Time and the Price Impact of a Trade. *The Journal of Finance, 55*(6), 2467-2498.

Easley, D., Engle, R. F., O'Hara, M., & Wu, L. (2008). Time-Varying Arrival Rates of Informed and Uninformed Trades. *Journal of Financial Econometrics, 6*(2), 171-207.

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

Easley, D., Hvidkjaer, S., & O’Hara, M. (2010). Factoring Information into Returns. *Journal of Financial and Quantitative Analysis, 45*(02), 1.

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

Easley, D., Kiefer, N. M., & O'Hara, M. (1997b). One day in the life of a very common stock. *Review of Financial Studies, 10*(3), 805-835.

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

Easley, D., Kiefer, N. M., & O’Hara, M. (1996b). Cream- • Skimming or Profit- • Sharing? The Curious Role of Purchased Order Flow. *The Journal of Finance, 51*, 811-833.

Easley, D., Lopez de Prado, M., & O'Hara, M. (2011). The microstructure of the "Flash Crash": flow toxicity, liquidity crashes, and the probability of informed trading. *Journal of Portfolio Management, 37*(2), 118-128.

Easley, D., Lopez de Prado, M., & O'Hara, M. (2012a). Bulk classification of trading activity. *Johnson School Research Paper Series*(8-2012).

Easley, D., Lopez de Prado, M., & O'Hara, M. (2012b). Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies, 25*(5), 1457-1493.

Easley, D., Lopez de Prado, M., & O'Hara, M. (2014). VPIN and the Flash Crash: A rejoinder. *Journal of Financial Markets, 17*, 47-52.

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

Easley, D., & O'Hara, M. (1992). Time and the Process of Security Price Adjustment. *The Journal of Finance, 47*(2), 577.

Easley, D., O'Hara, M., & Saar, G. (2001). How stock splits affect trading: A microstructure approach. *Journal of Financial and Quantitative Analysis, 36*(1), 25-51.

Easley, D., O'Hara, M., & Srinivas, P. S. (1998a). Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. *The Journal of Finance, 53*, 431-465.

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

Ederington, L. H., & Lee, J. H. (1993). How Markets Process Information - News Releases and Volatility. *The Journal of Finance, 48*(4), 1161-1191.

Ellis, K., Michaely, R., & O'Hara, M. (2000). The accuracy of trade classification rules: Evidence from Nasdaq. *Journal of Financial and Quantitative Analysis, 35*(4), 529-551.

Fang, L., & Peress, J. (2009). Media Coverage and the Cross-section of Stock Returns. *The Journal of Finance, 64*(5), 2023-2052.

Fang, L. H., Peress, J., & Zheng, L. (2014). Does Media Coverage of Stocks Affect Mutual Funds' Trading and Performance? *Review of Financial Studies, 27*(12), 3441-3466.

Fernandes, N., & Ferreira, M. A. (2009). Insider Trading Laws and Stock Price Informativeness. *Review of Financial Studies*, *22*(5), 1845-1887.

Finucane, T. J. (2000). A direct test of methods for inferring trade direction from intra-day data. *Journal of Financial and Quantitative Analysis*, *35*(4), 553-576.

Fuller, K. P., Ness, B. F., & Ness, R. a. (2009). Is information risk priced for NASDAQ-listed stocks? *Review of Quantitative Finance and Accounting, 34*, 301-312.

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics, 14*(1), 71-100.

Grammig, J., Schiereck, D., & Theissen, E. (2001). Knowing me, knowing you: Trader anonymity and informed trading in parallel markets. *Journal of Financial Markets, 4*(4), 385-412.

Green, T. C. (2006). The value of client access to analyst recommendations. *Journal of Financial and Quantitative Analysis, 41*(1), 1-24.

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

Hachmeister, A., & Schiereck, D. (2010). Dancing in the dark: post-trade anonymity, liquidity and informed trading. *Review of Quantitative Finance and Accounting, 34*(2), 145-177.

Harris, L. (1989). A day-end transaction price anomaly. *Journal of Financial and Quantitative Analysis, 24*(1), 29-45.

Hasbrouck, J. (1988). Trades, quotes, inventories, and information. *Journal of Financial Economics, 22*(2), 229-252.

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

Heidle, H. G., & Huang, R. D. (2002). Information-based trading in dealer and auction markets: An analysis of exchange listings. *Journal of Financial and Quantitative Analysis, 37*(3), 391-424.

Hillert, A., Jacobs, H., & Müller, S. (2014). Media Makes Momentum. *Review of Financial Studies, 27*(12), 3467-3501.

Holden, C. W., & Jacobsen, S. (2013). Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions. *The Journal of Finance*.

Holthausen, R. W., Leftwich, R. W., & Mayers, D. (1987). The effect of large block transactions on security prices: A cross-sectional analysis. *Journal of Financial Economics, 19*(2), 237-267.

Huang, R. D., & Stoll, H. R. (1997). The Components of the Bid-Ask Spread: A General Approach. *Review of Financial Studies, 10*(4), 995-1034.

Jackson, D. (2013). Estimating PIN for Firms with High Levels of Trading. *Journal of Empirical Finance, 24*, 116-120.

Kang, M. (2010). Probability of information-based trading and the January effect. *Journal of Banking & Finance, 34*, 2985-2994.

Kim, S. T., Lin, J. C., & Slovin, M. B. (1997). Market structure, informed trading, and analysts' recommendations. *Journal of Financial and Quantitative Analysis, 32*(4), 507-524.

Kumar, R., & Popescu, M. (2014). The implied intra-day probability of informed trading. *Review of Quantitative Finance and Accounting, 42*(2), 357-371.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). *What is Twitter, a social network or a news media?* Paper presented at the Proceedings of the 19th international conference on World wide web.

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

Lambert, R. A., Leuz, C., & Verrecchia, R. E. (2012). Information Asymmetry, Information Precision, and the Cost of Capital. *Review of Finance, 16*(1), 1-29.

