Topics in Empirical Market Microstructure:
Measuring the Informational Content of Order Flow

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Wirtschaftswissenschaftlichen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt von

Oliver Wünsche
aus Löbau

2010

Dekanin:	Prof. Dr. rer. pol. Kerstin Pull
Erstberichterstatter:	Prof. Dr. rer. pol. Joachim Grammig
Zweitberichterstatter:	Prof. Dr.-Ing. Rainer Schöbel
Tag der mündlichen Prüfung:	14. August 2008

Acknowledgments

Although there is a page to be submitted stressing out that this thesis is solely the authors own work (except of course potential co-authors), a lot of people are involved in the evolution of the final paper. Some are directly involved through subject matter discussions or in various support functions while some had a rather indirect effect. I could probably write a separate thesis naming all the people who I am grateful to. However, I rather focus to name a few that, I believe, had the most significant impact on this work.

First and foremost, I want to thank my supervisor Joachim Grammig for all the support and patience throughout my time in Tübingen University. I learned a lot from you, Jo, far beyond the academic context. I would also like to thank my colleagues Robert Jung, Kerstin Kehrle, Ramona Maier, Luis Huergo, Franziska Peter, Thomas Dimpfl, Peter Schmidt as well as my co-author Erik Theissen for all their constructive feedback during numerous presentations and brain-stormings from which I heavily benefitted. Further, I thank especially my colleague Stefan Frey who put tremendous effort in preparing the data set and setting up a sophisticated IT infrastructure. Without that achievement, the calculations and analyses in this thesis would have probably taken decades until finalization. Although she was sometimes hard to understand due to her Swabian dialect (in particular when I was not yet familiar), I strongly appreciate the contribution of Angelika Hutt who maintained the homepage and helped greatly in all matters of administration. I should also not forget to thank our awesome student assistants Anja, Julia, Uli, Natascha, Benno, Tobias, Felix and Miriam who did a great job in providing support and brought (not rarely) entertainment to the office.

My fascination for statistics and econometrics however, started already during my undergraduate studies, in particular during my time as student assistant at the chair of econometrics of Reinhard Hujer in Frankfurt. I am heavily indebted to all the people who introduced me into the mysteries of SAS, Gauss, LaTeX and the basic knowledge of empirical research, namely Marco, Dubi, Stefan, Christopher, Sandra, Stephan and last but not least my good friend Paulo with whom I pretty simultaneously developed an enthusiasm for the world of econometrics.

Finally, I want to thank my family who always supported me throughout the entire years of education. Without you, not a single page of the following text would have been written.

Oliver Wünsche

Zürich, 06. February 2010

Contents

List of Figures

List of Tables

Chapter 1

Introduction

When Garman (1976) coined the term ”market microstructure”, a new research direction in the broad field of finance and capital market research was born. In recent years, there was a surge in financial market microstructure research. Due to the new dimensions of computer technology and availability of data, especially, empirical studies are sprouting up. Also new trading platforms and trading mechanisms have evolved and gain more and more influence. Fully electronic limit order books play a very important role in today's stock exchange design. They differ from a traditional specialist market in terms of transparency, anonymity and the wide variety of order types from a traditional specialist market. While the classical capital market theory deals with equilibrium prices and equilibrium quantities, market microstructure rather tries to shed light on the *path* to equilibrium. How can agents benefit from not only watching the outcome of the trading process (e.g daily closing prices) but the trading process itself (e.g when and how much is traded on a transaction level)? How fast are prices reacting to news events? What is the probability that a market event was triggered by private information? How large is the impact of private information compared to pure noise trading on prices? What is the best market design to facilitate a profitable trading platform? All those important questions demand a more detailed look through the microscope at the trading process itself.

Referring to Madhavan (2000), one could say:

> ”Market microstructure is the area of finance that studies the process by which investors' latent demands are ultimately translated into prices and volumes.”

When talking about *empirical* market microstructure we usually talk about large data sets representing the fast-paced trading process. Compared to traditional daily or weekly data, *high-frequency data* poses an enormous challenge for the researcher. Generally spoken, the notion that the more information, i.e. the more data, the better the results has been found to be wrong in several respects. A famous example is measuring volatility more accurately by using intra-day data of price changes. Even though it was shown that using a finer time grid for the data could substantially enhance short-term forecasts there was a drawback. Naturally, seeing the aforementioned improvement we would suggest to make the time grid even finer and eventually take every available data point. The negative phenomenon related to this issue is well known as *microstructure noise*. If microstructure noise is left unaccounted for, increasing the frequency beyond a certain point can lead to serious flaws concerning the estimated parameters of interest (compare Aït-Sahalia, Mykland, and Zhang (2005)). On the other hand, if agents act rationally, prices of financial assets should adjust very quickly to their true values. Hence, it is desirable to use data on its highest frequency, so-called tick-by-tick data to learn something about price discovery.

Broadly speaking, one could say that from a theoretical microstructure perspective, *each* market event is informative. From a statistical point of view an irregularly spaced tick-by-tick data series is a marked point process. The time stamps of the events are the points and the realizations are the marks. Traditionally, time intervals were equally spaced and thus, did not convey additional information. When modeling high-frequency time series not only are the realizations of the variables of specific interest but their timing as well. Obviously, if the timing of market events is not purely random, it is desirable to find an adequate modeling approach describing the "timing process". Engle and Russell (1998) showed that the waiting time between market events is predictable and proposed to model the waiting times as an autocorrelated conditional duration (ACD) process. Since then, a plethora of econometric models has been proposed to account for the irregularly spaced time occurrence of market events. For example, Ghysels and Jasiak (1998) find that volatility has an impact on the time between transactions and that the persistence in GARCH models drops when trade durations are taken into account. Dufour and Engle (2000) analyze the price impact of trades taking into account the trade duration.

The vast amount of literature on the topic comprises, among others, the book of Harris

(2003) which focusses on the different trading and exchange mechanisms. Excellent surveys about theoretical and empirical models brought forward in market microstructure are O'Hara (1995), Madhavan (2000) and, more recently, Biais, Glosten, and Spatt (2005). A good share of the methodologies used in empirical market microstructure is summarized in Hasbrouck (2007).

How to measure information on financial markets in a microstructure setting

The traditional microstructure view explained price discovery mainly in the context of inventory models (see for example Ho and Stoll (1981) and Ho and Macris (1984)) meaning that the specialist who faces order flow uncertainty determines his quotes in order to optimize his inventory holdings. The last two decades, however, strengthened the view that information related trading is far more important for price formation (see Glosten and Harris (1988), Huang and Stoll (1997) or Madhavan, Richardson, and Roomans (1997) to name a few). A common basic assumption of the latter models is that there are two categories of traders, informed traders who possess superior information about the fundamental asset value and uninformed traders who merely trade for liquidity needs. The first significant contribution in empirically measuring the informational content of a trade has been provided by Hasbrouck (1991a). In his bivariate VAR approach he quantified the impact of a trade on the instantaneous quote revision after the trade. Further, Hasbrouck (1991b) derives an information measure based on a variance decomposition of the VAR allowing to compare the degree of information for different stocks.

In this thesis, I will focus on two types of microstructure models. The first model class comprises spread decomposition models, relating the price process to order flow (or trade direction) in order to decompose the bid-ask spread into an adverse selection component related to private information and into a component due to order processing costs. The second model class are sequential trade models, using order imbalances determined by aggregated order flow over a fixed time interval to estimate the probability of informed trading (PIN).

Spread Decomposition Models

Two views about the constituents of the bid-ask spread have dominated the literature in the past decades. In the late 1970's (e.g. Stoll (1978a) and Stoll (1978b)) and early 1980's (e.g. Amihud and Mendelson (1980) and Ho and Stoll (1981)), most studies dealing with microstructure theory pursued the view that, beside the institutional costs of order processing, the spread is the result of inventory optimizing behavior of the specialist. However, in the last 20 years, more and more exchanges have evolved running a fully electronic limit order book where liquidity is provided voluntarily via the submission of limit orders. For this type of exchange, inventory costs should play at best a minor role since nobody is obliged to take inventory. Bagehot (1971) was the first to distinguish between liquidity traders and informed traders and noted that the bid-ask spread consists of three components: order processing costs (including costs of exchange infrastructure etc.), inventory costs (to compensate for the risk of holding a sub-optimal portfolio) and adverse selection costs (to compensate for the risk of losing to a superior informed trader). In a newer study, Flood et al. (1998) find that in multiple dealer markets, search costs related to finding price quotes should be taken into account as an additional component. Drawing on the information asymmetry approach, several models have been proposed to disentangle and estimate those components. Some of the most popular and widely used models are Glosten and Harris (1988), George, Kaul, and Nimalendran (1991), Huang and Stoll (1997) and Madhavan, Richardson, and Roomans (1997).

In chapter 2, I will present an extension of the spread decomposition model of Madhavan, Richardson, and Roomans (1997). The novelty is to measure the impact of trade duration shocks extracted from an ACD model on the information content of a transaction. While Dufour and Engle (2000) find that transactions with short durations have a higher price impact and thus, are more informative, the results of Grammig, Theissen, and Wünsche (2007) indicate the opposite relationship.

Sequential Trade Models

Another strain of market microstructure literature does not focus on tick-by-tick event data but on aggregated order flow. Some famous models belonging to the class of sequential trade models are Copeland and Galai (1983), Kyle (1985)[1], Glosten and Milgrom (1985), Glosten (1987) and Easley, Kiefer, O'Hara, and Paperman (1996). The central assumption of this model class is that two types of agents, informed and uninformed, trade with a specialist who does not know the trader type he deals with. The specialist is therefore not only exposed to inventory costs but additionally to the risk of an adverse price movement when dealing with an informed trader. But she can infer information about the proportion of informed traders in the market by observing the order flow and adjust transaction costs accordingly. Chapter 3 and 4 deal with extensions of the sequential trade model proposed by Easley, Kiefer, O'Hara, and Paperman (1996). Chapter 3 provides an application of the EKOP model with time varying arrival rates proposed by Easley, Engle, O'Hara, and Wu (2002). In contrast to Easley et al., I apply the model to intra-day data, i.e. buys and sells are aggregated for each five minute interval of the trading day. This allows the estimation of an intra-daily pattern of the probability of informed trading together with intra-daily arrival rate dynamics. From an economic perspective, the model specification allows for strategic behavior of the two trader groups.

In chapter 4, I show that the traditional EKOP specification lacks empirical fit concerning the joint distribution of buy and sell counts. Venter and de Jongh (2004) propose to use a bivariate Poisson Inverse Gaussian mixture to model the joint process instead of an independent bivariate Poisson distribution. Although the empirical fit can be enhanced substantially, their model is rather intractable concerning the computational effort to compute the likelihood function. Therefore, I present the bivariate negative binomial distribution as a viable alternative. It not only delivers a very good empirical fit but can be easily computed and converges almost as rapidly as the traditional Poisson specification. Further, I show in a simulation study, if and how the estimated parameters, especially the probability of informed trading, are affected when the underlying count distribution is misspecified. The results show that the commonly used Poisson model tends to overestimate the PIN if the buy and sell counts are realizations of a mixed Poisson process.

[1] Kokot (2004) pointed out that the Kyle model is a Walrasian batch model rather than a sequential trade model because the market maker sets a single price (not a bid and an ask) for which all trades are executed. However, the model design is very similar.

Chapter 2

Revisiting the Role of Time for the Price Impact of a Trade

Dufour and Engle (2000) have shown that the duration between subsequent trade events carries informational content with respect to the evolution of the fundamental asset value. Their analysis supports the notion that "no trade means no information" derived from Easley and O'Hara's (1992) microstructure model. This paper revisits the role of time in measuring the price impact of trades using a structural model and provides challenging new evidence. For that purpose we extend Madhavan et al.'s (1997) model to account for time varying trading intensities. Our results confirm predictions from strategic trading models put forth by Parlour (1998) and Foucault (1999) in which short durations between trades are not related to the processing of private information. Instead, they are caused by strategic trading of impatient non-informed agents who use market orders more intensively when order book liquidity is high.

Chapter is based on the article Time and the Price Impact of a Trade - A Structural Approach by J. Grammig, E. Theissen and O. Wuensche (2007)

2.1 Introduction

Why do security prices change? What is the amount of price-relevant information contained in a trade event? And in which way is the phenomenon that the time intervals between trade events exhibit idiosyncratic patterns associated with information processing? The availability of financial markets transaction level data allows to address those questions. Valid answers are interesting for both academia and decision making of investors and designers of trading venues. It is thus not surprising that a vast literature has evolved, theoretical and empirical, in which these questions are addressed. Bringing together the empirical microstructure literature originating in the seminal papers by Hasbrouck (1991a,b) who introduced vector autoregressive models (VARs) in microstructure, and recent contributions to modeling the properties of financial duration processes (Engle and Russell (1998), Engle (2000)), Dufour and Engle (2000) investigate the role of time varying transaction intensities in measuring the informational content of trades. Their paper made a strong point for the "no trade means no information" prediction derived from Easley and O'Hara's (1992) microstructure model.

This paper revisits the role of time in measuring the price impact of trades using a structural framework and provides challenging new evidence. Instead of employing an agnostic VAR, we extend Madhavan et al.'s (1997) seminal model (MRR) to account for time varying trading intensities. To model the duration process we combine the MRR model and Engle and Russell's (1999) ACD framework. We estimate both the Dufour/Engle VAR and the extended MRR model for a cross section of stocks traded on one of the largest Continental European Markets and also, for robustness checks, on NYSE data. The results challenge the "no trade means no information" interpretation. Rather, our analysis corroborates predictions from strategic trading models put forth by Parlour (1998) and Foucault (1999), in which short durations between trades are not related to the processing of private information.

This paper connects to various streams in the literature which investigate the role of time in the trading process. The empirical analysis provides an empirical test of the predictions of theoretical microstructure models involving the role of time in the trading process. Due to their different inherent assumptions, these models deliver conflicting predictions. For instance, Diamond and Verrecchia (1987) predict that in the case of short sale constraints long intervals of trade inactivity are evidence for bad news. On the other hand, in absence of such restraints, the model put forth by Easley and O'Hara (1992) implies that long no-trade intervals indicate that there is no new information. In their model informed traders split up their orders in smaller chunks in order to disguise their trading motive. The order splitting strategy increases trading intensity, and leads to shorter durations between trades which then tend to be more informative w.r.t. the evolution of the asset price. Another classic reference is Admati/Pfleiderer (1988). In their model, non-informed liquidity traders cluster during periods of the trading day. This implies that high trading intensity would be associated with less informed trading. Although traders with private information may hide in the crowd of liquidity traders, the price impact of their trades is "cushioned" by the trading of the uninformed liquidity traders. Recent strategic trading models (Parlour (1998) and Foucault (1999)) provide more elaborate explanations for such a clustering process. In Parlour's (1998) model, large depth on the bid side increases buyer "aggressiveness", in other words: more buy market orders, more (buyer initiated) trading activity. The reasoning is symmetric for the sell side. Arguably, periods of high liquidity (measured as the depth at the best quotes) in limit order markets are associated with a lesser degree of private or public information in the market. Specifically, patient limit order traders, who are not trading for reasons of exploiting their superior information, and who are not afraid of being "picked off" by an informed order or adverse price movement, will supply ample liquidity. Thus, high liquidity in the order book emerges during non-informative periods. As liquidity traders gather together (via limit order submission, they supply liquidity) during non-informative periods, trading becomes more aggressive (more market orders triggering trades) when impatient traders strive to get priority over standing limit orders.

The second stream in the literature to which this paper connects to is the statistical modeling of time varying transaction intensities in financial markets. Engle and Russell's (1998) seminal contribution triggered a growing literature that proposes statistical methodologies to account for the idiosyncratic time series properties of financial duration processes (e.g. Engle (2000), Zhang, Russell, and Tsay (2001) Bauwens and Giot (2001), Fernandes and Grammig (2006)). Dufour and Engle (2000) linked this literature to a classical empirical microstructure methodology introduced by Hasbrouck (1991a, 1991b) who proposed to measure price impacts of trades via a VAR framework. Hasbrouck's VAR approach does not explicitly take into account the fact that the time between trades and quote updates is varying. The potential information contained in these no-trade intervals is neglected. Dufour and Engle (2000) showed that accounting for the time between trades does matter in measuring trade informativeness in a Hasbrouck-VAR framework. Dufour and Engle (2000) estimate their model on a cross section of stocks using the 1991 TORQ data base. Their study made a strong case for the "no trade means no information" prediction derived from Easley and O'Hara (1992). They found that trades after short durations have a significantly higher price impact than trades after long durations.

The third stream in the literature with which our contribution is connected, is the class of structural models introduced by Glosten and Harris (1988), Madhavan, Richardson, and Roomans (1997) and Huang and Stoll (1997). In contrast to Hasbrouck's VAR framework, these models contain structural equations for the evolution of the latent asset price which depends on the informational content of trade events. Furthermore, the anticipatory behavior of liquidity suppliers who take into account price impacts of trades (due to informed trading), order processing costs, and possible costs of holding unwanted inventory, is explicitly accounted for. As in Hasbrouck's VAR methodology, however, the information allegedly contained in no-trade intervals is not taken into account.

This paper contributes to the literature in the following way. We extend the Madhavan et al. (1997) (MRR) model to account for time-varying trade intensities, and revisit the role of time in measuring the informational content of a trade. To account for time varying trade intensities in a structural model we combine the MRR model with the autoregressive conditional duration (ACD) model introduced by Engle and Russell (1998). To our knowledge this paper is the first to provide a link of the structural models of market microstructure to the literature that deals with the modeling of dynamic duration processes. We show how structural parameters and the parameters of the ACD can be conveniently and simultaneously estimated using the Generalized Method of Moments (GMM). We estimate both the Dufour/Engle VAR and our extended MRR model on a cross section of stocks traded on one of the large European Stock markets, the Frankfurt Stock Exchange (FSE) which is operated as an automated auction market. For a robustness check the model is also estimated on a matched sample of NYSE traded stocks. One advantage of using the FSE data is their excellent quality. Problems that arise from misclassified trades, which can have severe consequences (see Boehmer, Grammig, and Theissen (2007)), are avoided. Furthermore, as open order book markets become increasingly important, it seems interesting to estimate these models using recent data generated within these market structures.

The main results are as follows. Estimating the extended MRR model on European and NYSE data we find that trades occurring after periods of inactivity (long durations between trades) are more informative than trades during active periods (short durations), a result that is also confirmed for the NYSE control sample. The adverse selection component of the spread is higher for trades after long durations. We also find that adverse selection costs of less actively traded stocks are more severely affected by the time between transactions than more actively traded stocks. These results challenge the "no trade means no information" result of Dufour and Engle (2000). Rather than supporting the predictions of the Easley and O'Hara (1992) model, our findings are more in accord with the models of strategic trading in limit order markets (Parlour (1998), Foucault (1999)). As noted above, a high trading intensity in those models is caused by the submission of market orders by impatient, yet uninformed traders who strive aggressively for priority for their orders when the liquidity on their own market side is high (small spread, large depth). However, liquidity supply is ample when limit order traders are not afraid of being picked off by an adverse price movement (be it induced by public or private information processing). These results emphasize the relevance of the Admati/Pfleiderer (1988) explanation that through clustering of liquidity-induced trading, short durations between trades are associated with a smaller price impact of trades.

Estimating Dufour and Engle's extended Hasbrouck-VAR on our data we broadly confirm their main conclusions. The contradictory results must therefore be attributable to the methodology used to measure the informational content of trade. Investigating this issue in greater detail we conjecture that the differences are caused by the way the econometric methodologies deal with partially filled market-to-limit orders which are a quite popular instrument used by traders and partially filled marketable limit orders.

The remainder of this chapter is organized as follows. Section 2.2 describes the data used for the analysis and the market structure. The empirical methodology employed in our study is presented in section 2.3. In section 2.4 we discuss the empirical results and section 2.6 concludes.

2.2 Market Structure and Data

In our empirical analysis, we use data from the automated auction system Xetra which is operated at various European trading venues, like the Vienna Stock Exchange, the Irish Stock Exchange, the Frankfurt Stock Exchange (FSE) and the European Energy Exchange.[1] Specifically, our data are the 30 DAX stocks traded at the FSE in the first quarter, 2004. In chapter 2, we also use TAQ data for a matched sample of NYSE traded stocks as a robustness check. Since the NYSE trading process has been outlined in many papers and textbooks (see e.g. Bauwens and Giot (2001) and Harris (2003) for lucid surveys), we refrain from adding another description. The Xetra trading system, however, warrants some explanations.

Xetra is a pure open order book system developed and maintained by the German Stock Exchange. It has operated since 1997 as the main trading platform for German blue chip stocks at the FSE. Since the Xetra/FSE trading protocol is the data generating process for this study we will briefly describe its important features.[2]

Between an opening and a closing call auction - and interrupted by another mid-day call auction - Xetra/FSE trading is based on a continuous double auction mechanism with automatic matching of orders based on the usual rules of price and time priority. During pre- and post-trading hours it is possible to enter, revise and cancel orders, but order executions are not conducted, even if possible. During the year 2004, the Xetra/FSE hours extended from 9 a.m. C.E.T to 5.30 p.m. C.E.T. For blue chip stocks there are no dedicated market makers like the Specialists at the New York Stock Exchange or the Tokyo Stock Exchange's Saitori. For some small capitalized stocks listed in Xetra there may exist so-called Designated Sponsors - typically large banks - who are required to provide a minimum liquidity level by simultaneously submitting competitive buy and sell limit orders.

In addition to the traditional limit and market orders, traders can submit so-called iceberg (or hidden) orders. An iceberg order is similar to a limit order in that it has pre-specified limit price and volume. The difference is that a portion of the volume is kept hidden from the other traders and is not visible in the open book.

Xetra/FSE faces some local, regional and international competition for order flow. The

[1] The Xetra technology was recently licensed to the Shanghai Stock Exchange, China's largest stock exchange.

[2] The Xetra trading system resembles in many features other important limit order book markets around the world like Euronext, the joint trading platform of the Amsterdam, Brussels, Lisbon and Paris stock exchanges, the Hong Kong stock exchange described in Ahn et al. (2001), and the Australian stock exchange, described in Cao et al. (2004).

FSE maintains a parallel floor trading system, which bears some similarities with the NYSE, and, like in the US, some regional exchanges participate in the hunt for liquidity. Furthermore, eleven out of the thirty stocks we analyze in our empirical study are also cross listed at the NYSE, as an ADR or, in the case of DaimlerChrysler, as a globally registered share. However, the electronic trading platform clearly dominates the regional and international competitors in terms of market shares, at least for the blue chip stocks that we study in the present work.

The Frankfurt Stock Exchange granted access to a database containing complete information about Xetra open order book events (entries, cancelations, revisions, expirations, partial-fills and full-fills of market, limit and iceberg orders) which occurred during the first three months of 2004 (January, 2nd - March, 31st). The sample comprises the thirty German blue chip stocks constituting the DAX30 index. Based on the event histories, we use a real time reconstruction of the sequences of best bid and ask prices and associated depths, and record a time stamped series of transactions (with transaction price and volume) initiated by market order or marketable limit order traders.[3] The resulting data are comparable to the Trade and Quote (TAQ) data supplied by the New York Stock Exchange. Contrary to the TAQ data set, we know the correct trade direction identifier and do not have to apply trade classification algorithms, e.g. Lee and Ready (1991).

Table 2.2.1 reports descriptive statistics for the thirty stocks that constitute the DAX30 index. The table also displays the sorting of the thirty stocks into four groups. The stocks are grouped according to their trading frequency (measured as the average number of trades per day). Group one contains the most frequently traded stocks, while group four the least frequently traded stocks. The table contains the market capitalization, the daily turnover and the average daily number of trades as well as the average midquote price, the quoted spread and the average relative quoted spread.

[3] We are indebted to Stefan Frey and Helena Beltran who performed the reconstruction of the order book.