Lee, C., & Radhakrishna, B. (2000). Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets, 3*(2), 83-111.

Lee, C., & Ready, M. J. (1991a). Inferring trade direction from intraday data. *The Journal of Finance, 46*(2), 733-746.

Lee, C. M. C., & Ready, M. J. (1991b). Inferring trade direction from intraday data. *The Journal of Finance, 46*(2), 733-746.

Lei, Q., & Wu, G. (2005). Time-varying informed and uninformed trading activities. *Journal of Financial Markets, 8*(2), 153-181.

Leung, H., Rose, A., & Westerholm, P. J. (2013). Systematic Trading Behavior and the Cross-Section of Stock Returns on the OMXH*. *Review of Finance*.

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

Mitchell, M. L., & Mulherin, J. H. (1994). The impact of public information on the stock market. *The Journal of Finance, 49*(3), 923-950.

Muntermann, J., & Guettler, A. (2007). Intraday stock price effects of ad hoc disclosures: the German case. *Journal of International Financial Markets, Institutions and Money, 17*(1), 1-24.

Nofer, M., & Hinz, O. (2014a). Are Crowds on the Internet Wiser than Experts? The Case of a Stock Prediction Community. *Journal of Business Economics, 84*(3), 303-338.

Nofer, M., & Hinz, O. (2014b). Using Twitter to Predict the Stock Market: Where is the Mood Effect? *Business & Information Systems Engineering, forthcoming*.

Novy-Marx, R. (2014). Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars. *Journal of Financial Economics, 112*(2), 137-146.

Nyholm, K. (2003). Inferring the private information content of trades: a regime-switching approach. *Journal of Applied Econometrics, 18*(4), 457-470.

O'Hara, M. (1995). *Market Microstructure Theory* (Vol. 108): Blackwell Cambridge, MA.

Odders-White, E. R. (2000). On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets, 3*(3), 259-286.

Patell, J. M., & Wolfson, M. A. (1984). The intraday speed of adjustment of stock prices to earnings and dividend announcements. *Journal of Financial Economics, 13*(2), 223-252.

Peress, J. (2014). The Media and the Diffusion of Information in Financial Markets: Evidence from Newspaper Strikes. *The Journal of Finance, 69*(5), 2007-2043.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies, 22*(1), 435-480.

Peterson, M., & Sirri, E. (2003). Evaluation of the biases in execution cost estimation using trade and quote data. *Journal of Financial Markets, 6*(3), 259-280.

Pöppe, T., Aitken, M., Schiereck, D., & Wiegand, I. (2015). A PIN per day shows what news convey: the intraday probability of informed trading. *Review of Quantitative Finance and Accounting*, 1-34.

Pöppe, T., Schiereck, D., & Zielinski, F. (2014). Das Google-Suchvolumen als Liquiditätsindikator des Aktienhandels: Evidenz vom weltweiten Agrarsektor. *Credit and Capital Markets, 47*(4), 611-640.

Savickas, R., & Wilson, A. J. (2003). On inferring the direction of option trades. *Journal of Financial and Quantitative Analysis, 38*(4), 881-902.

Scholtus, M., van Dijk, D., & Frijns, B. (2014). Speed, algorithmic trading, and market quality around macroeconomic news announcements. *Journal of Banking & Finance, 38*(0), 89-105.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2013). Tweets and trades: The information content of stock microblogs. *European Financial Management (Advance online publication)*.

Tanggaard, C. (2004). Errors in trade classification: Consequences and remedies. *Available at SSRN 420680.*

Tay, A., Ting, C., Tse, Y. K., & Warachka, M. (2009). Using High-Frequency Transaction Data to Estimate the Probability of Informed Trading. *Journal of Financial Econometrics, 7*(3), 288-311.

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

Theissen, E. (2001). A test of the accuracy of the Lee/Ready trade classification algorithm. *Journal of International Financial Markets, Institutions & Money, 11*(2), 147-165.

Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics, 82*(1), 103-133.

Visaltanachoti, N., & Yang, T. (2010). Speed of convergence to market efficiency for NYSE-listed foreign stocks. *Journal of Banking & Finance, 34*(3), 594-605.

Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance, 41*(0), 17-35.

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

Wu, K., Bethel, E. W., Gu, M., Leinweber, D., & Ruebel, O. (2013a). A Big Data Approach to Analyzing Market Volatility. *Available at SSRN 2274991*.

Wu, K., Bethel, E. W., Gu, M., Leinweber, D., & Ruebel, O. (2013b). Testing VPIN on Big Data – response to "Reflecting on the VPIN dispute". *Available at SSRN 2318259*.

Wu, K., Bethel, E. W., Leinweber, D., & Ruebel, O. (2012). Federal Market Information Technology in the Post Flash Crash Era: Roles for Supercomputing. *The Journal of trading, 7*(2), 9-25.

Wuensche, O., Theissen, E., & Grammig, J. (2011). Time and the Price Impact of a Trade: A Structural Approach. *Available at SSRN 968241*.

Yan, Y. (2009). A New Method to Estimate PIN (Probability of Informed Trading). *Available at SSRN 1895364*, 1-50.

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

Zhang, X., Fuehres, H., & 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*(0), 55-62.

Appendix

Table A.1: XETRA preprocessing of trading data 2005

This table shows the distribution of trades around the bid ask spread along the different stages of data pre-processing for all stocks in the DAX, MDAX and SDAX indices for the year 2005. The first two columns summarize the raw trading data from SIRCA. The first step reduces the number of trades by aggregating all trades occurring within 5 milliseconds at the same price level (5ms-Aggregation). The second step aggregates trades resulting from one large market order that clears several layers of the order book (Run-through). The trade-quote-match (TQM) rule re-aligns delayed reported trades with preceding quotes.

	Raw		**5ms-Aggregation**		**Run-through**		**TQM**	
	# trades	**% trades**	**# trades**	**% trades**	**# trades**	**% trades**	**# trades**	**% trades**
Total Trades	19,292,994	100%	18,971,182	98.3%	17,350,607	90%	17,350,607	90%
above ask	945,692	4.9%	940,486	5.0%	167,132	1.0%	160,211	0.9%
at ask	8,225,025	42.6%	8,099,037	42.7%	8,099,037	46.7%	8,132,850	46.9%
Inside spread	295,851	1.5%	290,469	1.5%	279,458	1.6%	241,998	1.4%
at midpoint	155,245	0.8%	152,118	0.8%	151,356	0.9%	133,933	0.8%
at bid	8,614,651	44.7%	8,458,648	44.6%	8,458,636	48.8%	8,493,575	49.0%
below bid	1,050,004	5.4%	1,026,572	5.4%	191,136	1.1%	184,188	1.1%

Table A.2: Accuracy of trade classification algorithms

This table displays the accuracy of several trade direction algorithms: the tick test, the quote rule, the algorithm by Lee-Ready (LR), Ellis-Michaely-O'Hara (EMO), Chakravarty-Li (ChaLi) and the trade-quote-match rule (TQM). The underlying trading data from the Australian Stock Exchange comprises all stocks listed in the ASX top100 index, October to December 2006. The accuracy for each algorithm was measured after aggregating trades within 5 milliseconds, aggregating run-throughs and after the trade-quote-match-rule had been applied.

	Accuracy after TQM					**Accuracy**
	Tick	**Quote**	**LR**	**EMO**	**ChaLi**	**TQM***
% of trades classified	92%	89%	92%	92%	92%	8%
above ask	87.1%	73.5%	73.5%	87.1%	87.1%	100%
at ask	79.3%	99.1%	99.1%	99.1%	99.1%	48%
inside spread	53.4%	38.3%	38.3%	53.4%	44.8%	100%
at midpoint	54.3%	-	54.3%	54.3%	54.3%	100%
at bid	79.3%	99.1%	99.1%	99.1%	99.1%	50%
below bid	86.0%	70.0%	70.0%	86.0%	86.0%	100%
Total Accuracy	78.2%	95.3%	96.7%	97.1%	96.9%	99.5%

*Accuracy of the trade-quote-match rule is measured relative to only those trades classified by the rule (9% of the total sample).

Table A.3: Extended PIN summary by indices

This table presents means, medians and standard deviations of parameter estimates by index membership for different trade intervals (TI in seconds) and bucket lengths (BS in minutes). PIN is the probability of informed trading, a composite variable of epsilon, mu and alpha. Epsilon is the probability an uninformed trader trades, mu is the probability of an informed trade, alpha is the probability of an information event and delta the probability information events are bad news.

			PIN		**epsilon**		**mu**		**alpha**		**delta**		**standard deviation**				
BS	**TI**	**conv.**	mean	med.	mean	med.	mean	med.	mean	med.	mean	med.	PIN	eps	mu	alpha	delta
DAX																	
8	**5**	96%	.199	.193	.239	.212	.164	.161	.326	.324	.466	.462	.050	.120	.058	.096	.212
	10	93%	.189	.182	.408	.381	.238	.239	.345	.343	.461	.456	.052	.172	.057	.099	.218
	20	93%	.169	.164	.632	.635	.298	.294	.376	.364	.451	.442	.058	.190	.065	.139	.240
	30	93%	.162	.159	.755	.784	.316	.317	.434	.388	.444	.430	.059	.167	.088	.222	.253
12	**5**	91%	.185	.180	.232	.208	.145	.141	.338	.337	.466	.458	.049	.111	.055	.108	.237
	10	87%	.176	.170	.397	.369	.215	.214	.352	.349	.462	.451	.051	.162	.059	.109	.244
	20	84%	.159	.153	.625	.610	.270	.267	.389	.374	.452	.440	.057	.190	.067	.157	.265
	30	85%	.151	.148	.754	.786	.284	.285	.456	.403	.440	.420	.060	.166	.092	.242	.279
15	**5**	82%	.180	.176	.219	.196	.132	.128	.346	.341	.470	.461	.049	.099	.049	.113	.246
	10	81%	.171	.166	.388	.361	.204	.203	.355	.351	.465	.457	.051	.155	.058	.115	.258
	20	77%	.155	.150	.618	.596	.259	.254	.395	.378	.452	.440	.056	.188	.067	.162	.277
	30	78%	.146	.144	.752	.787	.270	.271	.468	.415	.442	.424	.059	.166	.090	.251	.291
MDAX																	
8	**5**	90%	.335	.322	.030	.020	.054	.044	.278	.247	.524	.527	.129	.032	.043	.185	.290
	10	91%	.333	.319	.059	.041	.101	.087	.277	.248	.523	.525	.130	.059	.068	.181	.293
	20	92%	.329	.315	.113	.080	.175	.164	.283	.255	.521	.524	.131	.104	.099	.181	.295
	30	93%	.325	.309	.162	.119	.230	.226	.293	.264	.520	.524	.132	.140	.114	.185	.297
12	**5**	88%	.319	.303	.031	.021	.046	.037	.311	.275	.523	.523	.128	.031	.037	.195	.301
	10	90%	.317	.301	.061	.042	.086	.072	.310	.275	.523	.523	.128	.059	.064	.192	.304
	20	92%	.314	.297	.116	.083	.151	.137	.315	.282	.523	.523	.129	.104	.091	.193	.309
	30	93%	.310	.293	.165	.122	.201	.192	.324	.291	.521	.522	.131	.140	.108	.198	.312
15	**5**	87%	.309	.293	.031	.021	.042	.034	.325	.292	.527	.530	.126	.031	.035	.198	.308
	10	89%	.308	.291	.062	.043	.080	.067	.326	.293	.527	.531	.126	.060	.060	.199	.313
	20	90%	.304	.287	.118	.085	.142	.127	.331	.299	.525	.529	.128	.105	.088	.200	.316
	30	92%	.301	.283	.168	.126	.189	.179	.337	.306	.525	.529	.130	.141	.104	.203	.320
SDAX																	
8	**5**	76%	.399	.368	.006	.003	.021	.014	.300	.186	.538	.545	.216	.015	.026	.302	.378
	10	79%	.401	.368	.012	.006	.041	.027	.299	.184	.536	.541	.219	.023	.047	.301	.379
	20	80%	.402	.366	.024	.012	.078	.055	.302	.186	.534	.538	.222	.041	.079	.302	.380
	30	83%	.399	.361	.035	.018	.108	.079	.310	.192	.533	.535	.223	.057	.102	.307	.382
12	**5**	75%	.392	.365	.035	.003	.017	.011	.334	.221	.537	.546	.212	.057	.044	.305	.382
	10	77%	.393	.365	.013	.006	.034	.022	.330	.220	.537	.543	.215	.024	.039	.302	.382
	20	79%	.394	.364	.024	.012	.063	.043	.334	.223	.535	.539	.219	.041	.066	.302	.383
	30	81%	.393	.360	.036	.018	.090	.064	.339	.227	.532	.540	.219	.058	.091	.307	.386
15	**5**	74%	.390	.357	.006	.003	.015	.010	.346	.244	.543	.556	.210	.015	.019	.300	.382
	10	76%	.389	.355	.013	.006	.030	.020	.345	.241	.539	.545	.215	.024	.036	.300	.383
	20	77%	.391	.358	.025	.013	.058	.039	.344	.242	.540	.545	.217	.043	.061	.298	.383
	30	81%	.393	.355	.037	.019	.081	.057	.361	.250	.537	.544	.219	.059	.084	.310	.386