Ticker	Company Name	Daily Turnover (Mill.)	Market cap. (Mill.)	Daily nb. trades	Avg. Price (€)	Effective Spread (€)	Effective Spread (%)	Realized Spread (€)	Realized Spread (%)	Price Impact (€)	Price Impact (%)	Trade Activity Quartile
ALV	ALLIANZ	289.98	33805	4523	100.1	0.049	0.049	0.010	0.010	0.039	0.039	1
DTE	DEUTSCHE TELEKOM	350.63	34858	4445	15.7	0.011	0.072	0.005	0.031	0.006	0.041	
SIE	SIEMENS	321.70	52893	4418	64.0	0.026	0.041	0.004	0.006	0.022	0.035	
DBK	DEUTSCHE BANK	309.28	38228	3961	67.2	0.030	0.044	0.003	0.004	0.027	0.039	
MUV2	MUENCH. RUECKVERS.	207.35	16396	3425	93.9	0.046	0.049	0.005	0.005	0.042	0.045	
DCX	DAIMLERCHRYSLER	187.74	30316	3309	36.4	0.020	0.055	0.004	0.010	0.016	0.044	
EOA	E.ON	160.63	33753	2871	52.5	0.025	0.048	0.001	0.003	0.024	0.046	
SAP	SAP	184.63	27412	2806	131.5	0.065	0.049	0.002	0.001	0.063	0.048	2
IFX	INFINEON	146.46	4790	2799	11.6	0.012	0.104	0.005	0.040	0.007	0.064	
BAS	BASF	124.43	25425	2580	43.3	0.022	0.051	0.001	0.002	0.021	0.049	
VOW	VOLKSWAGEN	104.25	9688	2545	39.2	0.022	0.056	0.002	0.004	0.020	0.052	
BAY	BAYER	88.78	15911	2400	23.1	0.017	0.076	0.003	0.012	0.015	0.064	
RWE	RWE	97.66	12653	2314	33.8	0.021	0.062	0.001	0.002	0.020	0.060	
BMW	BMW	87.85	12211	2110	34.7	0.021	0.060	0.001	0.003	0.020	0.057	
HVM	HYPO-VEREINSBANK	98.35	6629	1937	18.7	0.018	0.098	0.003	0.019	0.015	0.079	
SCH	SCHERING	51.41	7055	1523	40.8	0.029	0.071	0.002	0.004	0.027	0.067	3
CBK	COMMERZBANK	53.17	7569	1450	15.4	0.015	0.100	0.004	0.023	0.012	0.077	
LHA	LUFTHANSA	43.95	4548	1352	14.2	0.016	0.111	0.003	0.022	0.012	0.088	
DPW	DEUTSCHE POST	43.84	6806	1315	18.2	0.018	0.097	0.003	0.018	0.014	0.079	
TKA	THYSSENKRUPP	37.89	6450	1262	15.9	0.018	0.111	0.005	0.029	0.013	0.083	
MEO	METRO	38.87	5018	1235	35.0	0.031	0.089	0.000	0.000	0.031	0.090	
ALT	ALTANA	30.99	3338	1095	48.6	0.039	0.079	0.004	0.008	0.035	0.071	
TUI	TUI	26.28	2025	1063	18.7	0.023	0.125	0.003	0.015	0.020	0.109	
MAN	MAN	27.69	2434	1057	27.7	0.027	0.096	0.001	0.003	0.026	0.094	4
CONT	CONTINENTAL	25.63	4060	1002	31.6	0.029	0.092	-0.003	-0.011	0.032	0.103	
DB1	DEUTSCHE BOERSE	35.70	4847	982	46.9	0.035	0.075	0.001	0.003	0.034	0.072	
ADS	ADIDAS-SALOMON	31.98	4104	980	92.6	0.065	0.070	-0.002	-0.002	0.067	0.072	
LIN	LINDE AG	22.38	3448	896	43.6	0.035	0.080	-0.004	-0.009	0.039	0.090	
HEN3	HENKEL	18.17	3682	702	65.9	0.050	0.077	0.003	0.005	0.047	0.072	
FME	FRESENIUS MEDICAL CARE	12.85	1944	621	54.0	0.053	0.098	0.006	0.010	0.047	0.088	
	Average	108.68	14076	2099	44.5	0.030	0.076	0.002	0.009	0.027	0.067	

Table 2.2.1: **Characteristics of the stocks in the sample (Xetra/DAX stocks).** The table reports characteristics of the stocks constituting the DAX30 index and our sample. The statistics are computed based on the data on the market events during the sample period January 2, 2004 to March 31, 2004 except for the column *Market cap.* which gives the market capitalization of the respective stock in million euros at the end of December 2003. *Daily turnover* is the total average turnover (in mill. euros) per trading day and *Daily nb. trades* is the average daily number of trades. *Price*, denotes the average midquote. *Effective Spread* (in euros) and *Effective Spread (%)* report the average effective spread and the average relative effective spread. *Realized Spread* (in euros) and *Realized Spread (%)* report the average realized spread and the average relative realized spread. *Price Impact* (in euros) and *Price Impact (%)* report the average price impact and the average relative price impact over the 3 months sample period.. The price impact was obtained by subtracting the realized spread from the effective spread. The stocks are sorted into four groups according to their trading frequency, i.e. by the column *Daily nb. trades*. The horizontal lines separate the four trading activity quartiles.

2.3 Empirical Methodology

2.3.1 The Dufour/Engle Approach

Before introducing the structural framework that we use to investigate the role of time and the price impact of trades let us briefly review Dufour and Engle's (2000) extension of Hasbrouck's (1991a,b) bivariate vector autoregressive model. To measure the price impact of trades and the role of duration between trade events, Dufour and Engle set up a VAR that contains two equations, one that accounts for the dynamics of the midquote revision process and one that models the evolution of the direction of trades:

$$R_i = \sum_{j=1}^{5} a_j R_{i-j} + \gamma_{open} D_i Q_i + \sum_{j=0}^{5} b_j Q_{i-j} + v_{1,i} \quad (2.1)$$

$$Q_i = \sum_{j=1}^{5} c_j R_{i-j} + \gamma_{open} D_{i-1} Q_{i-1} + \sum_{j=1}^{5} d_j Q_{i-j} + v_{2,i} \quad (2.2)$$

$$\text{where} \quad b_j = \gamma_j + \delta_j \ln(T_{i-j}). \quad (2.3)$$

Q_i is an indicator of the side of the trade taking the value 1 for a buyer initiated trade and -1 for a seller initiated trade. The counter-party of the trade is the liquidity supplier, either a dedicated market maker or the open limit order book. R_i denotes the instantaneous midquote revision after a trade. T_i measures the time interval (in seconds) between the i^{th} and the $i-1^{th}$ trade. Note that the i^{th} trade in time affects the midquote revision contemporaneously while Q_i is only affected by lagged midquote revisions. The model ticks in event time. A new trade increases i by one. The extension of Dufour and Engle (2000) to the Hasbrouck (1991a,b) model is the parameterization of the price impact b_j as a function of time between trades (duration) T_i.

The parameter b_0 is a raw measure of the informational content of the trade. The higher b_0, the larger the instantaneous price impact of a trade. Whether a longer duration between trades T_i leads to an increasing or decreasing price impact depends on the parameter δ_0. If δ_0 is negative a longer trade duration would be associated with a reduced price impact and hence, a less informative trade. In other words, a low trading frequency would be related to less informative trades if δ_0 is negative. Estimation of the model can be straightforwardly conducted via equation by equation OLS. The role of the price impact of trades can be assessed by the size of the parameter estimates of δ_j and, more sophisticated, by an impulse response analysis which, requires simulating future trade durations (see Dufour and Engle (2000) for details).

2.3.2 A Structural Approach

As an alternative to measuring the role of time in measuring the price impact of trades we resort to an alternative class of structural models which are extensively used in market microstructure. The most popular examples are the models proposed by Glosten and Harris (1988), Madhavan et al. (1997) and Huang and Stoll (1997). These models consist of structural equations for the evolution of the fundamental asset value and the behavior of liquidity suppliers (market makers or limit order traders) which post bid and ask quotes anticipating the price impact of trades and demand compensation for involuntary inventory taking and order processing costs. One of the advantages of these models is the clear theoretical background that allows to give the parameter estimates a structural interpretation and allows for an economically meaningful decomposition of the spread. Let us briefly review the basic contents of the Madhavan et al (1997) model that we will extend below to account for a time varying trade intensity.

In the MRR model there are two factors driving the fundamental value of a stock. First, we have the public news factor. The second factor is private information which can be inferred from order flow and consists of the surprise in order flow multiplied with a measure for the degree of asymmetric information. For the post-trade expected value of a stock, μ_i, results the following expression:

$$\mu_i = \mu_{i-1} + \theta(\cdot) \cdot (Q_i - E[Q_i|Q_{i-1}]) + \varepsilon_i \tag{2.4}$$

where $Q_i - E[Q_i|Q_{i-1}]$ measures the surprise in order flow and $\theta(\cdot)$ the degree of trade informativeness conveyed through a surprise in the order flow. ε_i denotes the public news impact which is assumed to be an i.i.d random variable with zero mean and variance σ_ε. Liquidity providers know μ_{i-1} and ε_i (public news accrued from $i-1$ to i) but not Q_i. But they can anticipate the effect of Q_i and set bid and ask prices accordingly. Bid and ask prices are set to reflect the expected value of the stock plus a fixed component $\phi(\cdot)$ which can be interpreted as a compensation for order processing or possible inventory holding:

$$\text{ask price: } P_i^a = \mu_{i-1} + \theta(\cdot)(1 - E[Q_i|Q_{i-1}]) + \phi(\cdot) + \varepsilon_i \tag{2.5}$$

$$\text{bid price: } P_i^b = \mu_{i-1} - \theta(\cdot)(1 + E[Q_i|Q_{i-1}]) - \phi(\cdot) + \varepsilon_i \tag{2.6}$$

Contrary to a market with a specialist where some transactions may be executed inside the spread, all buys (sells) with a smaller or equal volume than the best depth are executed at the prevailing best ask (bid) price. Trades inside the spread are not possible. It can easily be shown that $E[Q_i|Q_{i-1}] = \rho Q_{i-1}$ where ρ is the first order autocorrelation of the trade indicator series. The equation for the transaction price can be expressed as

$$P_i = \mu_i + \phi(\cdot) \cdot Q_i + \xi_i \tag{2.7}$$

where ξ_i is an i.i.d mean zero disturbance term which accounts for possible rounding errors due to price discretion. Combining (2.4) and (2.7) yields the following equation for transaction price changes:

$$\Delta P_i = \theta(\cdot)(Q_i - \rho Q_{i-1}) + \phi(\cdot)(Q_i - Q_{i-1}) + \varepsilon_i + \xi_i - \xi_{i-1}. \tag{2.8}$$

We extend the basic MRR model in the following way. In the spirit of Dufour and Engle (2000), we specify the MRR model parameters $\phi(\cdot)$ and $\theta(\cdot)$ as a function of time and the duration between trades. Both parameters are assumed to depend on time of day dummies $d_{m,i}$ which accounts for the stylized fact that the spread has a pronounced deterministic time of day pattern. Following Dufour and Engle (2000), we also allow the log-duration between the last and the current trade to determine the parameter $\theta(\cdot)$ which measures the price impact of a trade. Specifically, we write

$$\phi(t_i) = \gamma^\phi + \sum_{m=1}^{M} \lambda_m^\phi d_{m,i} \tag{2.9}$$

$$\theta(T_i, t_i) = \gamma^\theta + \sum_{m=1}^{M} \lambda_m^\theta d_{m,i} + \delta \ln T_i \tag{2.10}$$

where t_i is the time of event i and T_i denotes the duration between the trade in t_{i-1} and time t_i. As in Dufour and Engle (2000), we add one second to each duration before taking logarithms in order to avoid negative values. Incorporating a deterministic time of day pattern in the objective function directly instead of estimating the model separately for different periods of the day has the advantage that we can easily check for statistical significance of the estimated parameters λ_m^ϕ and λ_m^θ. Since the price in $t-1$ is $P_{i-1} = \mu_{i-1} + \phi(t_{i-1})Q_{i-1} + \xi_{i-1}$, equation (2.8) can now be written as:

$$\Delta P_i = \theta(T_i, t_i)(Q_i - \rho Q_{i-1}) + \phi(t_i)Q_i - \phi(t_{i-1})Q_{i-1} + \varepsilon_i + \xi_i - \xi_{i-1} \tag{2.11}$$

Moment conditions can be derived as follows. Denoting $u_i = \varepsilon_i + \xi_i - \xi_{i-1}$, we can write

$$\begin{aligned} u_i \;&=\; \Delta P_i - \Big[\Big(\gamma^\phi + \sum_{m=1}^{M} \lambda_m^\phi d_{m,i}\Big)Q_i - \Big(\gamma^\phi + \sum_{m=1}^{M} \lambda_m^\phi d_{m,i-1}\Big)Q_{i-1} \\ &\quad + \Big(\gamma^\theta + \sum_{m=1}^{M} \lambda_m^\theta d_{m,i} + \delta \ln T_i\Big) \cdot (Q_i - \rho Q_{i-1})\Big] \end{aligned} \tag{2.12}$$

Together with a vector of time-of-day dummy variables $\mathbf{d}_i = (\; d_{1,i}, \;\; \cdots \;\; , d_{M,i} \;)'$ and $\mathbf{z}_i = (\; Q_i, \;\; Q_{i-1} \;)'$, the resulting moment conditions are given by

$$E \begin{bmatrix} Q_i Q_{i-1} - \rho \\ u_i \\ u_i \mathbf{z}_i \\ u_i \mathbf{d}_i Q_i \\ u_i \mathbf{d}_{i-1} Q_{i-1} \\ u_i T_i \mathbf{z}_i \end{bmatrix} = 0 \tag{2.13}$$

The first three moment conditions are the same as in the standard MRR model. The next two moment conditions result from the inclusion of the time-of-day dummies and the last one is due to the inclusion of the duration.

Madhavan, Richardson, and Roomans (1997) have argued that the *surprise* in order flow rather than order flow itself affects the fundamental value of an asset. If order flow is predictable, using raw order flow would imply that the fundamental value μ_i depends on information of time t_{i-2}. Otherwise the fundamental asset value is not a martingale any longer. Engle and Russell (1998) have shown that trade durations are also predictable. They have a strong time-of-day (diurnal) component, but beyond that even diurnally adjusted durations exhibit a strong serial correlation. It thus seems sensible to account for predictability of trade durations. Instead of modeling the evolution of the fundamental asset value as a function of raw trade durations, we assume that duration *shocks* have an impact (via the trades) on the evolution of the asset value. This requires a decomposition of the trade duration sequence into a predictable and an unpredictable component. Following Engle and Russell (1998), we split up trade durations into three components, a diurnal time-of-day dependent component, a predictable component and a duration shock. Specifically, we have

$$T_i = \Phi(t_i) \cdot \psi_i \cdot \nu_i \tag{2.14}$$

where $\Phi(t_i)$ is the diurnal pattern of durations and ψ_i evolves as

$$\psi_i = \omega + \alpha \tilde{T}_{i-1} + \beta \psi_{i-1} \tag{2.15}$$

where $\tilde{T}_i = T_i/\Phi(t_i)$. $\Phi(t_i)\psi_i$ is the conditional expected duration $E(T_i|\mathcal{F}_{i-1})$ and ν_i is an i.i.d. duration shock with $E(\nu_i) = 1$.

The alternative specification for the trade informativeness parameter $\theta(\cdot)$ is then written as:

$$\theta(\tilde{\nu}_i, t_i) = \gamma^\theta + \sum_{m=1}^{M} \lambda_m^\theta d_{m,i} + \delta \ln \tilde{\nu}_i \tag{2.16}$$

where $\tilde{\nu}_i = \tilde{T}_i/\psi_i + 1$.

Two step estimation of the ACD model parameters is feasible by first estimating the intra-day pattern $\Phi(t_i)$ with polynomial trigonometric regression (see Eubank and Speckman (1990)). The seasonally adjusted durations can then be computed as $\tilde{T}_i = T_i/\hat{\Phi}(t_i)$. In the second step, the parameters of the ψ_i equation can be estimated by Maximum Likelihood. Joint estimation of ACD and structural parameters is also feasible. Grammig and Wellner (2002) show how estimation of ACD model parameters can be performed in a GMM framework. Drawing on their analysis, the moment conditions that estimate the extended MRR model parameters along with the ACD parameters are as follows:

$$E \begin{bmatrix} Q_i Q_{i-1} - \rho \\ u_i \\ u_i \mathbf{d}_i Q_i \\ u_i \mathbf{d}_{i-1} Q_{i-1} \\ u_i \mathbf{z}_i \\ u_i \tilde{\nu}_i \mathbf{z}_i \\ \nu_i - 1 \\ (\nu_i - 1)(\nu_{i-1} - 1) \\ \vdots \\ (\nu_i - 1)(\nu_{i-J} - 1) \end{bmatrix} = 0 \tag{2.17}$$

The first set of moment conditions are the same as in equation (2.13) except that T_i is now substituted with $\tilde{\nu}_i$. The other moment conditions identify the ACD parameters. They make use of the assumption that $E(\nu_i) = 1$ and that the covariance between ν_i and ν_{i-j} for all $j > i$ is assumed to be zero.

2.4 Results

Table 2.4.1 report the estimation results of the extended MRR model. We focus on the results obtained from the model in which the price impact of trades depend on duration shocks. The model in which raw durations are assumed to affect the price impact of trades delivers qualitatively the same results and supports the same conclusions (results are available upon request). To account for diurnal effects the time-of-day dummies in equations 2.10 and 2.16 are chosen to indicate the following six periods of the trading day: 9:00a.m - 9:30a.m; 9:30a.m - 11:00a.m; 11:00a.m - 2:00p.m; 2:00p.m - 3:30p.m; 3:30p.m - 5:00p.m, 5:00p.m - 5:30p.m. When constructing the time-of-day dummies, the reference period is the mid-day period ranging from 11:00a.m - 2:00p.m..

Table 2.4.1: **Estimation results of the extended MRR model with ACD shocks.** The first two columns report the first stage GMM estimates and p-values based on Newey-West standard errors averaged across all stocks. The third column reports the number of significant ($\alpha = 1\%$) parameters. The remaining columns show the same results for the four sub-samples sorted by trading activity. The spread components of the extended MRR model are specified as a function of the time of day. Additionally, the adverse selection component depends on the duration between the trade in t_i and t_{i-1}:

$$\phi(t_i) = \gamma^{\phi} + \Sigma_{m=1}^{M} \lambda_m^{\phi} D_{m,i}$$

$$\theta(\tilde{\nu}_i, t_i) = \gamma^{\theta} + \Sigma_{m=1}^{M} \lambda_m^{\theta} D_{m,i} + \delta \ln \tilde{\nu}_i$$

We included five dummy variables to capture the deterministic time of day pattern. The period from 11:00 a.m. to 2:00 p.m. is the reference for both equations.

	Overall			1^{st} Quartile (most active)			2^{nd} Quartile			3^{rd} Quartile			4^{th} Quartile (least active)		
	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]
δ	0.0043	(0.00)	[30, 0]	0.0040	(0.00)	[7, 0]	0.0034	(0.00)	[8, 0]	0.0031	(0.00)	[8, 0]	0.0069	(0.00)	[7, 0]
γ^{ϕ}	0.0052	(0.00)	[30, 0]	0.0063	(0.00)	[7, 0]	0.0048	(0.00)	[8, 0]	0.0045	(0.00)	[8, 0]	0.0056	(0.00)	[7, 0]
λ_1^{ϕ}	0.0030	(0.01)	[28, 0]	0.0020	(0.00)	[7, 0]	0.0017	(0.00)	[8, 0]	0.0024	(0.00)	[7, 0]	0.0060	(0.03)	[6, 0]
λ_2^{ϕ}	0.0003	(0.23)	[12, 0]	0.0003	(0.26)	[4, 0]	0.0004	(0.12)	[5, 0]	0.0004	(0.14)	[3, 0]	0.0002	(0.45)	[0, 0]
λ_4^{ϕ}	-0.0003	(0.29)	[3, 4]	-0.0002	(0.32)	[1, 1]	-0.0001	(0.32)	[2, 1]	-0.0005	(0.17)	[0, 1]	-0.0006	(0.37)	[0, 1]
λ_5^{ϕ}	-0.0005	(0.23)	[1, 7]	-0.0008	(0.34)	[0, 3]	-0.0002	(0.21)	[1, 1]	-0.0006	(0.05)	[0, 3]	-0.0004	(0.33)	[0, 0]
λ_6^{ϕ}	0.0007	(0.16)	[14, 2]	0.0003	(0.20)	[4, 1]	0.0006	(0.02)	[6, 0]	0.0001	(0.28)	[0, 1]	0.0018	(0.13)	[4, 0]
γ^{θ}	0.0040	(0.00)	[30, 0]	0.0033	(0.00)	[7, 0]	0.0033	(0.00)	[8, 0]	0.0031	(0.00)	[8, 0]	0.0066	(0.00)	[7, 0]
λ_1^{θ}	0.0051	(0.03)	[28, 0]	0.0032	(0.03)	[6, 0]	0.0029	(0.09)	[7, 0]	0.0043	(0.00)	[8, 0]	0.0103	(0.00)	[7, 0]
λ_2^{θ}	0.0012	(0.15)	[14, 2]	0.0011	(0.02)	[4, 1]	0.0004	(0.37)	[1, 1]	0.0009	(0.17)	[4, 0]	0.0024	(0.01)	[5, 0]
λ_4^{θ}	0.0002	(0.23)	[3, 5]	0.0003	(0.12)	[2, 1]	0.0000	(0.23)	[0, 1]	0.0002	(0.10)	[1, 3]	0.0003	(0.47)	[0, 0]
λ_5^{θ}	0.0002	(0.29)	[5, 4]	0.0009	(0.14)	[3, 1]	-0.0001	(0.34)	[0, 2]	0.0002	(0.36)	[1, 1]	0.0000	(0.32)	[1, 0]
λ_6^{θ}	-0.0003	(0.28)	[1, 11]	-0.0003	(0.08)	[0, 5]	-0.0008	(0.09)	[0, 5]	0.0003	(0.40)	[1, 1]	-0.0005	(0.55)	[0, 0]
ρ	0.2204	(0.00)	[30, 0]	0.2203	(0.00)	[7, 0]	0.2067	(0.00)	[8, 0]	0.2113	(0.00)	[8, 0]	0.2465	(0.00)	[7, 0]
ω	0.0721	(0.00)	[30, 0]	0.0842	(0.00)	[7, 0]	0.0714	(0.00)	[8, 0]	0.0641	(0.00)	[8, 0]	0.0700	(0.00)	[7, 0]
α	0.1252	(0.00)	[30, 0]	0.1544	(0.00)	[7, 0]	0.1354	(0.00)	[8, 0]	0.1121	(0.00)	[8, 0]	0.0994	(0.00)	[7, 0]
β	0.8050	(0.00)	[30, 0]	0.7659	(0.00)	[7, 0]	0.7960	(0.00)	[8, 0]	0.8248	(0.00)	[8, 0]	0.8320	(0.00)	[7, 0]

The estimation results confirm some well known stylized facts of financial market microstructure. The adverse selection component is significantly higher in the first half hour of the day, a result that is consistent with previous research studying the intra-day pattern of the spread. Furthermore, the order processing cost component is significantly higher at the end of the day for the vast majority of stocks. This indicates that liquidity providers demand a compensation for holding inventory overnight.

The vast majority of the dummy variables for the remaining time periods are not significantly different from the mid-day reference period. The ∪-shaped pattern of the effective spread in Xetra is therefore due to higher adverse selection costs as well as higher order processing costs in the morning and higher order processing costs shortly before closing.

Let us now focus on what the results imply regarding an effect of a duration shock on trade informativeness. The results reported in table 2.4.1 show that the estimates of the key parameter δ are positive and significantly different from zero for all 30 stocks. This implies that longer no-trade intervals are associated with increasing information related costs of a trade. As such this results stands in sharp contrast with the results reported by Dufour and Engle (2000). We will discuss the reasons and provide explanations for the contradicting results in the next section. Let us first focus on assessing the economic importance of the results beyond statistical significance.