Table A.4: Autocorrelation of sells

This table shows the autocorrelation of the number of sells per intraday bucket of 8, 12 and 15 minutes and per day, for six lags. The daily autocorrelation was calculated once per stock over the whole year, the intraday autocorrelation was calculated once per day per stock, resulting in 257 values per stock for the year 2005. For each lag, the mean autocorrelation per index and the share of autocorrelation coefficients significantly different from zero at the 10% confidence level is displayed. The final column displays the share of observations where the joint hypothesis of all autocorrelation coefficients being simultaneously zero is rejected, measured by the Q-statistic with a 10% confidence level.

			Mean Autocorrelation						% of AR coefficients sign. α=0.10						Q-Stat sign.
bucket	**obs**	**Lag**	1	2	3	4	5	6	1	2	3	4	5	6	**α=0.10**
DAX															
daily	28		.428	.235	.161	.175	.192	.127	100%	89%	57%	71%	75%	43%	100%
8 min	7196		.382	.285	.232	.200	.166	.144	83%	66%	55%	47%	39%	33%	81%
12 min	7196		.398	.283	.223	.172	.130	.092	77%	56%	44%	31%	20%	11%	71%
15 min	7196		.408	.285	.208	.154	.100	.045	74%	52%	35%	21%	9%	3%	65%
MDAX															
daily	39		.418	.263	.206	.176	.175	.145	100%	85%	79%	72%	69%	49%	100%
8 min	10023		.166	.099	.072	.053	.042	.029	38%	24%	18%	14%	11%	10%	34%
12 min	10023		.171	.094	.060	.038	.022	.009	33%	18%	13%	10%	7%	6%	26%
15 min	10023		.177	.091	.053	.028	.011	-.003	30%	16%	11%	7%	6%	4%	23%
SDAX															
daily	35		.444	.286	.223	.179	.159	.139	100%	97%	95%	71%	68%	47%	100%
8 min	8995		.066	.035	.023	.013	.009	.004	18%	12%	10%	8%	8%	7%	17%
12 min	8995		.066	.030	.013	.004	-.003	-.008	16%	10%	8%	6%	6%	5%	14%
15 min	8995		.067	.024	.008	-.003	-.010	-.016	15%	9%	7%	5%	4%	4%	11%

Table A.5: Autocorrelation of no-trade intervals

This table shows the autocorrelation of the number of no-trade intervals per intraday bucket of 8, 12 and 15 minutes and per day, for six lags. The length of the trade interval was fixed at 5, 10 and 30 seconds for the DAX, MDAX and SDAX respectively. The daily autocorrelation was calculated once per stock over the whole year, the intraday autocorrelation was calculated once per day per stock, resulting in 257 values per stock for the year 2005. For each lag, the mean autocorrelation per index and the share of autocorrelation coefficients significantly different from zero at the 10% confidence level is displayed. The final column displays the share of observations where the joint hypothesis of all autocorrelation coefficients being simultaneously zero is rejected, measured by the Q-statistic with a 10% confidence level.

		Mean Autocorrelation						% of AR coefficients sign. α=0.10						Q-Stat sign. α=0.10
bucket	obs	Lag 1	2	3	4	5	6	1	2	3	4	5	6	
DAX														
daily 5 sec	28	.491	.297	.218	.227	.213	.144	100%	100%	79%	82%	82%	50%	100%
8 min	7196	.482	.385	.325	.282	.241	.211	93%	85%	76%	69%	60%	53%	92%
12 min	7196	.501	.381	.305	.243	.190	.139	91%	77%	64%	51%	35%	20%	87%
15 min	7196	.512	.376	.286	.215	.147	.070	89%	72%	55%	36%	15%	3%	84%
MDAX														
daily 10 sec	39	.498	.334	.275	.238	.226	.194	100%	97%	87%	82%	82%	74%	100%
8 min	10023	.199	.137	.106	.085	.069	.052	47%	34%	25%	21%	17%	14%	45%
12 min	10023	.214	.137	.098	.067	.047	.028	43%	27%	19%	15%	10%	8%	37%
15 min	10023	.226	.136	.090	.056	.031	.006	41%	24%	16%	10%	7%	4%	33%
SDAX														
daily 30 sec	35	.500	.345	.267	.213	.191	.166	100%	100%	95%	82%	76%	55%	100%
8 min	8995	.083	.048	.034	.026	.018	.012	22%	16%	13%	11%	10%	8%	20%
12 min	8995	.086	.045	.028	.016	.008	.002	20%	14%	10%	8%	7%	6%	18%
15 min	8995	.086	.044	.020	.010	.000	-.010	19%	12%	9%	7%	5%	4%	16%