For the purpose of assessing the economic significance of our results, we split the adverse selection component $\theta(T_i, t_i)$ into a deterministic part

$$\theta(t_i) = \gamma^\theta + \sum_{m=1}^{M} \lambda_m^\theta d_{m,i}$$

and a part explained by the duration shock of the subsequent no-trade interval

$$\theta(\tilde{\nu}_i) = \delta \ln \tilde{\nu}_i.$$

Both terms constitute the complete adverse selection component, $\theta(\tilde{\nu}_i, t_i) = \theta(t_i) + \theta(\tilde{\nu}_i)$. We can then compute for each stock the adverse selection share of the spread

$$asr(\tilde{\nu}_i, t_i) = \frac{\theta(\tilde{\nu}_i, t_i)}{\theta(\tilde{\nu}_i, t_i) + \phi(t_i)},$$

Table 2.4.2: **Adverse selection in percent of the spread.** The table reports several average adverse selection shares for each trade activity quartile. In the first column, $asr(\tilde{\nu}_i, t_i) = \frac{\theta(\tilde{\nu}_i, t_i)}{\theta(\tilde{\nu}_i, t_i)+\phi(t_i)} \times 100$ denotes the average adverse selection share of the spread. Additionally, in the second column, $asr(\tilde{\nu}_i) = \frac{\theta(\tilde{\nu}_i)}{\theta(\tilde{\nu}_i, t_i)+\phi(t_i)} \times 100$ denotes the average adverse selection share of the spread explained by trade duration. $dasr$ denotes the average fraction of the adverse selection component which can be explained by trade duration or $dasr = \frac{\theta(\tilde{\nu}_i)}{\theta(\tilde{\nu}_i, t_i)} \times 100$. The three ratios are averaged over all trades for each trade activity quartile and for the whole sample. Note, that the adverse selection component $\theta(\tilde{\nu}_i, t_i)$ is the sum of the deterministic time of day component $\theta(t_i)$ and the duration dependent component $\theta(\tilde{\nu}_i)$.

Trading Activity Quartile	$asr(\tilde{\nu}_i, t_i)$	$asr(\tilde{\nu}_i)$	$dasr$
1^{st} Quartile (most active)	45.0	13.4	26.4
2^{nd} Quartile	48.2	14.1	26.6
3^{rd} Quartile	50.7	14.0	25.3
4^{th} Quartile (least active)	63.8	18.0	26.5
Overall	48.8	14.2	26.3

the adverse selection share of the spread due to duration

$$asr(\tilde{\nu}_i) = \frac{\theta(\tilde{\nu}_i)}{\theta(\tilde{\nu}_i, t_i) + \phi(t_i)},$$

and the share of adverse selection explained by duration to the total adverse selection component

$$dasr = \frac{\theta(\tilde{\nu}_i)}{\theta(\tilde{\nu}_i, t_i)}.$$

Table 2.4.2 shows that the information related share of the implied effective spread is highest (63.8%) for the least actively traded Xetra stocks. In contrast, the average adverse selection share of trade activity quartile 1 only amounts to 45.0%.[4].

Table 2.4.2 also shows that the effect of trade durations on trade informativeness is also stronger for less frequently traded stocks. We focus our attention on the indicator $asr(\tilde{\nu}_i)$ which measures the importance of the duration component relative to the complete spread. The mean of the indicator $asr(\tilde{\nu}_i)$ ranges from 13.4% for Xetra trade activity quartile 1 to 18.0% for Xetra trade activity quartile 4. Averaged across stocks, the share of the duration component relative to the spread amounts to 14.2%.

[4]Note, that this number is strongly influenced by the stock DTE which has an $asr(\tilde{\nu}_i, t_i)$ of 22.4% while all the other stocks in the quartile have an $asr(\tilde{\nu}_i, t_i)$ of 46.3% and higher. However, compared to trade activity 4, the information related component of the spread is substantially smaller.

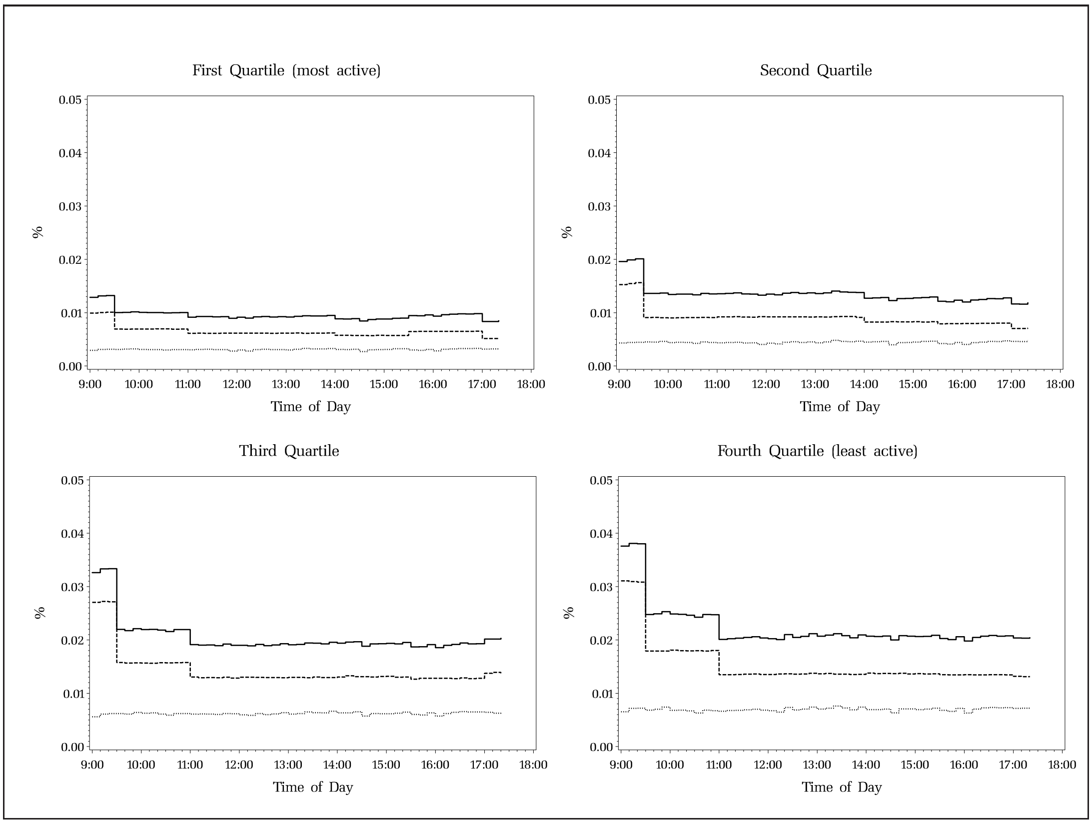

Figure 2.4.1: **Intra-day patterns for the estimated standardized adverse selection components.** The dotted line depicts the average standardized adverse selection component due to duration $\theta(\tilde{\nu}_i)$. The dashed line depicts the deterministic part of the average standardized adverse selection component $\theta(t_i)$. The solid line depicts the sum of $\theta(\tilde{\nu}_i)$ and $\theta(t_i)$, the complete adverse selection component $\theta(\tilde{\nu}_i, t_i)$. Top left: Intra-day patterns for the trade activity quartile 1. Top right: Intra-day patterns for the trade activity quartile 2. Lower left: Intra-day patterns for the trade activity quartile 3. Lower right: Intra-day patterns for the trade activity quartile 4.

Figure 2.4.1 provides a graphical illustration of the intra-day pattern of the adverse selection component. As above, we eliminate the price level effect by dividing the spread components by the average mid-quote of the respective stock. While the deterministic pattern was estimated for six periods of the day, the duration component varies with every trade. To capture any possible systematic intra-day variation in the duration component, but not overload the figure, we compute ten minute means for the duration component. One can see that the standardized duration component does not vary substantially throughout the day but rather floats around a constant mean. In contrast, the deterministic portion resembles the well known L-shaped intra-day pattern of the adverse selection component. Adding up the two parts yields the complete adverse selection component.

Table 2.4.3: **Correlations of the estimated standardized spread components with several relative spread measures, market capitalization and the daily number of trades.** The table reports the Pearson correlation coefficients of the estimated standardized spread components with several relative spread measures, market capitalization and the daily number of trades. $\tilde{\phi}(t_i)$ is the standardized order processing component, $\tilde{\theta}(\tilde{\nu}_i, t_i)$ is the standardized adverse selection component, $\tilde{\theta}(\tilde{\nu}_i)$ is the standardized adverse selection component due to duration and $\tilde{IS}_i = 2[\tilde{\theta}(\tilde{\nu}_i, t_i) + \tilde{\phi}(t_i)]$ denotes the implied spread. $asr(\tilde{\nu}_i, t_i) = \frac{\theta(\tilde{\nu}_i, t_i)}{\theta(\tilde{\nu}_i, t_i) + \phi(t_i)}$ denotes the average adverse selection share of the spread computed for each stock. $asr(\tilde{\nu}_i) = \frac{\theta(\tilde{\nu}_i)}{\theta(\tilde{\nu}_i, t_i) + \phi(t_i)}$ denotes the average adverse selection share of the spread explained by trade duration computed for each stock. $dasr = \frac{\theta(\tilde{\nu}_i)}{\theta(\tilde{\nu}_i, t_i)}$ denotes the average fraction of the adverse selection component which can be explained by trade duration. Correlations were computed across the sample of the 30 stocks constituting the DAX30. P-values for the correlation coefficients are in parentheses.

Variable	Effective Spread (%)	Realized Spread (%)	Price Impact (%)	Market cap. (Mill.)	Daily nb. trades
$\tilde{\phi}(t_i)$	0.763	0.881	0.373	-0.351	-0.153
	(0.000)	(0.000)	(0.043)	(0.057)	(0.419)
$\tilde{\theta}(\tilde{\nu}_i, t_i)$	0.782	-0.144	0.965	-0.802	-0.893
	(0.000)	(0.448)	(0.000)	(0.000)	(0.000)
$\tilde{\theta}(\tilde{\nu}_i)$	0.624	-0.320	0.884	-0.786	-0.890
	(0.000)	(0.085)	(0.000)	(0.000)	(0.000)
$\tilde{IS}_i$	0.996	0.505	0.845	-0.730	-0.653
	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)
$asr(\tilde{\nu}_i, t_i)$	-0.044	-0.827	0.410	-0.334	-0.565
	(0.816)	(0.000)	(0.024)	(0.071)	(0.001)
$asr(\tilde{\nu}_i)$	-0.286	-0.855	0.152	-0.191	-0.367
	(0.125)	(0.000)	(0.422)	(0.311)	(0.046)
$dasr$	-0.436	-0.486	-0.223	0.072	0.053
	(0.016)	(0.006)	(0.237)	(0.707)	(0.782)

Table 2.4.3 reports further evidence for the negative relationship between the duration effect and trading frequency. The correlation coefficient is approximately -0.5 and significant. Moreover, the correlations of the estimated standardized spread components with their observed "counterparts", the relative quoted spread, the relative effective spread, the relative realized spread and the relative price impact as well as the market capitalization, and the trading frequency measured in daily number of trades have the expected signs and are significant. For example, the adverse selection component is positively correlated with the price impact while the correlation of the estimated order processing cost component is, if at all, only weakly related to the price impact. Another expected result is that order processing costs, mainly consisting of institutional fees, are not related to size or trading frequency of

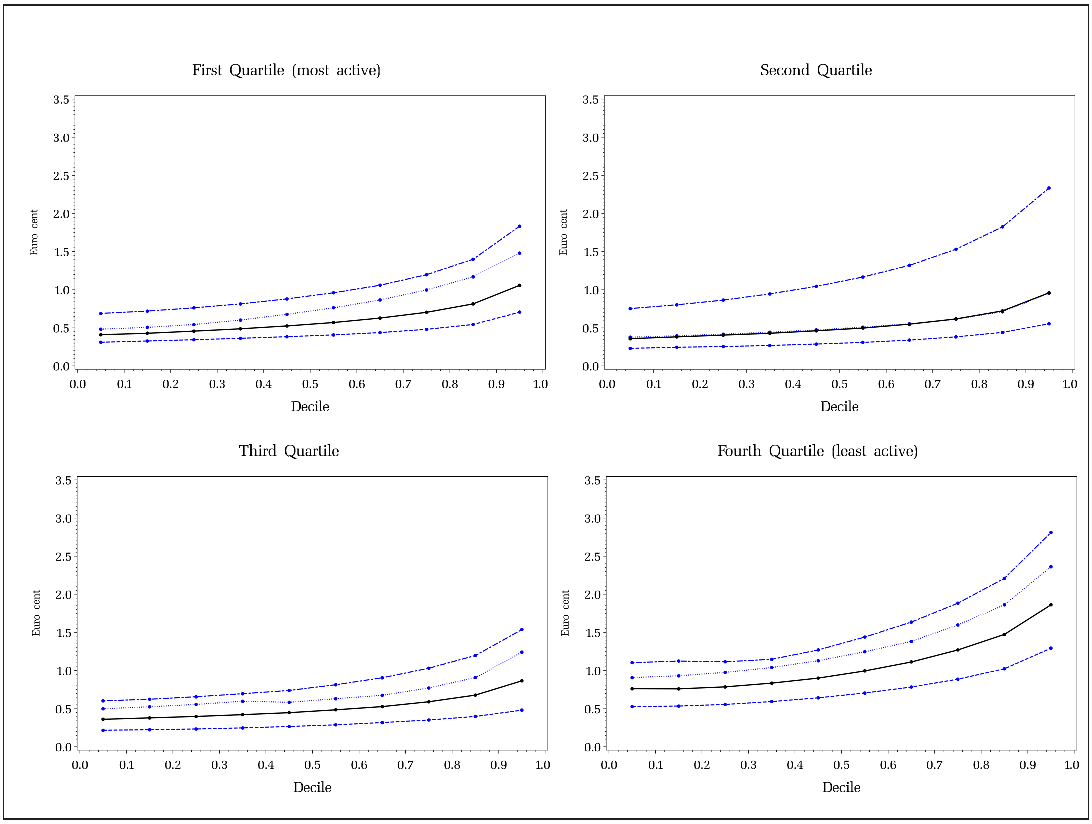

Figure 2.4.2: **Time between trades versus adverse selection component.** We sort (in ascending order) the trade duration shocks into deciles and compute the mean, 0.25, 0.75 and 0.9 quantile of the adverse selection component $\theta(\tilde{\nu}_i, t_i)$ in each decile and graphically display the results. The 0.25-quantiles are connected with dashed lines. The 0.75-quantiles are connected with dotted lines. The 0.9-quantiles are connected with dash-dotted lines. The decile means are connected with solid lines. All trade events of the stocks belonging to the same trading activity quartile are pooled. The top left panel displays the results for the group of most frequently trades stocks. The top right panel shows the results for the second and the lower left panel depicts the result for the third trading activity quartile. The lower right panel presents the results for the least frequently traded stocks.

the stock. In contrast, adverse selection is strongly negatively correlated with both, size and trading frequency.

The importance of trading intensity for the information content of a trade is further illustrated in figures 2.4.2 and 2.4.3. To produce these plots we have sorted all trade durations for groups of stocks into deciles. Decile 1 contains the smallest duration shocks while decile 10 contains the largest duration shocks. For each decile, we calculate the average standardized adverse selection component $\tilde{\theta}(\tilde{\nu}_i, t_i)$ and the average raw adverse selection component $\theta(\tilde{\nu}_i, t_i)$. The figures depict averages for each trade activity quartile. As can be seen in figure 2.4.3, even in the quartile with the most actively traded stocks the adverse selection component

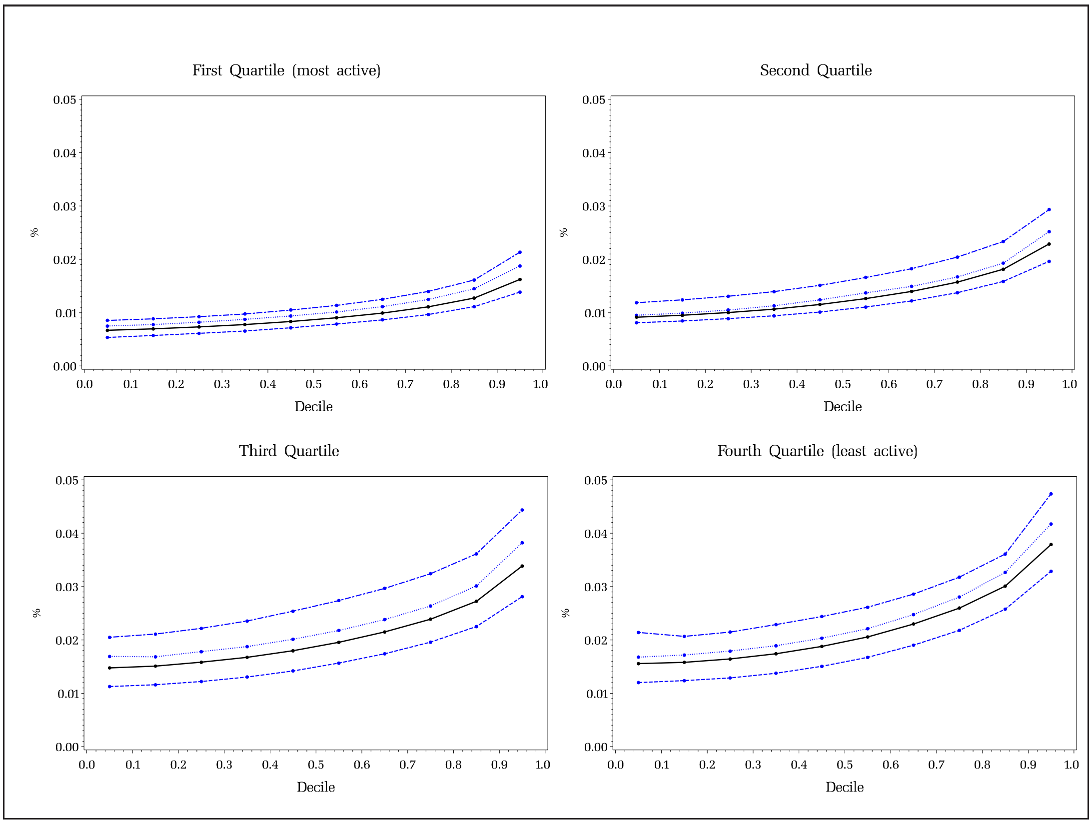

Figure 2.4.3: **Time between trades versus standardized adverse selection component.** We sort (in ascending order) the trade duration shocks into deciles and compute the mean, 0.25, 0.75 and 0.9 quantile of the standardized adverse selection component $\widetilde{\theta}(\tilde{\nu}_i, t_i)$ in each decile and graphically display the results. The 0.25-quantiles are connected with dashed lines. The 0.75-quantiles are connected with dotted lines. The 0.9-quantiles are connected with dash-dotted lines. The decile means are connected with solid lines. All trade events of the stocks belonging to the same trading activity quartile are pooled. The top left panel displays the results for the group of most frequently trades stocks. The top right panel shows the results for the second and the lower left panel depicts the result for the third trading activity quartile. The lower right panel presents the results for the least frequently traded stocks.

doubles from trade intensity decile 1 to decile 10. A much stronger effect can be observed for the fourth quartile containing observations of the least actively traded stocks. Here, the adverse selection component more than triples when comparing the shortest trade durations in decile 1 with the longest trade durations in decile 10. In all four trade activity quartiles especially very large duration shocks have a large impact on the asset price. The slope of the line connecting the mean of decile 9 with the mean of decile 10 is steeper in every quartile.

Note, that this result is not due to cross section or time-of-day variations of the trade durations. We have argued above that stocks traded less frequently tend to have higher adverse selection costs. To confirm that the upward sloping curve in figure 2.4.3 is not an

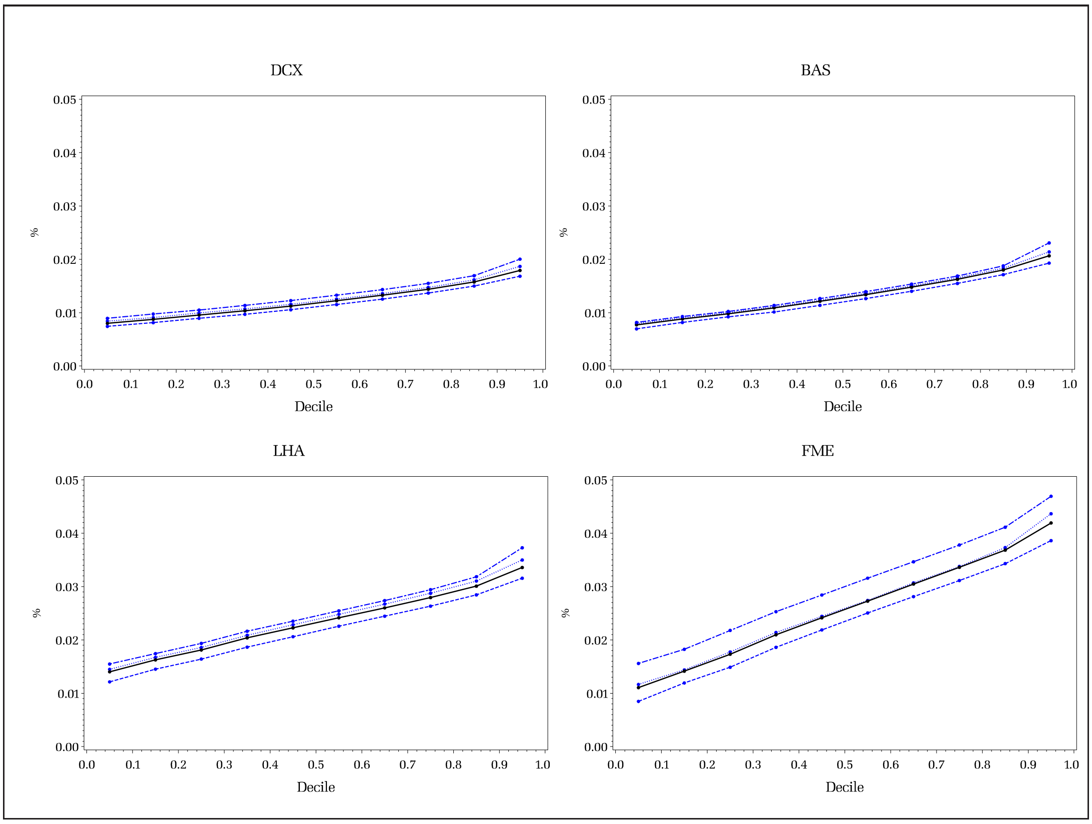

Figure 2.4.4: **Time between trades versus standardized adverse selection component for individual stocks.** We sort (in ascending order) the trade duration shocks into deciles and compute the mean, 0.25, 0.75 and 0.9 quantile of the adverse selection component $\theta(\tilde{\nu}_i, t_i)$ in each decile and graphically display the results. The 0.25-quantiles are connected with dashed lines. The 0.75-quantiles are connected with dotted lines. The 0.9-quantiles are connected with dash-dotted lines. The decile means are connected with solid lines. The top left panel displays the results for a representative stock of trade activity quartile 1. The top right panel shows the results for a representative stock of trade activity quartile 2. The lower left panel depicts the result for a representative stock of trade activity quartile 3 and the lower right panel presents the results for a representative stock of trade activity quartile 4. All trade events of the particular stock are pooled.

artefact caused by intra-group variation of trading frequency in each trade activity quartile, we provide additional figures for a selected representative stock in each trade activity quartile in figure 2.4.4. We find for individual stocks the same relation between duration and adverse selection.

Note that the duration effect can also not be ascribed to co-movements in the intra-daily pattern of the trade duration and the adverse selection component. We have seen in figure 2.4.1 that the information induced part of the spread is high in the first half hour and lower for the rest of the day.

Figure 2.4.5 shows that trade durations rather have an inverted ∪-shaped intra-day pat-

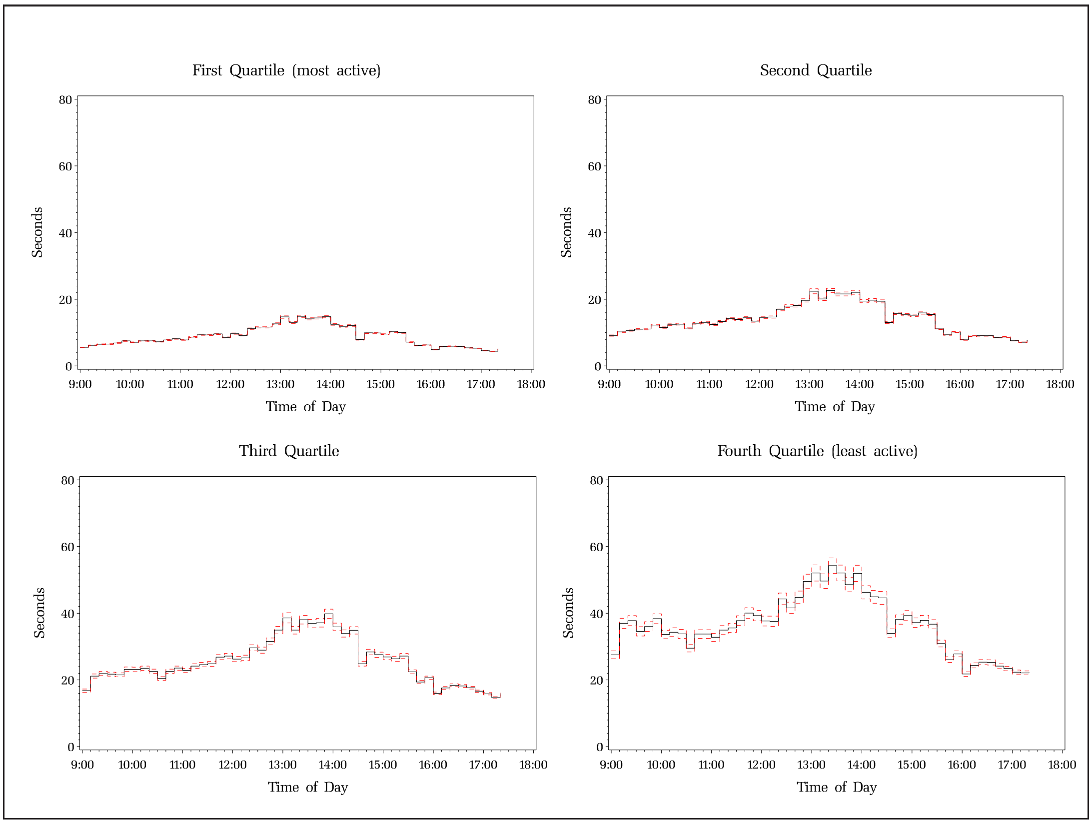

Figure 2.4.5: **Intra-day pattern of trade durations.** We compute for each ten minute interval of the day the average trade duration and plot the means against time of day. All trade events of the stocks belonging to the same trading activity quartile are pooled. The top left panel displays the results for the group of most frequently trades stocks. The top right panel shows the results for the second and the lower left panel depicts the result for the third trading activity quartile. The lower right panel presents the results for the least frequently traded stocks. The dashed lines are the 95% confidence intervals for the ten minute means.

tern. At the beginning of the trading day, when adverse selection costs are high, trade durations tend to be short. Hence, we would rather expect a dampening of the positive duration effect through the intra-day variation. Therefore, we conclude that neither intra-group variation in the average trading frequency of the stocks nor the intra-day pattern in the deterministic part of the adverse selection component is responsible for the strong impact of trade durations.

As a robustness check we also estimate the model for a matched sample of NYSE traded stocks. The matching variable is the daily trading volume. The NYSE stocks included in the matched sample and their Xetra counterparts are reported in Table 2.4.4.

Ticker	Company Name (Xetra)	Daily Turnover (Mill.€)	Ticker	Company Name (NYSE)	Daily Turnover (Mill. €)
DTE	DEUTSCHE TELEKOM	350.63	XOM	EXXON MOBIL	375.45
SIE	SIEMENS	321.70	JPM	J.P. MORGAN CHASE	334.96
DBK	DEUTSCHE BANK	309.28	JNJ	JOHNSON & JOHNSON	309.25
ALV	ALLIANZ	289.98	AIG	AMERICAN INT'L.	288.35
MUV2	MUENCH. RUECKVERS.	207.35	MWD*	MORGAN STANLEY	205.98
DCX	DAIMLERCHRYSLER	187.74	MDT	MEDTRONIC	188.80
SAP	SAP	184.63	WYE	WYETH	183.87
EOA	E.ON	160.63	ABT	ABBOTT LABS	160.45
IFX	INFINEON	146.46	KSS	KOHL'S	146.28
BAS	BASF	124.43	LMT	LOCKHEED MARTIN	123.88
VOW	VOLKSWAGEN	104.25	CAH	CARDINAL HEALTH	105.51
HVM	HYPO-VEREINSBANK	98.35	STJ	ST. JUDE MEDICAL	98.50
RWE	RWE	97.66	A	AGILENT TECHNOLOGIES	97.45
BAY	BAYER	88.78	ALL	ALLSTATE	88.28
BMW	BMW	87.85	HDI*	HARLEY DAVIDSON	88.26
CBK	COMMERZBANK	53.17	CVS	CVS	53.02
SCH	SCHERING	51.41	MHS	MEDCO HEALTH SOLUTIONS	51.30
LHA	LUFTHANSA	43.95	BDX	BECTON, DICKINSON	43.99
DPW	DEUTSCHE POST	43.84	RTN	RAYTHEON	43.83
MEO	METRO	38.87	JBL	JABIL CIRCUIT	38.76
TKA	THYSSENKRUPP	37.89	JCI	JOHNSON CONTROLS	37.93
DB1	DEUTSCHE BOERSE	35.70	BBT	BB & T	35.68
ADS	ADIDAS-SALOMON	31.98	DOV	DOVER	31.97
ALT	ALTANA	30.99	BNI	BURLINGTON NORTH. SANTA FE	30.93
MAN	MAN	27.69	MBI	MBIA	27.63
TUI	TUI	26.28	BCR	BARD (C.R.)	26.33
CONT	CONTINENTAL	25.63	BDK	BLACK & DECKER	25.68
LIN	LINDE AG	22.38	CBE	COOPER INDUSTRIES	22.32
HEN3	HENKEL	18.17	DYN	DYNEGY	18.09
FME	FRESENIUS MEDICAL CARE	12.85	TMK	TORCHMARK	12.92

Table 2.4.4: **Matched sample of NYSE traded stocks** For each DAX stock we compare the daily average traded volume to each NYSE traded stock of the S&P 500. We select the stock minimizing the absolute difference as a matching stock. * Both firms changed their ticker symbols in 2006. Here, we use the old ticker symbols available in our data set.

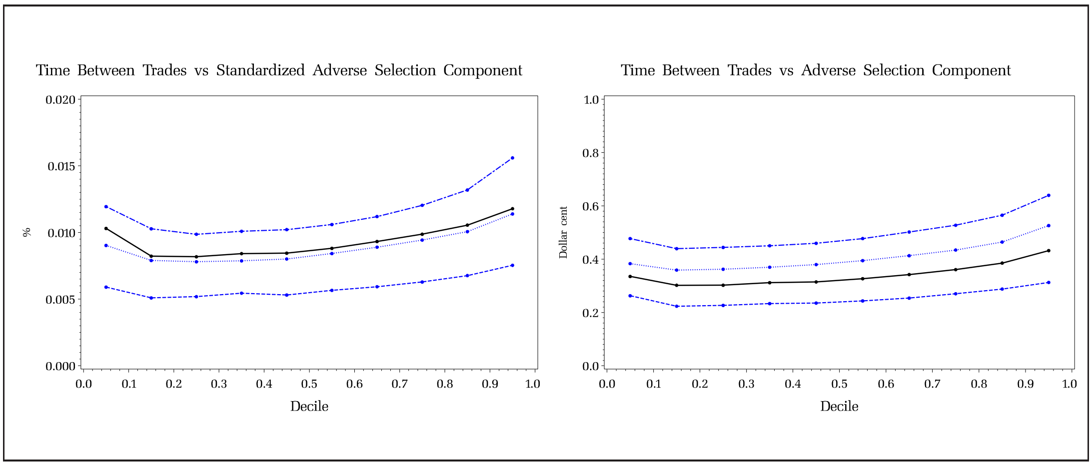

Figure 2.4.6: **Results for the NYSE traded matched sample.** We sort (in ascending order) the trade duration into deciles and compute the mean, 0.25, 0.75 and 0.9 quantile of the adverse selection component $\theta(\tilde{\nu}_i, t_i)$ (right panel) and the standardized adverse selection component $\tilde{\theta}(\tilde{\nu}_i, t_i)$ (left panel) in each decile and graphically display the results. The 0.25-quantiles are connected with dashed lines. The 0.75-quantiles are connected with dotted lines. The 0.9-quantiles are connected with dash-dotted lines. The decile means are connected with solid lines. All trade events of the 30 NYSE traded stocks are pooled.

Table A.1.1 shows that the conclusions for the Xetra/DAX stocks also hold for the NYSE sample: The adverse selection parameter δ is positive and significant for 22 of the 30 stocks. The plots in figure 2.4.6 show the same upward sloping curve that suggests that the informational content of trades is higher after relatively long no-trade periods. However, from an economic point of view, the informational importance of trade durations for the spread seems smaller for the NYSE stocks compared to the Xetra/DAX stocks.

2.5 Interpretation and Discussion

The statistical and economic significance of these results challenge the "no trade means no information" result of Dufour and Engle (2000) that is part of financial market microstructure's conventional wisdom. Rather than supporting the predictions of the Easley and O'Hara (1992) model, our findings are more in accord with the models of strategic trading in limit order markets (Parlour (1998), Foucault (1999)). Their contributions relate to a classic paper by Admati/Pfleiderer (1988). In the Admati/Pfleiderer model non-informed liquidity traders cluster during periods of the trading day which implies that high trading intensity would be associated with reduced trade informativeness. Traders with private information

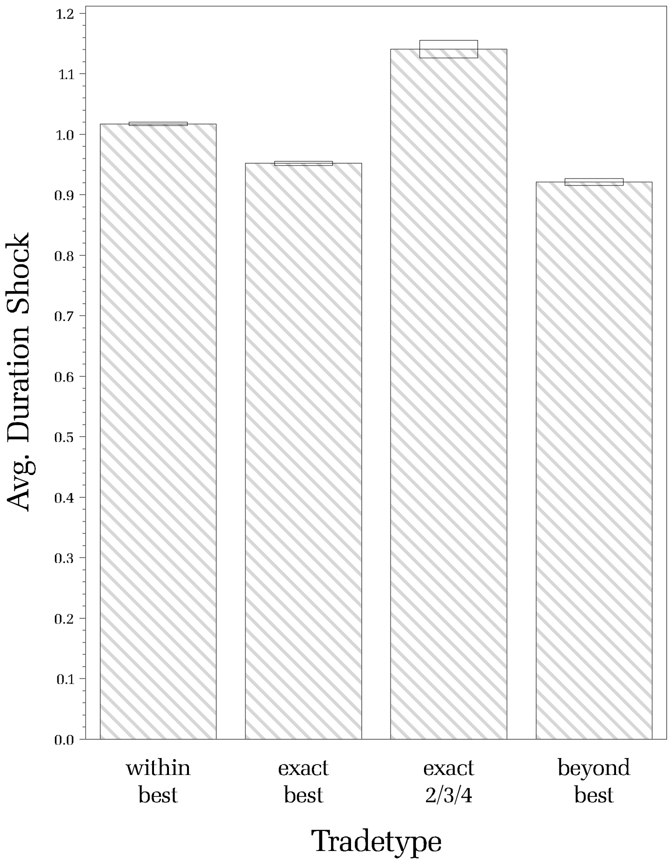

Figure 2.5.1: **Average duration shock for different trade categories.** The graph depicts the average duration shock for four different trade categories. We denote with 'within best' the trades with a volume smaller than the best depth. With 'exact best' we denote the trades consuming exactly the best depth. Trades consuming exactly the depth up to the second, third or fourth best quote are categorized as 'exact 2/3/4' and trades with a volume higher than the best depth but within any higher order quote are denoted as 'beyond best'.

may hide in the crowd of liquidity traders, the price impact of their trades is "cushioned" by the trading of the uninformed liquidity traders. The models by Parlour (1998) and Foucault (1999) provide more elaborate explanations for such a clustering process. In Parlour's (1998) model, large depth on the bid side increases buyer "aggressiveness", in other words: more buy market orders, more (buyer initiated) trading activity. The reasoning is symmetric for the sell side. Arguably, periods of high liquidity (measured as the depth at the best quotes) in limit order markets are associated with a lesser degree of private information present in the market. Patient limit order traders, who are not trading for reasons of exploiting their superior information, and who are not afraid of being "picked off" by an informed order or adverse price movement, will then supply ample liquidity. Thus, high liquidity in the order book emerges during non-informative periods. As liquidity traders gather together (via limit order submission, they supply liquidity to the market) during non-informative periods, trading becomes more aggressive (more market orders triggering trades) when impatient (yet not superiorly informed) traders strive to get priority over standing limit orders.

As a matter of fact, when estimating the Dufour/Engle model on our data, we find results that are, at least qualitatively, in line with their results (see table A.1.2) in that short trade durations are associated with informed trading. Hence, the difference must be due to the empirical methodology, i.e. the way these models process the data and not the differences of trading processes in a pure limit order market and a hybrid market. In the following we will offer an explanation for the differences. A convenient way to implement a strategy that aggressively strives for price priority a'la Parlour (1998) are so-called market-to-limit orders or limit orders with a limit price equal or better than the best quote (marketable limit orders). As a matter of fact, those order types are very popular and frequently used instruments. Both a partially filled buy market-to-limit order and a partially filled marketable limit order consumes the volume at the best ask and the volume that exceeds the depth at the best ask is entered, at the limit price, on top of the queue of the limit orders standing on the bid side.

As figure 2.5.1 shows, those trades that exactly consume the volume at the best quotes (from which roughly 45% are partially filled market-to-limit-orders or partially filled marketable limit orders) indeed have significantly smaller durations than trades that consume only part of the depth at the best quote. Note that in a basic MRR model framework, a market-to-limit order would be deemed uninformative while in a Hasbrouck-VAR setting the events triggered by a market-to-limit order would be regarded as highly informative. The numerical example depicted in table 2.5.1 makes that point clear: While the transaction price (P) in the sequence of events does not change at all, the midquote changes ΔMQ is dramatic. A market-to-limit order is a perfect tool for an impatient trader to gain price priority when own side liquidity is high. However, as the market-to-limit trader also immediately supplies liquidity (during times when liquidity is already high) it is hard to ascribe such a behavior to the exploitation of superior private information. It is rather the crowding-out effect in liquid markets described by Parlour (1998) and Foucault (1998) that we observe. In that perspective it seems more reasonable to view market-to-limit order trade events as a way to implement an aggressive trading strategy (without private information processing involved)

Table 2.5.1: **Numerical Example.**

	t_0	t_1	t_2	t_3
	book state at t_0	book state after small transaction (P=105)	book state after market-to-limit buy order with limit price $P = 105$ that takes best ask and improves the best bid	state of book after best bid is snatched ($P = 105$)
	110	110		110
	105	105	110	110
MQ	102.5	102.5	107.5	105
	100	100	105	100
	90	90	100	90
			90	
$\Delta\ MQ$		0	4.9%	-2.3%
ΔP		0	0	0

than deem these trades as highly informative with respect to the fundamental asset value.

2.6 Conclusion and Outlook

This paper provides new evidence regarding the role of time in measuring the informational content of trades. Two novelties characterize our contribution. First, instead of the vector autoregressive methodology employed by Dufour and Engle (2000), we advocate the use of a structural model and extend Madhavan et al's (1997) model to account for a time varying trade intensities. For that purpose we employ Russell and Engle's (1998) ACD model. We estimate the model on a cross section of stocks traded in an automated open order book market, the Xetra system maintained by the German Stock Exchange. For robustness checks, the models are also estimated on NYSE data. Xetra and NYSE trading processes are quite different. In Xetra there are no dedicated market makers, trading is anonymous, and a fully computerized trading protocol matches liquidity supply and demand using an open limit order book trading platform. As a matter of fact, these are the characteristic features of all large Continental European stock markets. One of the advantages of using data from an automated auction system is the excellent data quality. Misclassification of trades that haunts empirical microstructure analysis is not an issue.

Dufour and Engle's (2000) paper made a strong case for the argument that trading intensity carries informational content with respect to the price impact of a trade. Specifically, their results provide evidence for the notion that "no trade means no information" one of the key predictions implied by Easley and O'Hara's (1992) microstructure model and arguably part of the conventional wisdom of financial market microstructure. The results reported in this paper provide quite contrasting evidence. As in the Dufour/Engle paper we also find that "time matters", both from a statistical and an economic point of view. However, our results imply that the informational content of a trade increases with the duration since the last trade. Our analysis suggests that in an automated order book market, but also in an hybrid framework like the NYSE, the role of time in measuring trade informativeness is more in accord with the predictions derived from the Admati/Pfleiderer (1988) model, and models of strategic trading in limit order markets (Parlour (1998), Foucault (1999)). In these models, high trading intensity is caused by the submission of market orders by impatient yet uninformed traders who strive aggressively for priority of their orders when the liquidity on their own market side is high (small spread, ample depth). Liquidity supply is high, however, when limit order traders are not afraid of being picked of by an adverse price movement (be it induced by public or private information processing). These results revive the relevance of the Admati/Pfleiderer (1988) model's prediction that through clustering of liquidity-induced trading, short durations between trades are associated with smaller trade informativeness. We outline that the reason why the different methodologies (extended Hasbrouck-VAR and extended MRR model) deliver contradicting results is mainly rooted in the way market-to-limit orders are treated in the two econometric frameworks.

Besides providing an additional note to answer the question "Why do securities prices change and what is the role of trading intensity in the process?" our results have interesting policy implications for the design of trading systems. One could argue for the need of circuit breakers during high trading intensities as processing of (private) information may harm uninformed market participants. However, our results indicate that circuit breaking via a call auction mechanism or additional liquidity supply by a dedicated market maker, say, would rather be advisable after long non-trading intervals.

Appendix A

A.1 Tables

Table A.1.1: **Estimation results of the extended MRR model with ACD shocks for a matched sample of NYSE traded stocks.** The first two columns report the first stage GMM estimates and p-values based on Newey-West standard errors averaged across all stocks. The third column reports the number of significant ($\alpha = 1\%$) parameters. The spread components of the extended MRR model are specified as a function of the time of day. Additionally, the adverse selection component depends on the duration between the trade in t_i and t_{i-1}:

$$\phi(t_i) = \gamma^{\phi} + \Sigma_{m=1}^{M} \lambda_m^{\phi} D_{m,i}$$

$$\theta(\tilde{\nu}_i, t_i) = \gamma^{\theta} + \Sigma_{m=1}^{M} \lambda_m^{\theta} D_{m,i} + \delta \ln \tilde{\nu}_i$$

We included four dummy variables to capture the deterministic time of day pattern. The period from 11:30 a.m. to 2:00 p.m. is the reference for both equations.

	Overall		
	Avg. est.	Avg. p-val	# sig [pos, neg]
δ	0.0009	(0.06)	[25, 0]
γ^{ϕ}	0.0021	(0.00)	[30, 0]
λ_1^{ϕ}	0.0007	(0.05)	[19, 0]
λ_2^{ϕ}	0.0001	(0.20)	[13, 1]
λ_4^{ϕ}	0.0001	(0.20)	[7, 2]
λ_5^{ϕ}	0.0005	(0.09)	[19, 2]
γ^{θ}	0.0029	(0.00)	[30, 0]
λ_1^{θ}	0.0018	(0.03)	[25, 0]
λ_2^{θ}	0.0006	(0.05)	[19, 1]
λ_4^{θ}	-0.0001	(0.31)	[2, 6]
λ_5^{θ}	-0.0006	(0.05)	[0, 24]
ρ	0.2731	(0.00)	[30, 0]
ω	0.0457	(0.03)	[26, 0]
α	0.0468	(0.01)	[28, 0]
β	0.9077	(0.00)	[30, 0]

Table A.1.2: **Estimation results of the DE quote revision equation.** The first two columns report the p-values based on Newey-West standard errors averaged across all stocks. The third column reports the number of significant ($\alpha = 1\%$) parameters. The remaining columns show the same results for the four sub-samples sorted by trading activity. For convenience, we report only the parameters capturing the duration impact and the time of day dummy.

$$R_t = \sum_{j=1}^{5} a_j R_{t-j} + \gamma_{open} D_t Q_t + \sum_{j=0}^{5} b_j Q_{t-j} + v_{1,t}$$

$$\text{where} \quad b_j = \gamma_j + \delta_j \ln(T_{t-j})$$

For convenience, we report only the parameters capturing the duration impact δ_j and the time of day dummy γ_{open}.

	Overall			1^{st} Quartile (most active)			2^{nd} Quartile			3^{rd} Quartile			4^{th} Quartile (least active)		
	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]
δ_0	-0.0010	(0.00)	[0, 30]	-0.0009	(0.00)	[0, 7]	-0.0009	(0.00)	[0, 8]	-0.0010	(0.00)	[0, 8]	-0.0011	(0.00)	[0, 7]
δ_1	0.0002	(0.16)	[15, 0]	0.0003	(0.03)	[6, 0]	0.0003	(0.00)	[7, 0]	0.0002	(0.23)	[2, 0]	0.0001	(0.40)	[0, 0]
δ_2	0.0002	(0.14)	[16, 0]	0.0002	(0.00)	[7, 0]	0.0002	(0.09)	[5, 0]	0.0002	(0.13)	[3, 0]	0.0001	(0.35)	[1, 0]
δ_3	0.0001	(0.17)	[9, 0]	0.0001	(0.03)	[5, 0]	0.0002	(0.06)	[4, 0]	0.0001	(0.32)	[0, 0]	0.0001	(0.27)	[0, 0]
δ_4	0.0001	(0.15)	[10, 0]	0.0001	(0.07)	[5, 0]	0.0002	(0.03)	[3, 0]	0.0001	(0.22)	[0, 0]	0.0002	(0.30)	[2, 0]
δ_5	0.0002	(0.14)	[17, 0]	0.0002	(0.13)	[5, 0]	0.0002	(0.14)	[5, 0]	0.0003	(0.06)	[5, 0]	0.0002	(0.22)	[2, 0]
γ_{open}	0.0128	(0.00)	[30, 0]	0.0055	(0.00)	[7, 0]	0.0083	(0.00)	[8, 0]	0.0159	(0.00)	[8, 0]	0.0218	(0.00)	[7, 0]

A.2 Additional Results

This part of the appendix provides some additional results and presents a comparison of Xetra estimation results of the basic MRR model with their sample from NYSE.

Following MRR, we have been particularly interested in the intra-day patterns of spreads and its components. We refer to θ as the information asymmetry parameter as it gives the revision of the fundamental price due to a surprise in the order flow. ϕ is referred to as the non-informational related transaction cost element. It represents economic (opportunity) costs of market making. From the parameters ϕ and θ one can compute two indicators of interest. First, the implied (effective) spread $s = s_E = 2(\phi + \theta)$ and second, the share of the spread that is attributable to asymmetric information $r = \frac{\theta}{\phi+\theta}$. Table A.2.1 contains the first stage GMM estimates of θ and ϕ and ρ. As in MRR we report the cross sectional mean, median and standard deviation of the parameter estimates. The estimates are computed for six periods of the trading day, 9:00-9:30; 9:30-11:00; 11:00-14:00; 14:00-15:30;15:30-17:00, 17:00-17:30. The length and timing of these intervals is inspired from MRR's study and adapted to the Xetra trading hours.

Extending the MRR's aggregate view (who report only the aggregated results for their sample of 274 stocks), we also report the estimation results disaggregated for four groups with stocks sorted according to their trading intensity (see table 2.2.1 for the assignment of stocks into the trading frequency quartiles). We report the group mean and median as well as the within group standard deviation of the respective parameter estimate. This allows checking the homogeneity of the groups. Table A.2.2 reports the corresponding results for the implied spread $s = s_E = 2(\theta + \phi)$ as well as the share of the adverse selection component $r = \frac{\theta}{\phi+\theta}$.