Table A.6: Full company list with search strings and Twitter account information

This table lists all companies analyzed in this study, sorted by index membership. Column 3 to 5 lists the text strings used to filter and assign tweets for all companies in the sample and states the resulting action of a manual validation for false positives of the collected tweets. For some companies the text search resulted in too many false positives, hence this company was excluded from the text search based analysis. For some companies, narrowing the search string helped reduce false positives to an acceptable level. The last column indicates whether a Twitter account of the company could be identified that was active at the start and the end of the observation period.

		Text Search			
Index	**Company**	**RIC short**	**Final text search strings**	**Action after manual validation**	**Twitter Account**
DAX	Adidas	ADSG	Adidas	no changes required	Active
	Allianz SE	ALVG	---	deleted from text sample	Active
	BASF	BASF	BASF	no changes required	Active
	Bayer	BAYG	---	deleted from text sample	Active
	Beiersdorf	BEIG	Beiersdorf	no changes required	Active
	BMW	BMWG	BMW	no changes required	Active
	Commerzbank	CBKG	Commerzbank	no changes required	Invalid
	Continental	CONG	---	deleted from text sample	No account
	Daimler	DAIG	Daimler	no changes required	Active
	Deutsche Bank	DBKG	Deutsche Bank	no changes required	Active
	Deutsche Boerse	DB1G	Deutsche Börse	no changes required	Active
	Deutsche Post	DPWG	Deutsche Post	no changes required	Active
	Deutsche Telekom	DTEG	Deutsche Telekom	no changes required	Active

Index	Company	Text Search			Twitter Account
		RIC short	Final text search strings	Action after manual validation	
	E.ON SE	EONG	E.On	no changes required	Active
	Fresenenius Medical Care KGaA	FMEG	Fresenius Medical Care	no changes required	Active
	Fresenius SE+Co KGaA	FREG	Fresenius SE	switched to "Fresenius SE"	Active
	Heidelbergcement	HEIG	HeidelbergCement	no changes required	No account
	Henkel+Co KGaA	HNKG	---	deleted from text sample	Active
	Infineon Technologies	IFXG	Infineon	no changes required	Active
	K+S	SDFG	K+S	no changes required	No account
	Lanxess	LXSG	Lanxess	no changes required	Active
	Linde		---	deleted from text sample	Active
	Lufthansa	LHAG	Lufthansa	no changes required	Active
	Merck KGaA	MRCG	Merck KGaA	switched to "Merck KGaA"	No account
	Muenchner Rueckversicherung	MUVG	Münchener Rückversicherung; Münchener Rück	no changes required	Active
	RWE	RWEG	RWE AG	switched to "RWE AG"	Active
	SAP	SAPG	---	deleted from text sample	Active
	Siemens	SIEG	Siemens	no changes required	Active
	Thyssenkrupp	TKAG	ThyssenKrupp	no changes required	Active
	Volkswagen	VOWG	Volkswagen	no changes required	Invalid
MDAX	Aareal Bank	ARLG	Aareal Bank	no changes required	Active
	Airbus Group (EADS)	EAD	---	deleted from text sample ("heads up", "dreads", "reads")	Active
	Aurubis	NAFG	Aurubis	no changes required	Active
	Axel Springer SE	SPRG	Axel Springer	no changes required	No account
	Bilfinger SE	GBFG	Bilfinger	no changes required	Active
	Brenntag	BNRG	Brenntag	no changes required	Active
	Celesio	CLSG	Celesio	no changes required	Active
	Deutsche Euroshop	DTEQ	Deutsche EuroShop	no changes required	Active
	Deutsche Wohnen	DWNG	Deutsche Wohnen	no changes required	No account
	DMG Mori Seiki	GILG	Gildemeister	no changes required	No account
	Duerr	DUEG	Dürr AG	switched to "Dürr AG"	No account
	Elringklinger	ZILG	ElringKlinger	no changes required	Active
	Fuchs Petroluchs SE	FPEG	Fuchs Petrolub	no changes required	No account
	Gagfah S.A.	GFJG	GAGFAH	no changes required	No account
	Gea Group	G1AG	GEA Group	no changes required	Active
	Gerresheimer	GXIG	Gerresheimer	no changes required	Active
	Gerry Weber Internat	GWIG	Gerry Weber	no changes required	No account
	Hannver Rueck SE	HNRG	Hannover Rück	no changes required	No account
	Hochtief	HOTG	HOCHTIEF	no changes required	Active
	Hugo Boss		Hugo Boss	no changes required	Active
	Kabel Deutschland Holding	KD8G	Kabel Deutschland Holding	no changes required	Active
	Kloeckner + Co SE	KCOG	---	deleted from text sample (not hits at all)	No account
	Krones	KRNG	---	deleted from text sample	Active