Table A.2.1: **Estimation results for MRR for different periods of the day.** The table presents the mean coefficient estimate across stocks (first stage GMM) in the respective trade size quartile and across all thirty stocks, as well as standard deviation and median of the estimates within the respective group and across all thirty stocks. The estimates for θ and ϕ were multiplied with 100 in order to express the spread components in euro cents.

Trade Size Quartile		9:00-9:30	9:30-11:00	11:00-14:00	14:00-15:30	15:30-17:00	17:00-17:30
				θ			
1	Mean	0.91	0.67	0.57	0.61	0.65	0.57
	Median	0.91	0.67	0.55	0.56	0.59	0.52
	Std	0.49	0.34	0.29	0.31	0.35	0.29
2	Mean	0.83	0.58	0.50	0.53	0.52	0.47
	Median	0.76	0.51	0.44	0.47	0.47	0.45
	Std	0.60	0.44	0.38	0.40	0.33	0.30
3	Mean	0.90	0.57	0.50	0.51	0.51	0.53
	Median	0.66	0.45	0.42	0.40	0.44	0.44
	Std	0.51	0.30	0.24	0.27	0.24	0.26
4	Mean	2.02	1.26	1.06	1.07	1.03	1.02
	Median	1.67	1.25	0.94	1.05	1.07	0.99
	Std	0.77	0.40	0.38	0.33	0.32	0.30
all	Mean	1.15	0.76	0.65	0.67	0.66	0.64
	Median	0.90	0.64	0.54	0.56	0.58	0.53
	Std	0.75	0.45	0.39	0.39	0.36	0.35
				ϕ			
1	Mean	0.70	0.60	0.56	0.54	0.49	0.59
	Median	0.61	0.53	0.51	0.48	0.45	0.49
	Std	0.36	0.26	0.22	0.21	0.17	0.20
2	Mean	0.53	0.45	0.44	0.42	0.40	0.47
	Median	0.42	0.38	0.35	0.34	0.32	0.42
	Std	0.36	0.26	0.27	0.24	0.24	0.21
3	Mean	0.58	0.44	0.38	0.36	0.35	0.40
	Median	0.49	0.44	0.36	0.35	0.33	0.38
	Std	0.27	0.13	0.10	0.09	0.09	0.10
4	Mean	0.92	0.54	0.46	0.47	0.49	0.70
	Median	0.78	0.53	0.49	0.48	0.52	0.65
	Std	0.35	0.21	0.19	0.26	0.19	0.19
all	Mean	0.67	0.50	0.46	0.44	0.43	0.53
	Median	0.54	0.43	0.38	0.36	0.35	0.46
	Std	0.35	0.22	0.21	0.21	0.18	0.21
				ρ			
1	Mean	0.21	0.22	0.22	0.22	0.24	0.20
	Median	0.20	0.22	0.22	0.22	0.23	0.19
	Std	0.03	0.02	0.01	0.01	0.02	0.02
2	Mean	0.19	0.21	0.20	0.21	0.22	0.19
	Median	0.20	0.21	0.20	0.21	0.22	0.18
	Std	0.03	0.02	0.02	0.02	0.02	0.03
3	Mean	0.19	0.20	0.22	0.20	0.23	0.20
	Median	0.19	0.20	0.21	0.21	0.22	0.20
	Std	0.03	0.01	0.02	0.02	0.01	0.02
4	Mean	0.24	0.24	0.25	0.25	0.25	0.22
	Median	0.24	0.23	0.25	0.27	0.25	0.22
	Std	0.03	0.02	0.02	0.03	0.02	0.02
all	Mean	0.21	0.22	0.22	0.22	0.23	0.20
	Median	0.20	0.22	0.22	0.22	0.23	0.20
	Std	0.03	0.02	0.02	0.03	0.02	0.03

Table A.2.2: **Implied spread and adverse selection share of MRR for different periods of the day.** The table presents the mean coefficient estimate across stocks (calculated from the first stage GMM estimates) in the respective trade size quartile and across all thirty stocks, as well as standard deviation and median of the estimates within the respective group and across all thirty stocks. The estimates for s_E were multiplied with 100 in order to express the spread components in euro cents. The adverse selection share r is expressed in percent.

Trade Size Quartile		9:00-9:30	9:30-11:00	11:00-14:00	14:00-15:30	15:30-17:00	17:00-17:30
		$s_E = 2(\theta + \phi)$					
1	Mean	3.24	2.54	2.27	2.30	2.28	2.31
	Median	2.90	2.27	2.06	2.08	2.07	2.03
	Std	1.65	1.15	0.97	1.02	1.02	0.99
2	Mean	2.73	2.05	1.88	1.91	1.83	1.87
	Median	2.35	1.77	1.54	1.58	1.59	1.73
	Std	1.88	1.39	1.29	1.26	1.13	1.01
3	Mean	2.95	2.03	1.75	1.75	1.72	1.86
	Median	2.36	1.78	1.56	1.52	1.54	1.62
	Std	1.49	0.78	0.61	0.65	0.60	0.71
4	Mean	5.87	3.60	3.04	3.07	3.04	3.42
	Median	4.68	3.09	2.58	2.75	2.73	3.02
	Std	2.12	1.13	1.09	1.10	0.96	0.92
all	Mean	3.64	2.52	2.21	2.23	2.19	2.33
	Median	2.79	2.23	1.94	1.90	1.94	2.02
	Std	2.11	1.25	1.09	1.10	1.04	1.07
		$r = \frac{\theta}{\theta+\phi}$					
1	Mean	54.3	50.6	47.9	50.6	53.8	46.5
	Median	55.5	54.2	52.5	53.8	57.4	51.2
	Std	11.7	12.9	12.8	12.6	12.9	10.6
2	Mean	58.9	53.3	51.3	53.8	55.0	48.0
	Median	62.3	57.1	54.9	57.9	56.4	51.5
	Std	10.9	10.6	10.2	10.5	10.5	8.8
3	Mean	59.4	54.3	54.6	55.6	57.1	55.1
	Median	57.1	53.9	53.7	53.1	54.9	55.8
	Std	6.9	9.2	9.7	10.7	8.5	7.8
4	Mean	68.3	69.8	69.5	70.6	68.2	59.1
	Median	67.5	68.5	69.3	71.1	66.6	57.7
	Std	5.7	6.5	6.0	6.7	6.7	4.6
all	Mean	60.1	56.8	55.6	57.5	58.4	52.1
	Median	60.3	56.9	55.4	57.9	60.5	53.1
	Std	10.0	12.1	12.5	12.4	10.9	9.4

Table A.2.2 shows that the mean implied (effective) spread $s = s_E$ in Xetra exhibits a U-shape broadly comparable to the results reported for s by MRR. However, the drop after the first half hour of trading is more pronounced in our application. While s in MRR drops only by 10%, the decrease amounts to 31% in our application. The drop is largest for groups three and four containing the less frequently traded stocks. For group four, the implied spread drops by 39% after the initial half hour of Xetra trading. On the other hand, the increase of the implied spread during the final half hour of the trading process, is much less pronounced and in relative terms smaller than in the MRR study. MRR report an increase of the mean implied spread by 10% during the final half hour. The corresponding value in our application amounts to 6%. Again the the intra-day pattern is more pronounced for the group of least frequently traded stocks for which the average mean implied spread increases by 13% . Given these results, it may be more appropriate to describe the intra-day pattern of the implied spread in Xetra as a twisted U-shape or a "smirk". This is the main first difference between MRR and our study: A more pronounced intra-day pattern with a sharp decline of the implied spread after the initial half hour of the trading process with an even more distinct pattern for less frequently traded stocks.

We turn now to the explanations behind the intra-day pattern of the implied spread (s). By definition, s is composed of the information asymmetry parameter θ and the transaction cost parameter ϕ which is associated with opportunity costs market making. Table A.2.1 shows that, as in the MRR paper, the information asymmetry parameter θ drops sharply after the first half hour trading interval. The L-shape of the time of day pattern of the adverse selection cost component is broadly comparable to the MRR paper. The drop after the initial half hour of trading is somewhat more pronounced in our application. While MRR report a decrease of the mean θ of 23%, in our application the decrease amounts to 34% averaged across all stocks. The decrease is most pronounced for the group of least frequently traded stocks (38%) and smallest for the group of most frequently traded stocks (26%).

Our application allows comparing the level of the adverse selection costs measured in θ across stock groups that we sorted according their trading intensity. Table A.2.1 shows that, throughout the day, the information asymmetry parameter θ is higher for less frequently traded stocks. For example, the mean θ for the group of most actively traded stocks during the first half hour of the trading day amounts to 0.9 euro cent. The value for the group of least frequently traded stocks is more than double (2 euro cents). This pattern does not change during the trading day. For the last half hour, the adverse selection cost component averaged for the stocks in group one equals .57 euro cents while the value for group four amounts to 1.02 euro cents.

Although the levels of the adverse selection cost components are not directly comparable across the MRR study and ours, the overall adverse selection costs on the Xetra system appear considerably smaller in absolute terms. For example the mean θ across all DAX30 stocks for the first half hour amounts to 1.2 euro cent, while in the MRR study the average for their sample of 274 stocks amounts to 4 dollar cent. Of course, due to exchange rate effects, stock prices and the broader sample used by MRR these figures are not directly comparable, but the difference seems considerable anyway. These differences remain stable throughout the trading day.

In our application the serial autocorrelation of the trading indicator ρ is smaller than in MRR. Across the groups of stocks, and throughout the trading day, the correlation varies between 0.19 and 0.25, while for the MRR stocks the average correlations varies between 0.41 and 0.37.

While the time of day patterns regarding the adverse selection cost parameters are broadly comparable between the MRR study and ours, the differences concerning the non-informational related transaction cost component ϕ are more striking. MRR report an increase of ϕ during the day, with an increase of about 34% from the first half hour of the trading day to the final half hour. The jump of transaction costs reported by MRR is sharpest after the first half hour of the trading process (17% increase). As discussed above, ϕ can be interpreted as the (opportunity) costs of market making aside from adverse selection effects, a similar interpretation as the realized spread. In our application the intra day pattern of the costs of market making (here maybe better: opportunity costs of supplying liquidity) exhibit a symmetric U-shape which is most pronounced for the group of least frequently traded stocks. We start with a high level of transaction costs during the first half hour of the trading day and a subsequent 25% decrease (averaged across all stocks, 41% for the group of least frequently traded stocks) onto a level that remains roughly constant over the trading day, until the costs increase again during the final half hour (averaged across stocks by 23%, for the group of least frequently traded stocks by 43%).

The top right panel of figure A.2.1 depicts the pattern of the standardized $\tilde{\phi}$. The pattern is quite different in the MRR application. The Xetra pattern seems much more interesting and more pronounced. One could argue that the voluntary liquidity suppliers on Xetra demand a higher compensation for their market making efforts during the initial trading phase. The explanation for this compensation, though, cannot be found in adverse selection effects (these are captured by θ). The increase of the costs of liquidity supply during the final half hour could well be interpreted as an inventory cost effect. Liquidity suppliers (the limit order traders) demand a compensation for the risk (not adverse selection risk, but non-trading opportunity-risk) to carry the inventory overnight.

The possibly most important issue is the comparability of the absolute euro cent figures and the aggregation and averaging of parameter estimates and spreads and price impacts within the groups. Up to now, we followed MRR who also simply averaged the parameters over their 274 stocks. Although the differences in Xetra stock prices are less than they have been some years ago (SAP has the highest per share price of 131.5 Euro, Infineon with 11.6 is the smallest.) we could contemplate controlling for the price level. One possibility would be to divide θ and ϕ (both are in euro cent) by the average price of the respective stock before averaging within the group. In order to control for the price effect we have computed standardized MRR spread components $\tilde{\phi} = \phi/\bar{P}$, $\tilde{\theta} = \theta/\bar{P}$ and a standardized implied spread $\tilde{s}_E = 2(\phi + \theta)/\bar{P}$, where $\bar{P}$ is the average stock price (sample average). As can be seen in table A.2.3 the intra-daily pattern does not change for the standardized parameters.

We still observe an L-shape for the adverse selection component, a U-shape for the order processing cost component and a twisted U-shape for the implied spread (the rise of the implied spread at the end of the trading day is more pronounced for stocks in group 3 and 4). However, the advantage of standardizing can be seen by comparing the levels of the different groups. While for the unstandardized coefficients the second highest level of the implied spread and the adverse selection component occurs for group 1, the largest and most frequently traded stocks, standardized coefficients are ordered as we would expect. Group 3 and 4 exhibit a substantially higher level for the adverse selection component compared to group 1 and 2. The same is true for the implied spread. The differences in the ordering are due to a price effect in the unstandardized coefficients. Larger stocks are more likely to have a higher price which translates into a higher level of the spread and, hence, into a higher level

Table A.2.3: **Standardized adverse selection component, standardized order processing cost component and standardized implied spread.** The table presents the mean coefficient estimate standardized with the price level across stocks (calculated from the first stage GMM estimates) in the respective trade size quartile and across all thirty stocks, as well as standard deviation and median of the estimates within the respective group and across all thirty stocks.

Trade Size Quartile		9:00-9:30	9:30-11:00	11:00-14:00	14:00-15:30	15:30-17:00	17:00-17:30
		$\tilde{\theta}$ (%)					
1	Mean	0.014	0.011	0.009	0.010	0.010	0.009
	Median	0.015	0.011	0.009	0.010	0.010	0.009
	Std	0.003	0.002	0.002	0.002	0.002	0.001
2	Mean	0.021	0.015	0.013	0.014	0.014	0.012
	Median	0.022	0.014	0.012	0.013	0.014	0.012
	Std	0.004	0.002	0.002	0.002	0.002	0.002
3	Mean	0.034	0.023	0.020	0.020	0.020	0.021
	Median	0.034	0.022	0.019	0.019	0.020	0.021
	Std	0.005	0.004	0.004	0.004	0.005	0.004
4	Mean	0.040	0.025	0.021	0.021	0.021	0.021
	Median	0.042	0.027	0.022	0.023	0.022	0.020
	Std	0.004	0.004	0.003	0.003	0.003	0.003
all	Mean	0.027	0.018	0.016	0.016	0.016	0.016
	Median	0.028	0.018	0.015	0.016	0.016	0.016
	Std	0.011	0.007	0.006	0.005	0.005	0.006
		$\tilde{\phi}$ (%)					
1	Mean	0.012	0.011	0.011	0.010	0.010	0.011
	Median	0.012	0.010	0.009	0.008	0.007	0.009
	Std	0.005	0.006	0.006	0.006	0.005	0.005
2	Mean	0.016	0.014	0.014	0.013	0.012	0.015
	Median	0.012	0.011	0.010	0.010	0.009	0.012
	Std	0.008	0.007	0.007	0.007	0.007	0.007
3	Mean	0.024	0.019	0.017	0.016	0.016	0.018
	Median	0.025	0.021	0.020	0.019	0.017	0.020
	Std	0.005	0.006	0.005	0.005	0.005	0.005
4	Mean	0.019	0.011	0.009	0.009	0.010	0.014
	Median	0.017	0.010	0.009	0.010	0.010	0.014
	Std	0.006	0.003	0.002	0.002	0.003	0.004
all	Mean	0.018	0.014	0.013	0.012	0.012	0.015
	Median	0.015	0.011	0.010	0.010	0.010	0.012
	Std	0.007	0.006	0.006	0.006	0.006	0.006
		$\tilde{s}_E$ (%)					
1	Mean	0.054	0.044	0.040	0.040	0.040	0.040
	Median	0.055	0.042	0.037	0.036	0.037	0.039
	Std	0.008	0.009	0.010	0.009	0.009	0.009
2	Mean	0.074	0.057	0.053	0.053	0.052	0.054
	Median	0.070	0.052	0.045	0.046	0.046	0.051
	Std	0.018	0.016	0.016	0.016	0.016	0.017
3	Mean	0.116	0.084	0.073	0.072	0.072	0.077
	Median	0.115	0.083	0.074	0.072	0.072	0.077
	Std	0.015	0.015	0.014	0.013	0.014	0.013
4	Mean	0.117	0.072	0.060	0.061	0.061	0.070
	Median	0.109	0.071	0.061	0.059	0.060	0.069
	Std	0.017	0.010	0.007	0.005	0.008	0.012
all	Mean	0.091	0.065	0.057	0.057	0.057	0.061
	Median	0.095	0.064	0.057	0.058	0.057	0.061
	Std	0.031	0.020	0.017	0.016	0.017	0.019

of the spread components.

The level of transaction costs for the Xetra traded stocks is considerably smaller than in the MRR study. Even for group four, the group of least frequently traded stocks, our ϕ estimates are much smaller than the estimates reported in table 2 in Madhavan et al (1997). For example, the average transaction cost estimate for group four during the first half hour amounts to 0.9 euro cent (all stock average: 0.67 euro cent) while the corresponding average ϕ estimate in MRR is equal to 3.4 dollar cent. At least for the non-informational related transaction costs containing a compensation for taking unwanted inventory, we would expect smaller costs on Xetra since liquidity provision is voluntary. Therefore, limit order submitters providing liquidity should demand a smaller compensation for inventory risk.

Generally, the relative importance of the adverse selection cost component for the spread is higher for the Xetra traded stocks compared to the MRR stocks. Averaged across stocks, the ratio $r = \frac{\theta}{\theta+\phi}$ ranges from 0.6 (first half hour) to 0.52 (final half hour). Differences across stock groups are quite pronounced. For the group of least frequently traded stocks, the information component accounts for about 70% of the spread throughout the day until the ratio drops to 0.6 during the final half hour. The information component of the spread is considerably smaller and the intra-day pattern of r is more L-shaped for the more frequently traded stocks. For example, the relative importance of the adverse selection component is about 54% during the first half hour and drops to 46% during the final trading hour for stock group one containing the most frequently traded stocks.

However, the information component share ratios are considerably larger than in MRR's application. They report average ratios of about 0.38 (excluding the first half hour where the ratio is 0.51). Although the level of the adverse selection costs in Xetra seems to be smaller, the share of the spread attributable to the information component is higher. We can offer two explanations for this result. First, one could argue that the the anonymity of the trading system aggravates adverse selection effects. However, this concerns only the relative importance of adverse selection effects and not the level. Second, the smaller non-informational related transaction costs in the electronic trading system may render adverse selection costs relatively more important.

It seems also worth discussing further the intra-day pattern of the share of the information component as we have some striking differences compared to the MRR application. MRR report an L-shape of the intra-day pattern of $\frac{\theta}{\theta+\phi}$. Given the discussion it is not surprising that our pattern is somewhat different, but well explainable (see lower right picture in figure A.2.1). Although we also observe, averaged across all stocks, the initial drop of the information component share, but we also observe a further sharp decrease during the final half hour. This decrease can be explained by the time-of-day pattern of the non-informational related transaction costs which increase (allegedly due to inventory effects) during the final half Xetra hour, thus reducing the relative importance of the information component.

Another extension underpinning our argument in section 2.5 is the estimation of the standard MRR model for different trade categories. As in figure 2.5.1, trades are categorized according to how far they "walk up the book". In category 1, we include all trades with a volume smaller than the best depth, category 2 includes those trades consuming exactly the best depth, in category 3 we collect all trades larger than the best volume but within any higher order quote and finally, trades in category 4 consume exactly the volume behind the second, third or fourth quote. Very similar to Huang and Stoll (1997) who define different volume categories, we introduce a dummy variable for each trade category. The MRR model given in equation 2.8 can then be written as:

$$\Delta P_i = \sum_{k=1}^{4} D_k \theta_k (Q_i - \rho Q_{i-1}) + \phi(Q_i - Q_{i-1}) + \varepsilon_i + \xi_i - \xi_{i-1}. \tag{A.18}$$

where

$$D_k = \begin{cases} 1 & \text{if} \quad \text{trade of category k} \\ 0 & else \end{cases}$$

The estimation results of model A.18 are reported in table A.2.4. Not surprisingly, large volume trades of category 3 and 4 have a larger adverse selection component than smaller trades of category 1 and 2. An interesting result is that trades consuming exactly the best depth (category 2) have a lower information component than trades within the best depth. This result holds for the overall sample and for each trade activity quartile and even for every single stock.

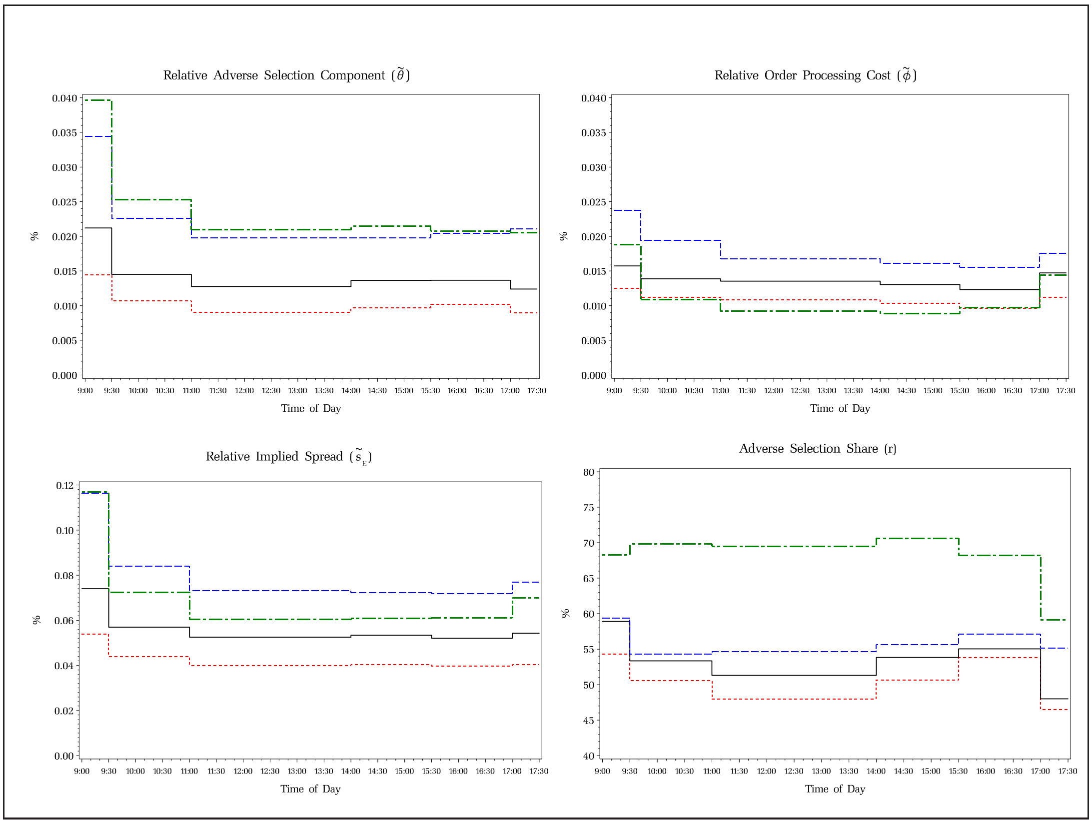

Figure A.2.1: **Intra-daily Patterns for the estimated standardized spread components.** The dotted line depicts trade size quartile 1, the straight line depicts trade size quartile 2, the dashed line depicts trade size quartile 3 and the dash-dotted line depicts trade size quartile 4. Top left: Intra-daily pattern for the standardized adverse selection component for each trade size quartile. Top right: Intra-daily pattern for the standardized order processing cost component for each trade size quartile. Lower left: Intra-daily pattern for the standardized implied spread for each trade size quartile. Lower right: Intra-daily pattern for the adverse selection share of the implied spread for each trade size quartile.

Table A.2.4: **Estimation results for MRR taking into account different trade types.** The table reports the estimation results for the standard MRR model including dummy variables for different trade types. θ_1 denotes the average price impact for trades within the best quote. θ_2 denotes the average price impact for trades hitting exactly the best quote. θ_3 denotes the average price impact for trades hitting exactly the second, third or fourth quote and θ_4 denotes the average price impact for trades beyond the best quote but within any other quote. Average p-values for the parameter estimates are in parentheses.