Index	Company	Text Search: RIC short	Text Search: Final text search strings	Text Search: Action after manual validation	Twitter Account
	Kuka	KU2G	---	deleted from text sample	Active
	Leoni	LEOG	---	deleted from text sample (not hits at all)	Active
	Man SE	MANG	---	deleted from text sample	Active
	Metro	MEOG	---	deleted from text sample	No account
	MTU Aero Engines	MTXG	MTU Aero Engines	no changes required	Active
	Norma Group SE	NOEJ	NORMA Group	no changes required	Active
	ProSiebenSat.1	PSMG	ProSiebenSat.1 Media	no changes required	Active
	Rational	RAAG	---	deleted from text sample	Active
	Rheinmetall	RHMG	Rheinmetall	no changes required	Active
	Rhoen-Klinikum	RHKG	RHÖN-KLINIKUM	no changes required	No account
	Salzgitter	SZGG	Salzgitter	no changes required	No account
	SGL Carbon SE	SGCG	SGL CARBON	no changes required	Active
	Sky Deutschland	SKYD	Sky Deutschland	no changes required	Active
	Stada Arzneimitt	STAG	---	deleted from text sample	No account
	Suedzucker	SZUG	Südzucker	no changes required	No account
	TAG Immobilien	TEGG	---	deleted from text sample	No account
	Talanx	TLXG	Talanx	no changes required	Invalid
TecDAX	Adva Optical Network SE	ADAG	Adva AG; ADVA Optical	no changes required	Active
	Aixtron SE	AIXG	Aixtron	no changes required	No account
	BB Biotech	BION	BB Biotech	no changes required	No account
	Bechtle	BC8G	Bechtle	no changes required	Active
	Cancom SE	COKG	Cancom	no changes required	Active
	Carl Zeiss Meditec	AFXG	Carl Zeiss	no changes required	Invalid
	Compugroup Med.	COPM	Compugroup	no changes required	No account
	Dialog Seminconductor	DLGS	Dialog Semicon	no changes required	Active
	Draegerwerk	DRWG	Drägerwerk; Dragerwerk	no changes required	Active
	Drillisch	DRIG	Drillisch	no changes required	Active
	Evotec	EVTG	Evotec	no changes required	No account
	Freenet	FNTG	Freenet AG	switched to "Freenet AG"	No account
	Jenoptik	JENG	Jenoptik	no changes required	Active
	Kontron	KBCG	Kontron	no changes required	Active
	LPKF Laser+Electronics	LPKG	LPKF Laser	no changes required	Active
	Morphosys	MORG	Morphosys	no changes required	Active
	Nemetschek	NEKG	Nemetschek	no changes required	Active
	Nordex SE	NDXG	Nordex	no changes required	No account
	Pfeiffer Vacuum	PV.D	Pfeiffer Vacuum	no changes required	Active
	PSI	PSAG	PSI AG	no changes required	Active
	Qiagen	QGEN	Qiagen	no changes required	Active
	QSC	QSCG	QSC AG	switched to "QSC AG"	Active
	Sartorius	SATG	Sartorius	no changes required	No account
	SMA Solar Technology	S92G	SMA Solar	no changes required	Active
	Software	SOWG	Software AG	no changes required	Active
	Stratec Biomedical	SBSG	Stratec Biomedical	no changes required	No account

Index	Company	Text Search: RIC short	Final text search strings	Action after manual validation	Twitter Account
	Telefonica Deutschland	O2Dn	Telefonica Deutschland	no changes required	Active
	United Internet	UTDI	United Internet	no changes required	Active
	Wirecard	WDIG	Wirecard	no changes required	No account
	XING	OBCG	---	deleted from text sample ("boxing again", "flexing", "mixing")	Active

Table A.7: Dollar/Euro tag sample test

This table displays results on an investigation into the use of $- or €-signs to flag tweets as relevant news for stocks. The first three columns show the text search criteria, the following three columns the corresponding number of tweets collected.

Initial Twitter search strings			Count of identified of tweets			
Company name	RIC (long)	RIC (short) & Ticker symbol	Company name	€-ticker OR $-ticker	Company name AND €/$-ticker	**Share of tweets identified only by €/$-ticker**
EADS	EAD.DE	EAD	123,129	405	398	0.006%
NORMA Group	NOEJ.DE	NOEJ	436	4	4	0%
United Internet	UTDI.DE	UTDI	1,181	9	9	0%

Table A.8: Pairwise correlations for daily Twitter activity and trading on XETRA (per company, twbyz)

This table shows the number of correlations calculated per company that are significant on a 5% and 1% level. Correlations are calculated between standardized daily Twitter activity and indicators of trading activity on XETRA. The total number of companies and hence the maximum possible count is 83. Results for both Pearson correlation and the Spearman rank-correlation are displayed. The third column of each section displays the average of the 83 correlation coefficients.