	Overall			1^{st} Quartile (most active)			2^{nd} Quartile			3^{rd} Quartile			4^{th} Quartile (least active)		
	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]	Avg. est.	Avg. p-val	# sig [pos, neg]
θ_1	0.0055	(0.00)	[30, 0]	0.0047	(0.00)	[7, 0]	0.0041	(0.00)	[8, 0]	0.0043	(0.00)	[8, 0]	0.0092	(0.00)	[7, 0]
θ_2	0.0030	(0.00)	[28, 2]	0.0021	(0.00)	[6, 1]	0.0022	(0.00)	[7, 1]	0.0022	(0.00)	[8, 0]	0.0056	(0.00)	[7, 0]
θ_3	0.0195	(0.00)	[30, 0]	0.0178	(0.00)	[7, 0]	0.0140	(0.00)	[8, 0]	0.0160	(0.00)	[8, 0]	0.0316	(0.00)	[7, 0]
θ_4	0.0212	(0.00)	[30, 0]	0.0194	(0.00)	[7, 0]	0.0168	(0.00)	[8, 0]	0.0174	(0.00)	[8, 0]	0.0324	(0.00)	[7, 0]
ϕ	0.0062	(0.00)	[30, 0]	0.0071	(0.00)	[7, 0]	0.0056	(0.00)	[8, 0]	0.0051	(0.00)	[8, 0]	0.0073	(0.00)	[7, 0]
ρ	0.2204	(0.00)	[30, 0]	0.2203	(0.00)	[7, 0]	0.2067	(0.00)	[8, 0]	0.2113	(0.00)	[8, 0]	0.2465	(0.00)	[7, 0]

In figure 2.5.1 we have seen that trades of category 2 tend to occur when trading activity is higher (durations are shorter). Together with the results of model A.18 this finding supports our previous hypothesis that non-informed liquidity traders herd together when trading costs are low. Due to the increase in trading activity they increase their order aggressiveness, for example by using market-to-limit orders or marketable limit orders. If those orders are only partially filled the remaining order size is entered on the other side of the market as a limit order with the respective limit price. Further, we have seen in an example in section 2.5 that this order type can yield a different assessment of the information content depending on the model specification. Specifically, the chosen variable to analyze the informational content (Dufour/Engle: instantaneous midquote revision, MRR: bid-ask spread) plays a major role. If this order type constitutes a considerable share of all transactions (see figure A.2.2), different results concerning the role of time for the informational content of a trade are not surprising.

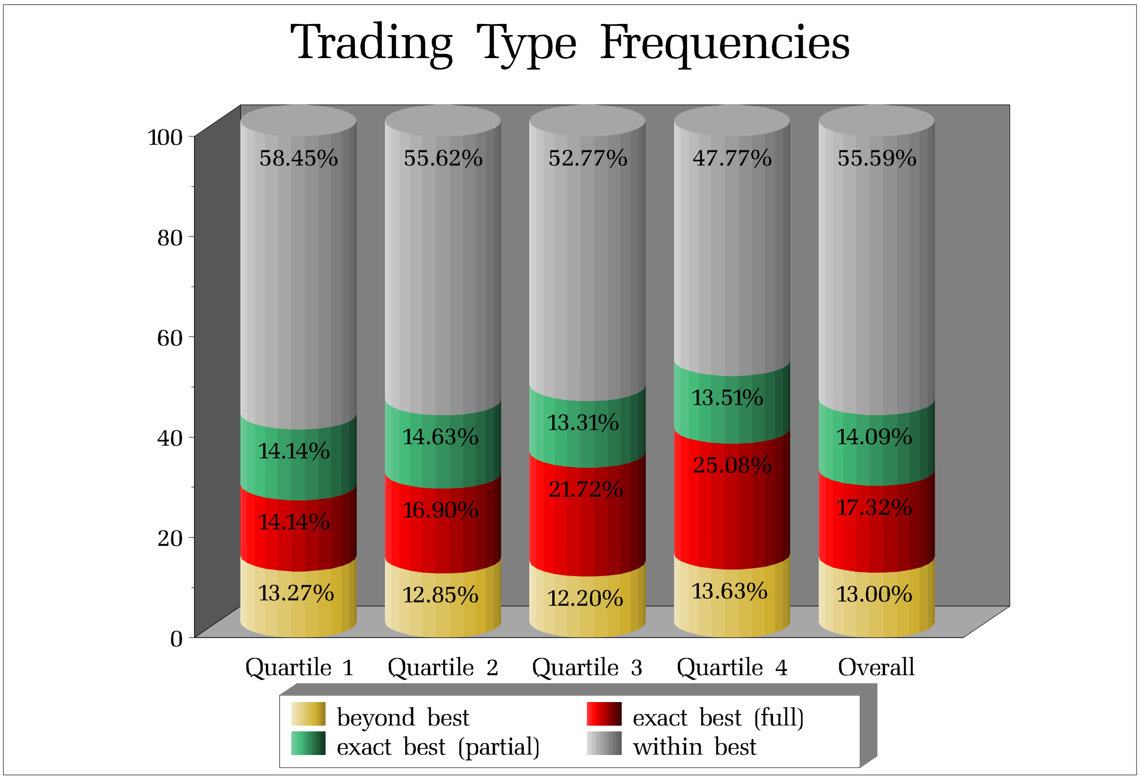

Figure A.2.2: **Frequencies of different order types.** The graph depicts the frequencies of transactions for four different trade categories. We denote with 'within best' the trades with a volume smaller than the best depth. With 'exact best (partially)' we denote the trades consuming exactly the best depth but are only partially filled. Fully filled orders consuming exactly the best depth are categorized as 'exact best (full)' and trades with a volume higher than the best depth 'beyond best'.

Chapter 3

Modeling Time Varying Arrival Rate Dynamics of Informed and Uninformed Traders

The dynamic EKOP model of Easley, Engle, O'Hara, and Wu (2002) which allows for time varying trading intensities can be used to reveal some insights about the strategic behavior of informed as well as uninformed traders. While my results for German intra-day trade data confirm that informed traders try to hide their status by trading more heavily when uninformed traders are present, the behavior of uninformed traders seems to be dependent on the trading frequency of the stock. For frequently traded stocks uninformed traders try to avoid informed traders while for infrequently traded stocks uninformed traders herd when informed traders are present. Generally, the reaction of uninformed traders is weaker. Furthermore, I perform a cross-sectional analysis in order to shed light on the relation between the probability of informed trading and various microstructure variables.

Chapter is based on the article Modeling Time Varying Arrival Rate Dynamics of Informed and Uninformed Traders by O. Wuensche (2005)

3.1 Introduction

In recent years the availability of high-frequency data has increased dramatically. As a consequence, empirical tests of hypotheses in the field of financial market microstructure play a vital role in modern economic research. The idea is that every single transaction contains valuable information. One strain of the empirical literature copes with the question *when* a trade takes place whilst another strain asks *how many* trades arrive in a certain interval of time. Although those two questions are closely related, modeling the trade process differs because of the different data structure. The latter question is important when dealing with a sequential trade model belonging to the class of EKOP models first introduced by Easley, Kiefer, O'Hara, and Paperman (1996). In those models information about the amount of informed and uninformed trading can be inferred from the daily number of buys and sells. This information can be combined in a single measure, the probability of informed trading (PIN). Since then numerous studies have focussed on calculating the PIN to answer questions concerning market microstructure. For example, Grammig, Schiereck, and Theissen (2001) use the PIN measure to test if informed traders prefer to trade on markets providing anonymity and Odders-White and Ready (2004) explain some portion of a firms credit rating with the firms PIN.

The vast majority of the literature (e.g. Easley, Kiefer, O'Hara, and Paperman (1996), Easley, Engle, O'Hara, and Wu (2002)) has used daily numbers of buys and sells to estimate the EKOP model and calculate the PIN. Further, in most applications the estimated PIN and the estimated arrival rates of informed and uninformed traders have no time dimension and are calculated as constant numbers. Although the cross-sectional difference in size might be interesting in various respects it is also important not to neglect the time varying structure. One exception is the work of Easley, Engle, O'Hara, and Wu (2002) who allow for time varying arrival rates but they also use only daily aggregated numbers of buys and sells. Hence, a potentially existing intra-day pattern of the arrival rates and/or the PIN cannot be revealed. Moreover, the arrival rates are closely related to the number of trade events and the order imbalance in a certain interval. But order imbalances as well as the overall number of transactions have been shown to be autocorrelated (see Chordia, Roll, and Subrahmanyam (2002) or Chordia, Roll, and Subrahmanyam (2005)). Thus, models neglecting this time dependence deliver questionable results. Lei and Wu (2005) have recently proposed a Markov switching EKOP model to capture the dynamics of arrival rates. In a different structural framework, Nyholm (2002) uses a modified trade indicator model based on Huang and Stoll (1997) to estimate an intra-day pattern of the PIN. The following questions could be addressed when having a time varying PIN: Does the opening of the NYSE have an effect on domestic trading? Is the spread and the overall trading activity after the opening so high because of overnight information? Are large or small orders more likely to stem from informed traders?[1]

With additional information of dynamic interaction between informed and uninformed traders I can find answers to the following questions: Do informed traders time their trades to hide behind high liquidity?[2] Do uninformed traders avoid informed traders or is there any evidence for herding? My results indicate that the reaction of uninformed traders to informed trading is generally weaker than the reaction of informed traders to uninformed trading. Additionally, I find evidence that the reaction of uninformed traders to informed trading depends on the trading frequency of the traded stock. In all cases where uninformed trading declines as a reaction to informed trading the stocks were rather frequently traded. Informed traders on the other hand tend to enter the market in times of high liquidity and follow the uninformed to cover their informational advantage. I will test if some important market microstructure theories hold for XETRA trade data. First, it can be confirmed by comparing the average PIN for a cross-section of stocks that smaller stocks are more exposed to informed trading than larger stocks (e.g. Easley, Kiefer, O'Hara, and Paperman (1996)). Further, I find that the PIN decreases with increasing average volume per trade, thus, there is evidence that informed investors tend to split their trades. On the other hand, the PIN is positively related to overall trading activity measured as the total volume traded. Moreover, my results indicate that the beginning of US trading has a significant positive effect on the PIN of most of the stocks traded on XETRA. I also observe that informed traders hide behind liquidity since the PIN is lower when liquidity is low.

The remainder of this chapter is organized as follows. In section 3.2, I will provide a brief review of the basic EKOP model followed by a critical assessment presenting some caveats of the basic model. Then, an extended model is presented taking into account the autocorrelation and the cross-correlation of arrival rates of informed and uninformed market

[1] For example, Brown, Thomson, and Walsh (1999) find that informed traders are more likely to choose smaller orders than uninformed traders.

[2] Kyle (1985) and Admati and Pfleiderer (1988) brought forward the argument that informed traders make use of the camouflage noise traders provide.

participants. Section 3.4 introduces an application of this model to the DAX30 stocks of the fully electronic trading system XETRA. First, I present the arrival rate dynamics and second, I perform a cross-sectional study relating the PIN to various microstructure variables in order to address some of the questions formulated above. Finally, section 3.6 provides a summary of the results and concludes.

3.2 Modeling Trade Arrival Rates

3.2.1 Reviewing the EKOP Model

The traditional approach of modeling the arrival rates of informed and uninformed traders is based on a sequential trading model, the EKOP model, first introduced by Easley, Kiefer, O'Hara, and Paperman (1996). They assume that traders, trading with a market maker, are split into two subgroups, informed traders and uninformed traders. The market maker does not know the type of the trader but she can infer some information from the trading process. At the beginning of the day an information event occurs with the probability α. Given an information event has occurred, δ denotes the probability that it was a *bad news* event and $1-\delta$ the probability that it was a *good news* event. Only the informed traders know if there is good, bad or no information, hence, they are not expected to buy on a bad news day or sell on a good news day. In contrast, uninformed traders do not have such superior information. Therefore, one would expect them to trade in every information state. The resulting buy and sell trading intensities for this simple trading process, λ_i, for each possible state are as follows. Denoting the trading intensity of the informed traders with μ and for uninformed traders with ε one gets the trading intensities λ_i as depicted in figure 3.2.1. In the most basic model it is assumed that the trading intensity of uninformed traders as well as informed traders is the same for buys and sells. Furthermore, it is assumed that the trading intensity of both groups is equal across all possible types of information events.[3] This assumption can easily be relaxed by introducing different arrival rates for buyer and seller without loss of generality.

The number of buys and sells per pre-specified time interval is sufficient to estimate the parameters of the model via straightforward Maximum Likelihood techniques. Assuming

[3]It has already been mentioned that informed traders only trade in two states. They buy when there are good news and they sell when there are bad news.

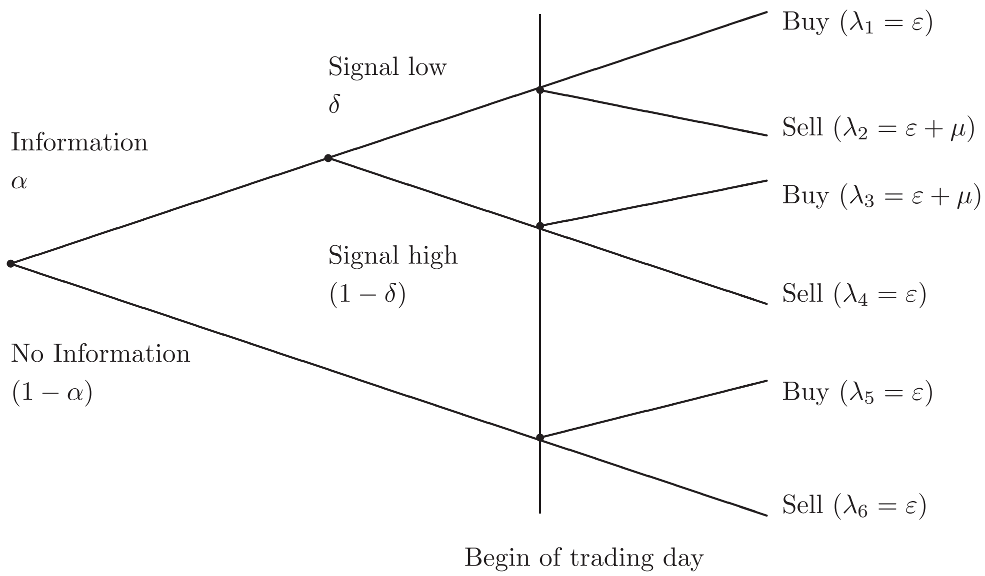

Figure 3.2.1: Tree representation of the EKOP model.

the number of buys and sells to be independently Poisson distributed, the likelihood contribution of interval t can be constructed as follows. It consists of three products of two independent Poisson distributions where each product represents one possible state of nature. Consequently, the intensity parameter of buys and sells is different for each product. Those products are weighted with the respective probability for each information state. The probability of observing a specific sequence $\{B_t, S_t\}_{t=1}^{T}$ of buys and sells is simply the product of all individual likelihood contributions, i.e.

$$\begin{aligned}\mathfrak{L}_{EKOP}(\boldsymbol{\theta}) &= \prod_{t=1}^{T} \alpha\delta e^{-(2\varepsilon+\mu)} \frac{\varepsilon^{B_t}(\varepsilon+\mu)^{S_t}}{B_t!S_t!} + \alpha(1-\delta)e^{-(2\varepsilon+\mu)} \frac{\varepsilon^{S_t}(\varepsilon+\mu)^{B_t}}{B_t!S_t!} \\ &\quad +(1-\alpha)e^{-2\varepsilon} \frac{\varepsilon^{B_t}\varepsilon^{S_t}}{B_t!S_t!}, \qquad (3.1)\end{aligned}$$

where $\boldsymbol{\theta}$ denotes the parameter vector and T is the number of time intervals (e.g. days) in the sample. However, in empirical applications it is often more convenient to use a slightly modified version of equation (3.1). For example, Venter and de Jongh (2004) suggest the following log-likelihood:

$$\begin{aligned}\mathfrak{l}_{EKOP}(\boldsymbol{\theta}) &= \sum_{t=1}^{T} -2\varepsilon + (B+S)\ln\varepsilon - \ln B! - \ln S! \\ &\quad + \ln\{(1-\alpha) + \alpha\delta\exp[S\ln(1+\mu/\varepsilon) - \mu] \qquad (3.2) \\ &\quad + \alpha(1-\delta)\exp[B\ln(1+\mu/\varepsilon) - \mu]\} \qquad (3.3)\end{aligned}$$

The advantage of using equation (3.2) instead of just taking the log of equation (3.1) is that numerical problems, occurring frequently from large numbers of buys and sells, can be circumvented.

Note, that all terms involving the factorial have been dropped. This is possible because the exclusion of the factorials does not alter the the location of the parameters maximizing the likelihood function. Instead, it only shifts the value of the likelihood function for any value of $\boldsymbol{\theta}$.[4]

3.2.2 Critical Assessment

Unfortunately, the EKOP model in its simplest form has various drawbacks. Probably the most important problem is that it does not fit empirical data very well. This is partly due to a) the strict assumptions concerning the independence of μ and ϵ, meaning that there is no interaction between the arrival of informed and uninformed traders and b) the time invariance of μ and ϵ.

The first assumption seems to be quite unrealistic since from finance theory we know that uninformed traders try to avoid informed traders because they "lose" when they trade against superior informed agents (see Foster and Viswanathan (1990)). On the other hand, one could also imagine some kind of herding behavior (see Froot, Scharfstein, and Stein (1992)) which would imply more uninformed trading when informed trading is high. Imagine a case where informed traders received a bad news signal and sell extensively. As a consequence, the stock price drops. Now, either the noise traders conclude that there is bad information and sell themselves or they conclude that the lower price is a speculative buying opportunity. Even if it is not clear which of the above hypotheses is true, there is theoretical evidence for

[4] The likelihood value for any choice of $\boldsymbol{\theta}$ resulting from the modified likelihood in equation 3.2 cannot be directly compared with the likelihood value resulting from the log of equation 3.1. A derivation of equation 3.2 is given in Appendix B.1.

interdependence between the trading intensities of the two trader groups.

The second assumption of time invariant arrival rates is also problematic with regard to observed data. The number of trades, and hence, the number of buys and sells in a certain time interval, is not independent of the number of trades in the previous interval. For example, Engle and Russell (1998) introduce the autoregressive conditional duration (ACD) model to capture the autoregressive structure of the trade process. If the time between trades exhibits strong autocorrelation, the number of trades in a certain period will also not be independent from past observations. In the next section I will show that the order imbalance drives the arrival rate of the informed traders whilst the number of balanced trades drives the arrival rate of the uninformed traders. As was already mentioned above, order imbalances have been shown to be autocorrelated. From our earlier argumentation it follows that balanced trades are also expected to be autocorrelated.

Table 3.2.1 reports the Ljung-Box statistic for autocorrelations of order imbalances and balanced trades for the DAX30 stocks. The large values of the Ljung-Box statistics confirm the presence of autocorrelation also for the German stock exchange. Additionally, in the fourth column I find evidence for the interdependence between the order imbalance and the balanced trades. Hence, if the arrival rates are driven by those variables it might be useful to allow for interdependence when estimating the arrival rates. In the following section, I will present a model capable of allowing for time dependence and interacting behavior of informed and uninformed traders.

3.3 The EKOP Model with Time Varying Trading Intensities

In the previous section I presented some main caveats of the basic EKOP model. In order to relax the rather strict assumptions Easley, Engle, O'Hara, and Wu (2002) have developed an extended model which allows for interaction between the arrival rates and takes into account the observed autocorrelation in the sequence of buys and sells.

Let us first reconsider some implications from section 3.2.1. I can obtain the expected number of total trades (hereafter TT) from

$$E(TT) = \alpha\delta(\mu + 2\varepsilon) + \alpha(1-\delta)(\mu + 2\varepsilon) + (1-\alpha)(2\varepsilon) = \alpha\mu + 2\varepsilon \tag{3.4}$$

Table 3.2.1: **Ljung-Box statistics for the order imbalance and the number of balanced trades.** The second column presents the Ljung-Box statistic for the order imbalance ($|OI|$) of the individual stocks as listed in column 1. The third column contains the Ljung-Box statistic for the number of balanced trades (BT). $\rho(OI, BT)$ is the Pearson correlation coefficient for the correlation between order imbalance and the number of balanced trades. The last column classifies the stocks according to their trade activity into quartiles.

Ticker	$LB(OI)$	$LB(BT)$	$\rho(OI, BT)$	Trade activity quartile
ALV	1685.22	24983.43	0.21	1
DBK	2082.59	25165.88	0.20	
SIE	1247.51	21261.75	0.20	
DCX	935.56	14942.15	0.20	
MUV2	1059.46	17656.79	0.18	
EOA	990.20	15189.23	0.16	
DTE	1383.02	23718.59	0.13	
SAP	1166.62	13927.71	0.21	2
RWE	1249.97	16790.25	0.19	
IFX	6140.95	21223.48	0.17	
HVM	2404.28	20000.01	0.17	
BAS	1216.64	10863.51	0.16	
VOW	1052.12	9900.30	0.16	
BAY	1768.04	18902.90	0.16	
BMW	903.45	11441.34	0.08	
MEO	777.39	7658.01	0.13	3
SCH	1531.71	14045.07	0.13	
TUI	2923.62	25895.53	0.11	
ALT	1393.03	12006.63	0.10	
LHA	1357.08	15910.20	0.09	
CBK	1175.05	10484.18	0.07	
TKA	807.49	9905.73	0.07	
DPW	947.34	9886.71	0.03	
LIN	876.65	5314.55	0.06	4
MAN	828.35	6630.81	0.06	
ADS	757.71	8052.39	0.06	
DB1	1060.65	6630.06	0.05	
CONT	1314.62	8777.95	0.05	
HEN3	661.68	3635.70	0.04	
FME	478.69	3959.90	0.01	

Further, I define the order imbalance $|OI|$ as $|S - B|$. Since the observed order imbalance stems from one of the three possible states of nature, one can write:

$$E|OI| = \alpha\delta E|OI_{bn}| + \alpha(1-\delta)E|OI_{gn}| + (1-\alpha)E|OI_{nn}|.^5 \tag{3.5}$$

Since OI is the difference between two Poisson variables, the analytical computation of the expected value $E|OI|$ is a rather non-trivial task. Katti (1960) has derived analytical

[5] The index bn stands for bad news, gn for good news and nn for no news.

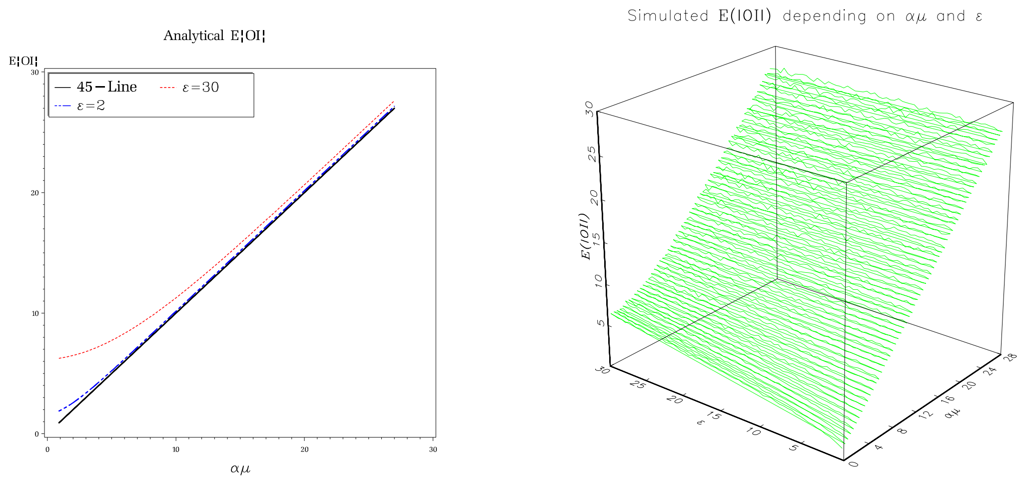

Figure 3.3.1: **Analytical vs. simulated expected order imbalance.** The left panel shows the analytical expected value of $|OI|$ depending on $\alpha\mu$ and different values for ε. The right panel shows the simulated expected value of $|OI|$ depending on $\alpha\mu$ and ε. In both graphs the underlying numerical values are $\alpha = 0.9$ and $\delta = 0.5$.

expressions for the moments of the absolute value of the difference between, inter alia, two Poisson variables. I forbear from presenting those expressions and will instead show some simulation results and compare them with the analytical results in figure 3.3.1.

Both graphs show that at least for large values of μ the approximate relation

$$E|OI| \doteq \alpha\mu \tag{3.6}$$

holds. Further, it even holds for small values of μ if ε is small. Looking at previous empirical studies, a combination of a small μ and a large ε is not observed, so, overall this is a quite good approximation. Using this result together with equation (3.4) it is easy to see that

$$E(TT - |OI|) = E(BT) \doteq 2\varepsilon \tag{3.7}$$

where BT are the balanced trades. Now I have established two observable variables, the order imbalance and the balanced trades, as the driving force for the process of the two unobservable arrival rates μ and ε. The next step is using this relation to model the arrival rates as a bivariate vector process instead of assuming that they are constant. After collecting the two arrival rates in a vector $\boldsymbol{\zeta}_t = (\alpha\mu_t, 2\varepsilon_t)'$, Easley, Engle, O'Hara, and Wu (2002) suggest the following model:

$$\boldsymbol{\zeta}_t = \omega + \sum_{k=1}^{p} \boldsymbol{\Phi}_k \boldsymbol{\zeta}_{t-k} + \sum_{j=0}^{q-1} \boldsymbol{\Gamma}_j \mathbf{Z}_{t-j} \tag{3.8}$$

where $\mathbf{Z}_t = (|OI|_t, BT_t)'$. The expression in (3.8) is very similar to GARCH or ACD type models. As for the latter models arrival rates can only take on values greater than zero which is not guaranteed in the above model specification. In order to take this fact into account a logarithmic version of (3.8) could be used which reads as:

$$\ln \boldsymbol{\zeta}_t = \omega + \sum_{k=1}^{p} \boldsymbol{\Phi}_k \ln \boldsymbol{\zeta}_{t-k} + \sum_{j=0}^{q-1} \boldsymbol{\Gamma}_j g(\mathbf{Z_t}) \tag{3.9}$$

where g is any suitable function of $\mathbf{Z_t}$. For example, one could use $\ln \mathbf{Z}_t$ as suggested by Geweke (1986) in his GARCH specification or $\frac{\mathbf{Z}_t}{\boldsymbol{\zeta}_{t-1}}$ as suggested by Bauwens and Giot (2000) in their Log-ACD$_2$ model. I use a slightly modified version of the latter function resembling the EGARCH model of Nelson (1991). Defining $g(\mathbf{Z_{it}}) = \mathbf{M}_{it} = \frac{\mathbf{Z}_{it}}{\boldsymbol{\zeta}_{i(t-1)}} - 1$ and plugging this into (3.9), one can write:

$$\ln \boldsymbol{\zeta}_t = \omega + \sum_{k=1}^{p} \boldsymbol{\Phi}_k \ln \boldsymbol{\zeta}_{t-k} + \sum_{j=0}^{q-1} \boldsymbol{\Gamma}_j \mathbf{M}_t. \tag{3.10}$$

In order to control for overnight dynamics I extend the model in (3.10) by introducing a dummy variable for the first five minute interval of each day. Hence, the estimated model with $p = 1$ and $q = 1$ now looks like:

$$\ln \boldsymbol{\zeta}_t = \omega + (1 - d_t)\boldsymbol{\Phi} \ln \boldsymbol{\zeta}_{t-1} + d_t \tilde{\boldsymbol{\Phi}} \ln \boldsymbol{\zeta}_{t-1} + (1 - d_t)\boldsymbol{\Gamma}\mathbf{M}_t + d_t \tilde{\boldsymbol{\Gamma}}\mathbf{M}_t \tag{3.11}$$

where

$$d_t = \begin{cases} 1 & \text{if} \quad \text{first observation of the day} \\ 0 & \text{else} \end{cases}$$

Estimation of the parameters is accomplished by using the log-likelihood function given in (3.2).

3.4 Time Varying Arrival Rates on XETRA

The data set used for the empirical analysis is extensively described in section 2.2. For this specific model, I counted the buy and sell transactions for each five minute interval on each trading day. From the aggregated numbers of buys and sells for each five minute interval, I calculate the order imbalance $|OI|_t$ and the balanced trades BT_t.

3.4.1 Intra-day Behavior of Different Types of Traders

Figure B.2.1 plots the estimated arrival rates for the uninformed traders ε_t and figure B.2.2 depicts the estimated arrival rates for the informed traders μ_t throughout a trading day. Therefore, I compute averages of both trading intensities for each five minute interval across all trading days. Both types of traders tend to enter the market predominantly at the beginning and at the end of the trading day. Not surprisingly, the estimated arrival rate patterns are closely related to the average number of balanced trades and the average order imbalance. The PIN for each five minute interval can be calculated as

$$PIN_t = \frac{\alpha\mu_t}{\alpha\mu_t + 2\varepsilon_t}. \tag{3.12}$$

As can be seen in figure 3.4.1, the PIN is roughly constant throughout the day. In contrast to Nyholm (2002) I do not observe a higher PIN shortly after the opening.[6] This finding supports the theory that much of the high trading activity at the beginning and at the end of the trading day stems from portfolio-rebalancing activities and does not reflect specific information.[7] Although the order imbalance and hence, the arrival rate of informed traders is higher at the beginning and the end of the trading day, this is compensated by an increase in uninformed trading at the same time. Overall, the similar intra-day behavior of informed and uninformed traders yields a roughly constant probability of informed trading throughout the day. To shed light on the simultaneous appearance of informed and uninformed trading, I have to explore the dependency structure of the arrival rates.

[6] The results of Nyholm (2002) are mainly driven by the typical L-shaped intra-day spread pattern. In our model, the trading activity throughout the day drives the results.

[7] Brock and Kleidon (1992) and Gerety and Mulherin (1992) argue that investors with different risk aversion profiles exchange the exposure to overnight uncertainty prior to the periodical market closure. Symmetrically, investors trade at the opening of the following day in order to reacquire their assets, resulting in volume that is mainly unrelated to unanticipated overnight information.

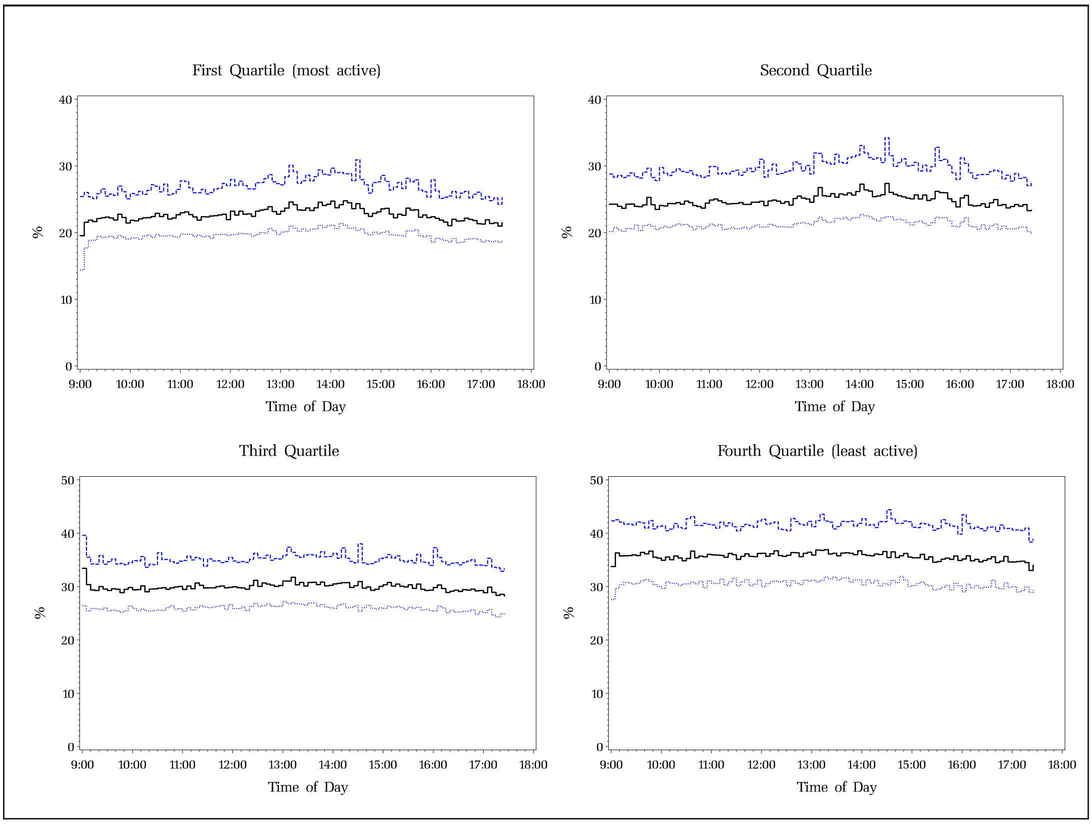

Figure 3.4.1: **Intra-day pattern of the PIN.** The estimated PIN_t values are averaged across all stocks in the particular trade activity quartile for each five minute interval. The probability of informed trading is calculated as $\text{PIN}_t = \frac{\alpha\mu_t}{\alpha\mu_t + 2\varepsilon_t}$. The solid line depicts the median PIN_t. The dotted line depicts the 25% percentile and the dashed line depicts the 75% percentile. The upper left panel depicts the intra-day pattern of the PIN for trade activity quartile 1. The upper right panel depicts the intra-day pattern of PIN for trade activity quartile 2. The lower left panel depicts the intra-day pattern of PIN for trade activity quartile 3. The lower right panel depicts the intra-day pattern of PIN for trade activity quartile 4.

3.4.2 Arrival Rate Dynamics

One main advantage of using the model specification proposed by Easley, Engle, O'Hara, and Wu (2002) is that conclusions about the strategic behavior of informed as well as uninformed traders can easily be drawn. The $\mathbf{\Phi}$ matrix in equation (3.11) contains the dynamics of the arrival rates for informed and uninformed traders. While the diagonal elements capture the autoregressive part, ϕ_{12} measures the reaction of informed traders to the arrival of uninformed traders and ϕ_{21} measures the reaction of uninformed traders to the arrival of informed traders. From table 3.4.1 it can be seen that for all stocks informed traders try to enter the market when noise trading is higher.

Table 3.4.1: **Estimation results of the arrival rate dynamics.** The table reports Maximum Likelihood estimates of the arrival rate dynamics for each of the 30 stocks constituting the DAX30:

$$\ln \boldsymbol{\zeta}_t = \omega + (1-d_t)\boldsymbol{\Phi} \ln \boldsymbol{\zeta}_{t-1} + d_t \tilde{\boldsymbol{\Phi}} \ln \boldsymbol{\zeta}_{t-1} + (1-d_t)\boldsymbol{\Gamma}\mathbf{M}_t + d_t \tilde{\boldsymbol{\Gamma}}\mathbf{M}_t$$

The dummy variable d_t is 1 for the first observation of each day and 0 otherwise. To conserve space we will not report the results for the matrices $\tilde{\Gamma}$ and $\tilde{\Phi}$. P-values based on robust standard errors are reported in parentheses.

Stock	ω_1	ω_2	γ_{11}	γ_{12}	γ_{21}	γ_{22}	ϕ_{11}	ϕ_{12}	ϕ_{21}	ϕ_{22}	α	δ
ALV	6.3401	6.2584	3.8075	3.6642	0.2789	5.7670	0.4734	0.4842	0.0811	0.8109	0.8551	0.3782
	(0.000)	(0.000)	(0.000)	(0.000)	(0.051)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
DBK	7.8530	6.2277	3.6534	4.9539	0.7349	5.6758	0.5444	0.3529	-0.0685	0.9870	0.8908	0.5303
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
DCX	7.5634	3.5844	3.0757	5.1399	0.4333	3.2910	0.5186	0.3565	-0.1118	1.0544	0.7903	0.5631
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
DTE	9.2570	5.2786	4.2339	5.8848	0.8843	4.6809	0.4540	0.4273	-0.0368	0.9934	0.9512	0.4582
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.217)	(0.000)	(0.000)	(0.000)
EOA	6.0174	6.7784	3.8583	2.6981	0.6288	6.0945	0.4987	0.3866	0.0967	0.8001	0.8754	0.4643
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MUV	7.3461	6.6833	3.6246	4.6427	0.8870	5.8023	0.4714	0.4450	0.0195	0.8802	0.9170	0.4179
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.696)	(0.000)	(0.000)	(0.000)
SIE	7.2720	6.1015	3.6200	4.0453	0.6553	5.6042	0.5803	0.2726	-0.0076	0.9381	0.8517	0.4895
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.776)	(0.000)	(0.000)	(0.000)
BAS	5.6433	6.2859	3.5365	2.7949	0.7335	5.2889	0.5411	0.3816	0.1430	0.7344	0.8549	0.5032
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BAY	7.9369	5.7469	3.4382	5.2912	0.7753	5.0169	0.5414	0.3504	-0.0465	0.9482	0.9212	0.4425
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.272)	(0.000)	(0.000)	(0.000)
BMW	6.9218	6.6307	4.0265	3.6465	0.7504	5.3498	0.5331	0.3856	0.0429	0.7528	0.9375	0.4430
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.494)	(0.000)	(0.000)	(0.000)
HVM	6.9135	6.3440	4.1144	3.3617	0.7282	5.6488	0.4044	0.4393	0.0559	0.8474	0.9237	0.4173
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
IFX	8.2982	3.8646	4.3173	4.8028	0.3564	3.6624	0.6281	0.2982	-0.0241	0.9820	0.9300	0.3957
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RWE	5.5575	6.4454	3.6215	2.7697	0.5649	5.6999	0.4588	0.4716	0.1121	0.7678	0.8684	0.5570
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 3.4.1: (continued)

Stock	ω_1	ω_2	γ_{11}	γ_{12}	γ_{21}	γ_{22}	ϕ_{11}	ϕ_{12}	ϕ_{21}	ϕ_{22}	α	δ
SAP	8.9713	4.0554	3.7285	6.0648	1.0532	3.2733	0.4303	0.4260	-0.1112	1.0651	0.9119	0.5117
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
VOW	8.9421	5.0823	3.9494	5.2341	1.2906	3.8535	0.7289	0.0891	-0.1209	1.0169	0.9055	0.4895
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ALT	5.5484	6.4803	3.7976	2.4403	0.8731	5.2985	0.5261	0.4051	0.1523	0.6722	0.9177	0.4791
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CBK	7.1304	5.7543	3.8345	3.4004	0.8127	4.9123	0.5178	0.2618	0.0306	0.8595	0.9429	0.4799
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.443)	(0.000)	(0.000)	(0.000)
DPW	6.3120	6.7590	3.8860	2.8965	0.8276	5.6589	0.5509	0.3233	0.1083	0.7234	0.9491	0.5027
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LHA	6.8650	6.0072	3.9847	3.3345	0.8672	5.0654	0.3553	0.4719	0.0695	0.8091	0.9460	0.5220
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.014)	(0.000)	(0.000)	(0.000)
MEO	5.9689	6.0911	3.7885	2.8443	0.7002	4.8007	0.4531	0.4518	0.1489	0.6340	0.9033	0.5063
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SCH	5.3659	6.3771	3.6200	2.6733	0.6174	5.4168	0.4461	0.5157	0.1588	0.6831	0.9165	0.4750
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TKA	5.0790	6.8978	3.7730	1.6848	0.6735	6.0136	0.5294	0.3396	0.1229	0.7203	0.9569	0.4839
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TUI	5.7392	5.6898	3.7037	2.7829	0.4788	4.8437	0.5097	0.4258	0.0795	0.7395	0.9115	0.4221
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)
ADS	5.8916	6.4861	3.9644	2.1912	0.7065	5.2454	0.5190	0.3185	0.0350	0.7179	0.9610	0.4693
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.060)	(0.000)	(0.000)	(0.000)
CON	5.6035	5.7469	3.9557	2.3996	0.7425	4.7195	0.5699	0.3864	0.0651	0.7626	0.9364	0.5214
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
DB1	5.8716	6.4169	3.9534	1.8640	1.1273	4.8208	0.4919	0.2931	0.0891	0.6792	0.9455	0.5107
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)
FME	5.7157	5.5229	3.6933	1.8230	0.8683	4.2558	0.4370	0.2795	0.1220	0.6600	0.9771	0.4944
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HEN	4.7804	5.7811	3.5394	1.6652	0.6617	4.1764	0.6420	0.2839	0.2030	0.4575	0.9057	0.5043
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LIN	5.8298	5.8094	3.8072	2.2643	0.8347	4.5959	0.5359	0.3085	0.0970	0.6999	0.9308	0.5275
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MAN	5.7710	6.3798	3.8804	2.1857	0.8594	5.0788	0.4765	0.3747	0.1580	0.6326	0.9346	0.5283
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

This supports the aforementioned theory that informed traders make use of the camouflage noise trading provides. In contrast the reaction of noise traders to the arrival of informed traders is ambiguous. While for some of the large stocks noise traders seem to delay their trades when informed agents enter the market (ϕ_{21} is negative), for small stocks a herding effect prevails (ϕ_{21} is positive). This result supports the model of Foster and Viswanathan (1990). Private information about smaller companies is likely to be revealed much slower than information about larger ones. Hence, for small companies, the informed traders will carry on a portion of their information in order not to reveal their type via a too conspicuous order flow. If this is anticipated by the noise traders they will follow the signals they observe to participate in the remaining information. However, for large companies, informed traders try to exploit their advantage immediately. If noise traders enter the market they would lose to the informed and cannot expect to participate in any remaining information. Thus, they try to avoid the informed traders.

3.5 Cross-Sectional Analysis

The second part of my empirical analysis relates the estimated PIN to some common microstructure variables in a pooled regression. First, I choose the market capitalization to take into account that larger stocks are expected to have a lower PIN.[8] Then, I include the sum of the traded volume and the volume per trade.[9] I expect a positive sign for the total volume traded and a negative sign for volume per trade. Further, I include the effective spread and the realized spread. I calculate the effective spread as

$$\text{EFFSPREAD} = \begin{cases} 2 \cdot (TP - MQ_t) & \text{if} \quad \text{buy} \\ 2 \cdot (MQ_t - TP) & \text{if} \quad \text{sell} \end{cases}$$

where TP is the transaction price and MQ_t denotes the prevailing bid-ask midpoint. The index t is the time of the trade in minutes. The realized spread is calculated as

$$\text{REALSPREAD} = \begin{cases} 2 \cdot (TP - MQ_{t+5}) & \text{if} \quad \text{buy} \\ 2 \cdot (MQ_{t+5} - TP) & \text{if} \quad \text{sell} \end{cases}$$

[8] Market capitalization only varies across stocks but not over time.

[9] Some authors use the term volume for the number of traded shares times the transaction price. Here, I refer to volume as the number of traded shares.

Since our estimation results concerning the dynamics of the arrival rates suggest that informed traders try to exploit the camouflage of high liquidity I also include a measure for liquidity as an explanatory variable. Therefore, I use the XETRA liquidity measure which is very similar to the Cost of Round Trip measure developed by Irvine, Benston, and Kandel (2000). For a given buy or sell order size V one can calculate a weighted average price P with the information a limit order book provides. The buyer and seller XLM(V) measures are defined as[10]

$$XLM_{B,t}(V) = 10000\frac{P_{B,t} - MQ_t}{MQ_t} \quad \text{and} \quad XLM_{S,t}(V) = 10000\frac{MQ_t - P_{B,t}}{MQ_t}.$$

To obtain the cost of a roundtrip the two components are simply added up to obtain

$$XLM(V) = XLM_{B,t}(V) + XLM_{S,t}(V).$$

When comparing two XLM measures with the same order size it can be stated that a higher XLM implies less liquidity since the price impact is larger. If there is a camouflage strategy I would expect a negative relationship between the PIN and the XLM measure.

To control for some interesting time of day effects I introduce three dummy variables to the model. The first dummy concerns the period shortly after continuous trading begins (9:00-9:30). Nyholm (2002) finds a higher PIN for this period in his intra-day pattern. But if the trading activity at the open is high predominantly due to portfolio-rebalancing I should expect no effect at all. Furthermore, a dummy for the interval beginning at 2:30 p.m. and a dummy for the interval beginning at 4:00 p.m. are included.[11]

All time-varying variables are averaged across all stocks belonging to the respective trade size quartile for each five minute interval to estimate the following equation.

$$\begin{aligned} PIN_{i,t} &= \beta_0 + \beta_1 MKTCAP_i + \beta_2 DUM230 + \beta_3 VOLSUM_{i,t} \\ &\quad + \beta_4 VOLPERTRADE_{i,t} + \beta_5 XLM_{i,t-1} + \beta_6 REALSPREAD_{i,t} \\ &\quad + \beta_7 EFFSPREAD_{i,t-1} + \beta_8 DUM400 + \beta_9 DUMOPEN \end{aligned} \tag{3.13}$$

The effective spread and the realized spread are expressed in percent of the midquote to

[10] More details about the XLM measure can be found in Gomber, Schweickert, and Theissen (2004).

[11] Note, that index futures trading starts at 2:20 p.m. and the NYSE starts trading at 3:30 p.m.. Gomber, Schweickert, and Theissen (2004) report the same pattern for their XLM measure.

eliminate the level effect of the stock price. Since the variables $VOLSUM$, $VOLPERTRADE$, XLM and $EFFSPREAD$ exhibit a distinct intra-day pattern, they are standardized by subtracting their mean and dividing by their standard deviation for each five minute interval. Estimation results are presented in table 3.5.1. Market capitalization has the expected negative sign in every trade activity quartile confirming the results of Easley, Kiefer, O'Hara, and Paperman (1996). The two dummy variables for the opening of U.S. trading are all positive, although for the less frequently traded stocks in quartile 3 and quartile 4 they are smaller. But overall the beginning of US trading seems to have a positive effect on informed trading. The dummy controlling for the first half hour of the trading day is hardly significant for the smaller stocks but negative for the larger stocks. Heavier trading measured as total volume traded also indicates more informed trading. This is in line with the findings of Engle and Russell (1998). In contrast, there is evidence that informed traders rather submit smaller orders because the parameter for volume per trade is negative in any case. This supports the findings of Brown, Thomson, and Walsh (1999) and the results of a more recent study by Beltran, Grammig, and Menkveld (2005). Furthermore, the sign of the effective spread is somewhat counterintuitive since one would expect that larger spreads are associated with a higher PIN which is only true for the stocks in trade activity quartile 4. However, the effect seems to be rather small anyway. On the other hand, the realized spread always has the expected negative sign. Since the realized spread is the counterpart of the price impact of a trade this result is quite intuitive. The higher the price impact the lower the realized spread and the higher the probability of informed trading. From the size of the parameters I conclude that the realized spread is a better indicator of informed trading than the effective spread. In the above presentation of the XLM measure I argue that a higher XLM measure stands for less liquidity. The negative sign in all four regressions supports the theory that less liquidity is not preferred by informed investors because they are less able to hide their informed status. In general, except for the effective spread, all variables have the expected influence on the probability of informed trading. Even for the less frequently traded stocks the results are robust though some effects like the start of US trading play a minor role in explaining the PIN.

Table 3.5.1: **Estimation results for pooled regressions.** We perform OLS regressions with the pooled series $\text{PIN}_{i,t}$ as dependent variable. The regressions are conducted separately for each trade activity quartile. The effective spread and the realized spread are expressed in percent of the midquote to eliminate the level effect of the stock price. The variables $VOLSUM$, $VOLPERTRADE$, XLM and $EFFSPREAD$ are standardized by subtracting their mean and dividing by their standard deviation for each five minute interval. The test statistics are based on heteroscedasticity consistent standard errors.