n = 83 Companies		**Same day**						**Lag 1 day**						**Future 1 day**					
		Pearson			**Spearman**			**Pearson**			**Spearman**			**Pearson**			**Spearman**		
Variables		Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01	Avg rho	#sig .10	#sig .01
Volume	lnVolume	.233	62	45	.169	51	33	.115	38	15	.096	35	8	.103	29	12	.122	39	16
Volatility	highlowVola	.152	42	24	.088	25	10	.043	16	3	.022	15	7	.060	22	8	.069	19	7
	15minVola	.166	49	29	.083	24	11	.049	21	6	.013	16	8	.067	22	11	.062	26	7
	15minVolaMax	.142	39	20	.085	22	7	.035	14	2	.018	19	3	.060	22	7	.065	19	6
Spread	effspread	.119	35	15	.098	29	13	.038	12	4	.045	16	6	.078	22	8	.078	25	7
	qspread	-.012	14	4	-.016	11	4	-.039	16	1	-.016	11	1	-.010	11	2	-.009	12	1
	qspreadeqw	.071	39	15	-.006	28	8	-.046	24	5	-.048	27	8	.011	22	8	-.011	30	8
Small-Large Trades	xt3share	-.100	26	10	-.106	34	18	-.066	23	7	-.078	30	10	-.054	22	8	-.079	30	14
	xt5share	-.086	26	7	-.089	30	17	-.057	17	6	-.068	27	11	-.047	17	3	-.067	25	9
	xt3lnratio	-.112	31	16	-.118	33	20	-.080	29	10	-.091	35	12	-.060	26	7	-.089	29	14
	xt5lnratio	-.115	34	16	-.117	32	20	-.081	28	12	-.091	33	13	-.056	25	7	-.087	31	17
	slt50	.213	58	40	.140	44	24	.101	29	13	.074	29	7	.094	26	10	.101	33	13
	slt500	.209	55	36	.167	52	26	.103	34	13	.090	25	7	.096	24	11	.111	31	9
	slt500p	.117	25	15	.066	13	5	.037	10	1	.036	12	3	.036	11	6	.030	11	2
	sl100k	.215	58	40	.142	45	24	.102	29	13	.075	28	7	.095	26	10	.102	35	13
	sl1mio	.187	52	30	.143	38	17	.082	21	10	.071	20	4	.086	23	8	.100	26	8
	sl1mioplus	.073	15	6	.042	14	1	.028	10	1	.026	11	2	.007	8	2	.006	8	2
Return	retln	.026	31	11	.029	15	4	-.016	8	0	-.021	8	0	.016	9	3	.020	15	1
	retNeg	-.074	23	13	.002	11	2	-.025	11	5	-.023	8	3	-.016	9	4	-.001	13	1
	retPos	.117	32	19	.044	16	1	-.001	11	2	-.016	7	0	.043	13	2	.030	13	2
PIN	PIN	-.086	21	4	-.051	17	6	-.048	12	2	-.031	16	3	-.030	10	1	-.019	17	4
	mu	.114	36	11	.072	20	6	.060	14	3	.036	11	2	.070	23	3	.058	19	7
	epsilon	.170	50	26	.109	36	14	.102	35	12	.063	26	6	.074	23	6	.073	27	6
	delta	-.033	11	2	-.038	15	2	-.032	12	5	-.017	13	3	-.024	9	1	-.029	12	2
	alpha	-.020	8	0	.010	5	1	.006	6	1	.006	8	0	-.015	5	0	.017	8	1

Table A.9: Differences in trading indicators - top vs. bottom tercile of Twitter activity (twbyz)

This table display compares indicators of trading activity on days in the top tercile of Twitter activity to days in the bottom tercile of Twitter activity. Terciles are calculated per single stock over the whole observation period based on the variable *twbyz* The first three columns compare trading and Twitter activity on the same day. The following columns base the comparison on Twitter activity of the previous day (lagged); the last columns are based on the following day's Twitter activity. In each section "MW" is the p-value of a Mann-Whitney mean comparison, "MM" the p-value of a Moods-Median comparison and "Delta Avg" is the difference of the averages in the bottom vs. the top tercile.

Terciles based on twbyz top vs. bottom		**Same day**			**Lag 1 day**			**Future 1 day**		
		sig. of difference		Delta	sig. of difference		Delta	sig. of difference		Delta
Variable		MW	MM	Avg	MW	MM	Avg	MW	MM	Avg
Volume	lnVolume	.	.	69%	.	.	57%	.	.	61%
Volatility	15minVola	.	.	12%	.199	.253	2%	.	.001	8%
	15minVolaMax	.	.	9%	.069	.143	2%	.	.004	7%
	highlowVola	.	.	11%	.081	.272	3%	.	.	9%
Spread	effspread	.164	.435	-2%	.	.026	-6%	.012	.074	-3%
	qspread	.	.	-12%	.	.	-13%	.	.	-11%
	qspreadeqw	.	.	-12%	.	.	-14%	.	.	-12%
Small-Large Trades	xt3lnratio	.	.	-9%	.	.	-7%	.	.	-7%
	xt3share	.	.	-4%	.	.	-3%	.	.	-3%
	xt5lnratio	.	.	-11%	.	.	-8%	.	.	-8%
	xt5share	.	.	-4%	.	.	-3%	.	.	-3%
	slt50	.	.	44%	.	.	34%	.	.	38%
	slt500	.	.	80%	.	.	62%	.	.	68%
	slt500p	.	.	9%	.	.	6%	.	.	6%
	sl100ks	.	.	45%	.	.	34%	.	.	39%
	sl1mio	.	.	57%	.	.	42%	.	.	49%
	sl1mioplus	.	.	3%	.	.	2%	.043	.047	1%
Return	retln	.001	.015	0%	.047	.052	0%	.056	.272	0%
	retNeg	.617	.	0%	.031	.	0%	.855	.	0%
	retPos	.	.01	0%	.102	.098	0%	.004	.161	0%
PIN	PIN	.	.	-8%	.	.	-8%	.	.	-7%
	mu	.	.	5%	.	.01	2%	.	.	5%
	epsilon	.	.	18%	.	.	13%	.	.	15%
	delta	.001	.012	-3%	.157	.294	-2%	.034	.042	-2%
	alpha	.047	.041	1%	.091	.142	1%	.056	.135	1%

Table A.10: Regression on trade metrics (twbyz)

This table display results of fixed-effects panel regressions with *lnVolume*, *15minVola*, *effspread* and *retln* as dependent variables. For each of the dependent variables the first regression is run on contemporaneous data without any lags. The second regression is run including 5 lags of the dependent variable and 5 lags of the Twitter variable. Fixed effects are per company and standard errors clustered by date. Significance is indicated by asterisks: * p<0.05, ** p<0.01, *** p<0.001

Group	Variable	lnVolume		15minVola		effspread		retln	
Twitter	twbyz	0.043***	0.045***	0.008	0.015**	0.034***	0.022***	0.001	0.001*
	L.twbyz		-0.008		-0.017***		-0.012***		0.000
	L2.twbyz		0.003		-0.008**		-0.006*		0.000
	L3.twbyz		0.002		-0.007		-0.001		0.000
	L4.twbyz		-0.003		-0.004		-0.002		0.000
	L5.twbyz		0.002		-0.007		0.001		0.000
Control	lnVolume			0.221***	0.235***	0.003	0.000	0.000	0.000
	15minVola	0.235***	0.238***			0.235***	0.202***	-0.002	-0.002*
	effspread	0.003	0.008	0.163***	0.154***			-0.004*	-0.003
	retln	0.014	-0.076	-0.500	-0.846*	-1.432*	-1.667***		
PIN	PIN	-1.287***	-1.136***	0.265*	0.333**	0.299*	0.207	-0.001	-0.002
	epsilon	2.471***	2.228***	1.017***	0.699***	-0.369**	-0.170*	-0.011	-0.015*
	mu	2.931***	2.810***	0.350**	0.339**	0.033	-0.089	0.009	0.011
	alpha	1.511***	1.394***	-0.025	-0.064	-0.223**	-0.236**	0.005	0.006
	delta	-0.023	-0.022	0.002	-0.003	-0.068**	-0.045**	-0.025***	-0.026***
Lags dep. var.	L1 dep. var.		0.066***		0.157***		0.254***		-0.054
	L2 dep. var.		0.041***		0.080***		0.168***		-0.048
	L3 dep. var.		0.013		0.042**		0.110***		-0.063*
	L4 dep. var.		0.025*		0.037**		0.093***		-0.031
	L5 dep. var.		0.017		0.018		0.099***		0.018
	Constant	15.92***	13.50***	-8.85***	-7.01***	-5.60***	-0.62*	-0.024	-0.030
Observations		10467	10065	10467	10065	10467	10004	10467	10065
R2-Adj		0.969	0.970	0.622	0.647	0.754	0.821	0.099	0.112
Clustered std. errors		date	date	date	date	date	date	date	date
Firm FE		yes	yes	yes	yes	yes	yes	yes	yes

Table A.11: Regression on PIN (twbyz)

This table display results of fixed-effects panel regressions with *PIN* and its components *epsilon*, *mu* and *alpha* as dependent variables. In this table *twbyz* measures Twitter activity. For each of the dependent variables the first regression is run on contemporaneous data without any lags. The second regression includes 5 lags of the dependent variable and 5 lags of the Twitter variable. Fixed effects are per company and standard errors clustered by date. Panel A displays regular coefficients, Panel B standardized coefficients. Significance is indicated by asterisks: * p<0.05, ** p<0.01, *** p<0.001

		PIN		**epsilon**		**mu**		**alpha**	
Panel A: regular coefficients									
Twitter	twbyz	-0.001	-0.001	0.000	0.001	-0.001	0.000	-0.003***	-0.004***
	L.twbyz		0.000		-0.001		0.000		0.001
	L2.twbyz		0.001		-0.003***		0.001		-0.001
	L3.twbyz		0.001		0.000		0.000		0.002
	L4.twbyz		-0.001		0.000		0.000		0.000
	L5.twbyz		0.000		-0.001*		0.000		0.000
Control	lnVolume	-0.037***	-0.032***	0.119***	0.105***	0.052***	0.051***	0.018***	0.016***
	15minVola	-0.006	-0.006	0.037***	0.029***	0.016***	0.017***	-0.005	-0.006
	effspread	0.006*	0.005	-0.012***	-0.008**	0.004*	0.004*	-0.009*	-0.009*
	retln	0.094*	0.094*	-0.093	-0.128	-0.008	0.009	0.106	0.100
Lags dep. var.	L1 dep. Var.		0.166***		0.225***		0.102***		0.113***
	L2 dep. var.		0.030*		0.028*		-0.005		-0.008
	L3 dep. var.		0.017		0.017		0.007		0.023
	L4 dep. var.		-0.003		0.011		0.004		-0.015
	L5 dep. var.		0.018		0.025**		0.027*		0.001
	Constant	0.822***	0.687***	-1.488***	-1.372***	-0.510***	-0.498***	-0.055	-0.068
Panel B: standardized coefficients									
Twitter	twbyz	-.011	-.014	.000	.009	-.009	-.007	-.032***	-.035***
	L.twbyz		.001		-.008		.001		.008
	L2.twbyz		.006		-.019***		.012		-.006
	L3.twbyz		.010		-.003		-.007		.021
	L4.twbyz		-.008		-.001		-.004		.003
	L5.twbyz		.000		-.009*		-.003		-.002
Control	lnVolume	-.656***	-.567***	1.398***	1.240***	1.360***	1.226***	.281***	.238***
	15minVola	-.028	-.031	.121***	.102***	.115***	.124***	-.022	-.028
	effspread	.046*	.034	-.060***	-.039**	.038*	.042*	-.062*	-.056*
	retln	.018*	.019*	-.012	-.017	-.002	.003	.018	.017
Lags dep. var.	L1 dep. var.		.166***		.225***		.102***		.112***
	L2 dep. var.		.030*		.028*		-.005		-.008
	L3 dep. var.		.017		.017		.007		.022
	L4 dep. var.		-.003		.011		.004		-.015
	L5 dep. var.		.018		.025**		.027*		.001
Observations		10467	8708	10467	8708	10467	8708	10467	8708
R2-Adj		54.9%	54.5%	84.6%	85.7%	40.5%	41.5%	5.5%	5.6%
Clustered std. errors		date	date	date	date	date	date	date	date
Firm FE		yes	yes	yes	yes	yes	yes	yes	yes

Table A.12: Ad-hoc announcements sample description

This table describes the sample of ad-hoc announcements. Panel A lists the number of ad-hoc announcements, Panel B the number of companies with at least one ad-hoc announcement within our observation period per index. The columns next to the total indicate how many of those fall within or outside of trading hours and for how many announcements we could collect tweets based on the company name ("Text") and from the Twitter account operated by the company ("Account").

		Trading hours		**Corresponding tweets available**	
Index	**Total**	Within	Outside	Text	Account
Panel A: Number of news					
DAX	34	13	21	28	14
MDAX	52	26	26	35	6
TecDAX	50	23	27	36	7
Total	136	62	74	99	27
Panel B: Companies covered					
DAX	19	10	15	15	9
MDAX	24	14	16	16	4
TecDAX	20	14	15	18	5
Total	63	38	46	49	18