Parameter	Coefficient	Std Err	t-value	$Pr > \|t\|$	Parameter	Coefficient	Std Err	t-value	$Pr > \|t\|$
1^{st} Quartile					3^{rd} Quartile				
CONST.	0.2520	0.0016	158.58	< .0001	CONST.	0.3452	0.0015	238.37	< .0001
MKTCAP	-2.63E-7	4.62E-8	-5.70	< .0001	MKTCAP	-5.84E-6	2.59E-7	-22.54	< .0001
DUM230	0.0333	0.0055	6.04	< .0001	DUM230	0.0129	0.0049	2.62	0.0089
VOLSUM	0.0166	0.0009	18.44	< .0001	VOLSUM	0.0066	0.0008	8.64	< .0001
VOLPERTRADE	-0.0107	0.0006	-18.86	< .0001	VOLPERTRADE	-0.0027	0.0005	-5.39	< .0001
XLM	-0.0028	0.0004	-6.55	< .0001	XLM	-0.0023	0.0004	-5.35	< .0001
REALSP	-5.6550	0.5544	-10.20	< .0001	REALSP	-1.8968	0.3676	-5.16	< .0001
EFFSP	-0.0003	0.0005	-0.64	0.5228	EFFSP	-0.0012	0.0004	-2.65	0.0081
DUM400	0.01169	0.0051	2.27	0.0231	DUM400	0.0170	0.0053	3.21	0.0013
DUMOPEN	-0.0104	0.0019	-5.34	< .0001	DUMOPEN	-0.0033	0.0020	-1.68	0.0939
2^{nd} Quartile					4^{th} Quartile				
CONST.	0.2824	0.0013	223.38	< .0001	CONST.	0.3805	0.0023	168.62	< .0001
MKTCAP	-1.24E-6	6.824E-8	-18.19	< .0001	MKTCAP	-3.73E-6	6.349E-7	-5.87	< .0001
DUM230	0.0411	0.0056	7.36	< .0001	DUM230	0.0147	0.0060	2.45	0.0144
VOLSUM	0.0173	0.0009	19.24	< .0001	VOLSUM	0.0065	0.0009	7.28	< .0001
VOLPERTRADE	-0.0110	0.0006	-18.76	< .0001	VOLPERTRADE	-0.0010	0.0006	-1.71	0.0868
XLM	-0.0020	0.0005	-4.13	< .0001	XLM	-0.0023	0.0005	-4.15	< .0001
REALSP	-3.2700	0.6563	-4.98	< .0001	REALSP	-3.7167	0.4208	-8.83	< .0001
EFFSP	-0.0020	0.0005	-4.35	< .0001	EFFSP	0.0017	0.0005	3.10	0.0019
DUM400	0.0193	0.0049	3.91	< .0001	DUM400	0.0104	0.0063	1.65	0.0990
DUMOPEN	-0.0108	0.0022	-4.94	< .0001	DUMOPEN	0.0006	0.0027	0.24	0.8110

3.6 Conclusion

In this chapter I used an extended EKOP model developed by Easley, Engle, O'Hara, and Wu (2002) to estimate a time-varying probability of informed trading. The specification allows us to take into account the autocorrelation and cross-correlation of the arrival rates of informed and uninformed agents on the market. An application of the model for the DAX30 stocks traded on the German fully electronic trading system XETRA reveals evidence for the strategic timing of trades for the informed investors in order to make use of the camouflage provided by noise traders.

Our results indicate that the reaction of uninformed traders to informed trading is generally weaker than the reaction of informed traders to uninformed trading. Additionally, I find evidence that the reaction of uninformed traders to informed trading depends on the trading frequency of the traded stock. Noise traders try to avoid informed traders when trading in large and more frequently traded stocks while they exhibit herding behavior when trading in less frequently traded stocks. When constructing an intra-day pattern for the PIN the picture differs from former studies especially at the beginning of the day. In contrast to Nyholm (2002), I do not find a high and decreasing PIN after the opening but a rather constant PIN throughout the day. In the next step I relate the PIN to various microstructure variables and find that the PIN is positively related to the total volume traded and the opening of US trading. Further, it is negatively related to market capitalization, the volume per trade, the realized spread and the XETRA liquidity measure.

To explore the intra-day arrival rates in greater detail it would be desirable to allow for a time varying probability that an information event occurs. This probability might be roughly constant if daily aggregated numbers are used but certainly not if five minute intervals are used. Letting α depend on time-of-day dummies might be a possible improvement.

Appendix B

B.1 Derivation of Stable Likelihood

The likelihood function of the EKOP model with an independent bivariate Poisson distribution for buys and sells is given by:

$$\begin{aligned}
\mathfrak{L}_{EKOP}(\boldsymbol{\theta}) &= \prod_{t=1}^{T} \alpha\delta e^{-(2\varepsilon+\mu)} \frac{\varepsilon^{B_t}(\varepsilon+\mu)^{S_t}}{B_t!S_t!} + \alpha(1-\delta)e^{-(2\varepsilon+\mu)} \frac{\varepsilon^{S_t}(\varepsilon+\mu)^{B_t}}{B_t!S_t!} \\
&\quad + (1-\alpha)e^{-2\varepsilon} \frac{\varepsilon^{B_t}\varepsilon^{S_t}}{B_t!S_t!},
\end{aligned}$$

where $\boldsymbol{\theta}$ is a vector of parameters and t the index of time intervals.

Factoring out yields:

$$\begin{aligned}
\mathfrak{L}_{EKOP}(\boldsymbol{\theta}) &= \prod_{t=1}^{T} e^{-2\varepsilon} \frac{\varepsilon^{B_t}\varepsilon^{S_t}}{B_t!S_t!} \\
&\quad \times \left[(1-\alpha) + \alpha\delta e^{-\mu}(1+\mu/\varepsilon)^{S} + \alpha(1-\delta)e^{-\mu}(1+\mu/\varepsilon)^{B}\right]
\end{aligned}$$

Applying exp and log to the power terms inside the brackets and taking the logarithm of the whole function, we receive the log-likelihood function (denoting $\ln \mathfrak{L}_{EKOP} := \mathfrak{l}_{EKOP}$):

$$\begin{aligned}
\mathfrak{l}_{EKOP}(\boldsymbol{\theta}) &= \sum_{t=1}^{T} -2\varepsilon + (B+S)\ln\varepsilon - \ln B! - \ln S! \\
&\quad + \ln\left\{(1-\alpha) + \alpha\delta \exp[S\ln(1+\mu/\varepsilon) - \mu] + \alpha(1-\delta)\exp[B\ln(1+\mu/\varepsilon) - \mu]\right\}
\end{aligned}$$

Note, that the terms including the factorials can be dropped if necessary, since they only shift the level of the likelihood value but not the location of the maximum. ■

B.2 Intra-day Pattern of Arrival Rates

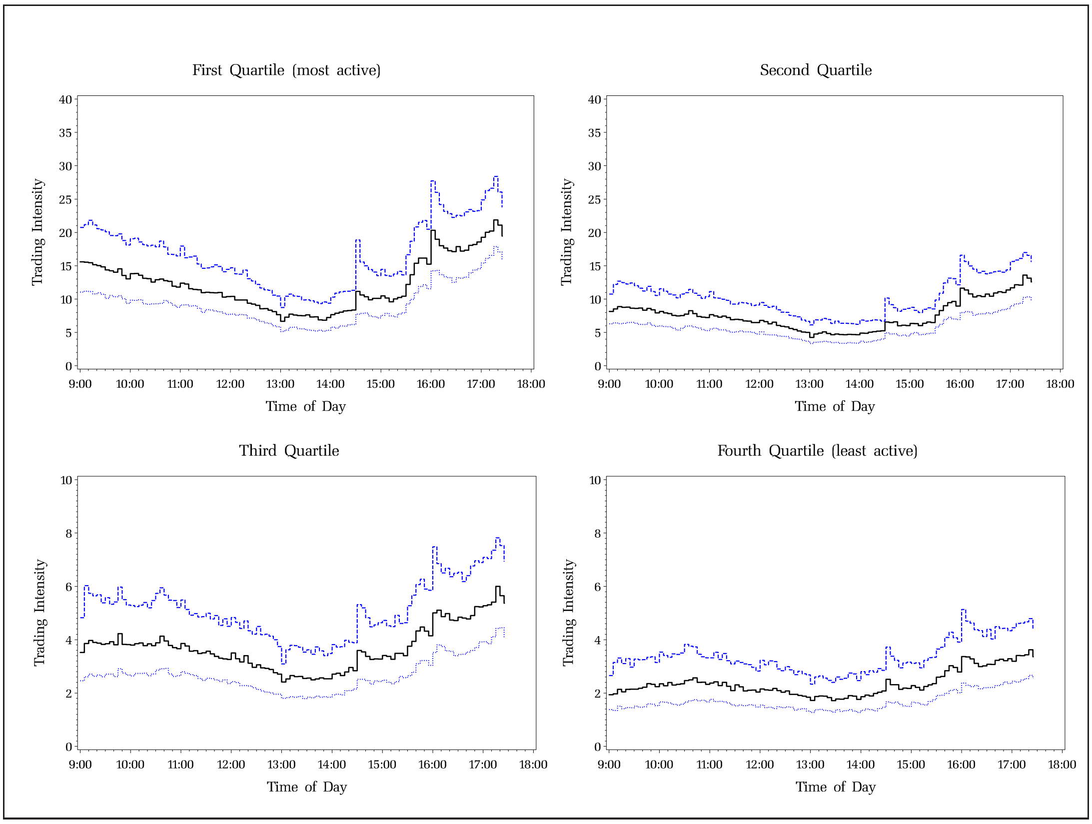

Figure B.2.1: **Intra-day pattern for the arrival rate of the uninformed traders.** The arrival rates ε_t are averaged across all stocks in the particular trade size quartile for each five minute interval. The solid line depicts the median ε_t. The dotted line depicts the 25% percentile and the dashed line depicts the 75% percentile. The upper left panel depicts the intra-day pattern of the average uninformed arrival rate for trade activity quartile 1. The upper right panel depicts the intra-day pattern of the average uninformed arrival rate for trade activity quartile 2. The lower left panel depicts the intra-day pattern of the average uninformed arrival rate for trade activity quartile 3. The lower right panel depicts the intra-day pattern of the average uninformed arrival rate for trade activity quartile 4.

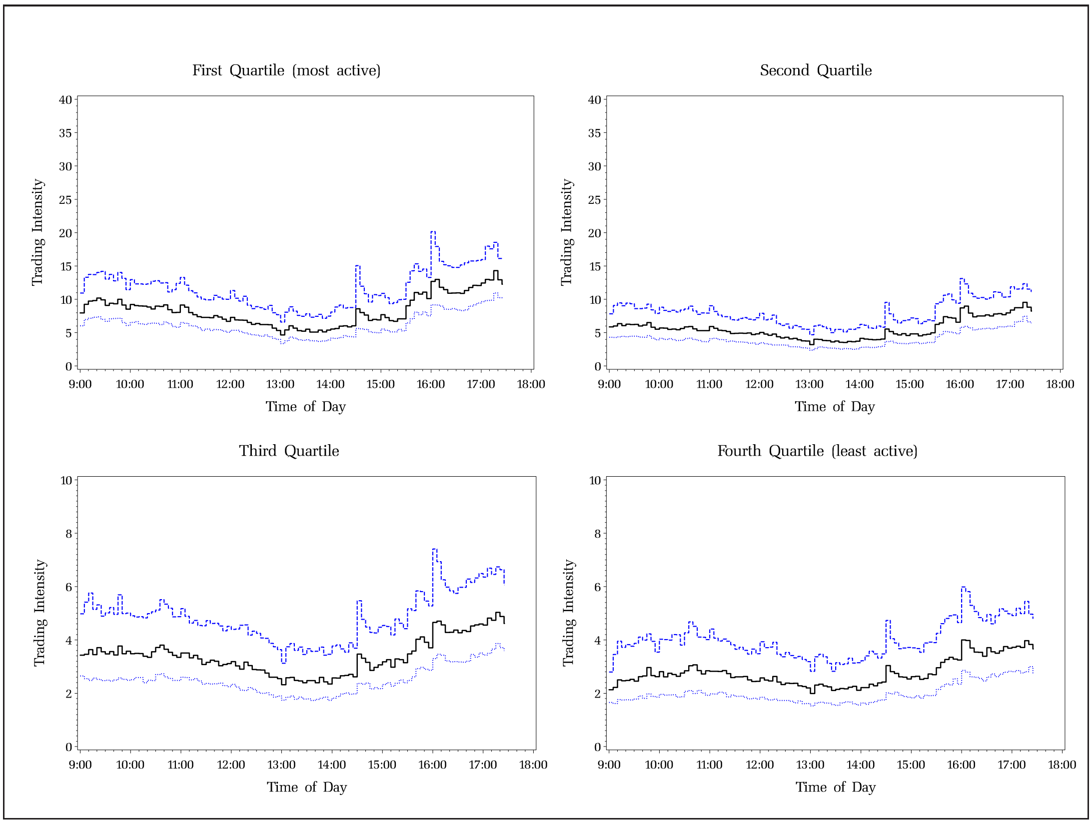

Figure B.2.2: **Intra-day pattern for the arrival rate of the informed traders.** The arrival rates μ_t are averaged across all stocks in the particular trade size quartile for each five minute interval. The solid line depicts the median μ_t. The dotted line depicts the 25% percentile and the dashed line depicts the 75% percentile. The upper left panel depicts the intra-day pattern of the informed arrival rate for trade activity quartile 1. The upper right panel depicts the intra-day pattern of the informed arrival rate for trade activity quartile 2. The lower left panel depicts the intra-day pattern of the informed arrival rate for trade activity quartile 3. The lower right panel depicts the intra-day pattern of the informed arrival rate for trade activity quartile 4.

Chapter 4

Using Mixed Poisson Distributions in Sequential Trade Models

The standard sequential trade model of Easley, Kiefer, O'Hara, and Paperman (1996) (EKOP) lacks to fit empirical data. Specifically, the counts of buys and sells exhibit overdispersion, are serially correlated and cross-correlated. Hence, making use of independent Poisson distributions to model the trade data does not seem to be appropriate. Instead, I propose to use mixed Poisson distributions, specifically the bivariate negative binomial, in order to capture the stylized facts of the data. I show in a simulation study that estimating the probability of informed trading with the standard Poisson model when the true data generating process comes from a more flexible distribution yields misleading results. An empirical study for DAX30 stocks traded on the Xetra platform confirms the simulation results.

Chapter is based on the article Using Mixed Poisson Distributions in Sequential Trade Models by O. Wuensche (2007)

4.1 Introduction

The model of Easley, Kiefer, O'Hara, and Paperman (1996) (henceforth EKOP) has been widely used by researchers to estimate the probability of informed trading. The measure stems from a simple and intuitive sequential trading model based on the number of buys and sells traded in a certain time interval.

Unfortunately, recent research has shown that several problems are associated with the standard model. For example, Boehmer, Grammig, and Theissen (2007) show that trade misclassification as proposed in Lee and Ready (1991) can lead to seriously biased estimates of the EKOP model. If one is merely interested in estimating the PIN, this problem can be circumvented by using aggregate trade data instead of disaggregated buy and sell data (Kokot 2004). But even for data where trade classification is not necessary, the EKOP model has various drawbacks. To understand those problems we have to go back to the basic assumptions of the model. In the EKOP model, buys and sells are assumed to be independently Poisson distributed conditional on a certain information regime. This implies that buys and sells are neither cross-correlated nor serially correlated. Further, one important characteristic of the Poisson distribution is that the mean of the distribution is equal to its variance. All three features of the model seem questionable when looking at empirical order flow data. As a result, the standard Poisson-EKOP model lacks empirical fit. Venter and de Jongh (2004) successfully proposed a bivariate Poisson inverse Gaussian mixture to enhance the empirical fit.

Drawing on the latter approach, this chapter provides a new distributional specification which improves the empirical fit drastically and is less demanding concerning the computational effort than comparable specifications. Further, it shows via a simulation study that if the data generating process of buys and sells comes from a more complex distribution than the Poisson, the standard EKOP model cannot estimate the PIN as well as the other structural parameters of the EKOP model consistently. Additionally, we show that the posterior classification in news and no news days of the new model confirms results that two-sided trade clustering is not associated with asymmetric information (Sarkar and Schwartz (2006)).

Several papers utilized the standard EKOP model addressing questions in microstructure research. Easley, Kiefer, and O'Hara (1997a) analyze the informational content of different trade sizes and Easley, Kiefer, and O'Hara (1997b) estimate the informational content of no trade intervals. Brown, Thomson, and Walsh (1999) and Easley, O'Hara, and Saar (2001) extend the basic model to allow for different order types, i.e. market versus limit order. Other noticeable applications were contributed by Easley, Kiefer, and O'Hara (1996) and Grammig, Schiereck, and Theissen (2001) who analyze stocks traded on different markets. The latter find that a stock traded in a market with trader anonymity, for example an open limit order book market, has a higher probability of informed trading (henceforth PIN) than the same stock on a non-anonymous market. Generally, a plethora of structural extensions and interesting applications has found its way to the literature. Any thinkable extension of the model is mainly restricted by the lack of appropriate data.

The already mentioned, mainly statistical, problem of interdependence between buys and sells as well as serial dependence has been dealt with in several ways. Easley, Engle, O'Hara, and Wu (2002) propose a bivariate vector autoregressive scheme for the conditional mean of buys and sells, thus capturing the time series dynamics of the trading process and the interaction between trading groups. However, they still use the Poisson distribution with time varying parametrization. Lei and Wu (2005) use a Markov Switching approach to obtain time varying parameters and induce a dependence between the trade intensities of informed and uninformed traders resulting in a dependence for the data generating process of buys and sells. Another possibility to tackle the problem of independence is to replace the standard Poisson distribution with a mixed Poisson distribution where the bivariate Poisson for buys and sells is mixed with the same distribution. Venter and de Jongh (2004) proposed the bivariate Poisson Inverse Gaussian (BPIG) and showed that the empirical fit of the model could be enhanced substantially. Kokot (2004) suggested a Negative Binomial (or Negbin) distribution to account for overdispersion. However, the Negbin model has not been applied to trade data thus far.

An introduction to the basic EKOP model has already been provided in section 3.2.1. In the next section I proceed with presenting a model taking into account unobserved heterogeneity within the trading groups and provide some distributional extensions and their implications. Section 4.3 reviews simulation results for the estimated EKOP parameters when the true data generating process comes from different mixed Poisson distributions. Then, in section 4.4, I apply the new bivariate Negbin model to 30 blue chips traded on the Xetra trading platform and assess the implications of the different outcomes. A short summary is given in section 4.5.

4.2 Heterogeneity Within the Trading Group and Mixed Poisson Distributions

From a theoretical point of view, making use of a mixed Poisson distribution can be justified by assuming not only heterogeneity *between* groups of traders, i.e. informed and uninformed traders, but additionally *within* a given group (see Kokot (2004)). Heterogeneity between groups is already accounted for in the standard Poisson specification by assuming different trading intensities of the two distinct trading groups. For an illustration of modeling intra-group heterogeneity, consider, as an example, the distribution of the random variable y with intensity λ. If we replace the constant conditional mean with the random variable $\tilde{\lambda}$, then:[1]

$$\tilde{\lambda}_k = \lambda w_k$$

with $E(w_k) = 1$ such that

$$E(\tilde{\lambda}_k) = E(\lambda w_k) = \lambda E(w_k) = \lambda$$

Note, that the distribution conditioned on w_k is still Poisson. But by definition, we cannot observe w_k. Hence, we have to integrate out the unobservable variable:

$$f(y; \lambda) = \int_0^\infty f(y|w; \lambda) \cdot g(w; \theta) dw \tag{4.1}$$

where f and g denote the pdf's of y and w, respectively.

Additionally, one could assume different characteristics for buyers and sellers in general by allowing the parameter of the mixing distribution to differ for buyer and seller such that:

$$f(B; \lambda_B) = \int_0^\infty f(B|w_B; \lambda_B) \cdot g(w_B; \theta_B) dw \tag{4.2}$$

$$f(S; \lambda_S) = \int_0^\infty f(S|w_S; \lambda_S) \cdot g(w_S; \theta_S) dw \tag{4.3}$$

[1] The example is analogous to the more general discussion about Poisson regression with unobserved heterogeneity in Long (1997) or Cameron and Trivedi (1998).

From a statistical point of view, Mixed Poisson distributions usually allow for different variance structures than the simple Poisson model where the mean is equal to the variance. Translated to the EKOP model it could be sensible to allow for different variances for the number of buys and sells in different news regimes (see for example Ranaldo (2006)). Thus, the parameter of the mixing distribution could be allowed to be regime dependent.

If the joint process of buys and sells is affected by the *same* factor, we can model the bivariate process of buys and sells as:

$$f(B, S; \lambda_B, \lambda_S, \theta) = \int_0^\infty f(B|w; \lambda_B) f(S|w; \lambda_S) \cdot g(w; \theta) dw \tag{4.4}$$

From a theoretical perspective, this common factor could be anything which affects both trader groups in the same fashion, e.g. important macro news, analyst reports (see Venter and de Jongh (2004)) or simply a common sentiment towards the market situation (see Henke (2004)). From a statistical point of view, the model described by equation (4.4) imposes the desired (because frequently observed) positive dependence of buys and sells. Note, that buys and sells are still Poisson distributed with expectation $\tilde{\lambda}_B = \lambda_B w$ and $\tilde{\lambda}_S = \lambda_S w$. Thus, the constant trading intensity for buyer and seller is now multiplied with the same stochastic factor. The dependence of buys and sells is a stylized fact when looking at aggregate buy and sell data and is left unaccounted for in the standard EKOP model. In fact, it will be shown in section 4.4 that more general models as the one in equation (4.4) are capable of generating data very similar to observed data while the standard model fails to produce a good empirical fit.

4.2.1 The Bivariate Poisson Inverse Gaussian Model

Depending on the choice of the mixing distribution g, one receives a corresponding Mixed Poisson distribution.[2] Venter and de Jongh (2004) proposed the bivariate Poisson Inverse Gaussian (BPIG) mixture as a possible alternative to the standard model. The BPIG uses as a mixing distribution the unit inverse Gaussian[3] with density function

$$g(w) = \frac{\psi e^{\psi^2}}{\sqrt{2\pi}} w^{-\frac{3}{2}} \exp\left[-\frac{1}{2}\psi^2(w^{-1} + w)\right], \quad w > 0 \tag{4.5}$$

[2] An extensive overview of mixed Poisson distributions can be found in Karlis and Xekalaki (2005).

[3] The distribution is also known as the standard form of the Wald distribution (see Johnson and Kotz (1970)).

As above for the Gamma distribution, we have $E(w) = 1$. The resulting joint distribution of buys and sells can be written as:

$$f_{BPIG}(B, S; \lambda_B, \lambda_S, \psi) = \frac{\lambda_B^B}{B!}\frac{\lambda_S^S}{S!}\left[\frac{\psi^2}{\psi^2 + 2(\lambda_b + \lambda_S)}\right]^{\frac{B+S}{2}} e^{\psi^2 - \psi\sqrt{z}} \times K_{B+S-\frac{1}{2}}(\psi\sqrt{z})/K_{\frac{1}{2}}(\psi\sqrt{z}) \quad (4.6)$$

where $z = \psi^2 + 2(\lambda_b + \lambda_S)$ and $K_n(\cdot)$ denotes the modified Bessel function of the third kind.[4]

Because $E(w) = 1$, the expectation of buys and sells in the j^{th} information regime are still, as in the standard Poisson case, $E_j(B) = \lambda_{j,B}$ and $E_j(S) = \lambda_{j,S}$. But the variance is now quadratic in the mean:

$$Var_j(B) = \lambda_{j,B} + \frac{1}{\psi^2}\lambda_{j,B}^2 \qquad Var_j(S) = \lambda_{j,S} + \frac{1}{\psi^2}\lambda_{j,S}^2 \quad (4.7)$$

For $\psi \to \infty$, we receive a standard independent bivariate Poisson model.

Although neglecting possible serial correlation in the time series of trades, the BPIG has been shown to fit empirical data quite well. however, a drawback is the rather complex likelihood function. Especially for large values of buys and sells, numerical maximization can be quite time consuming since the modified Bessel function is computed recursively.

4.2.2 The Bivariate Negbin Model

As an alternative to the BPIG presented above, we propose a bivariate Negbin distribution (BNB) for buys and sells. It has very similar features concerning the mean and the variance. The main advantage over the BPIG is that maximum likelihood estimation is hardly slower than estimation of the standard EKOP model.

Consider the case, that instead of a unit inverse Gaussian, w has a Gamma density with parameter ν:

$$g(w) = \frac{\nu^\nu}{\Gamma(\nu)} w^{\nu-1} e^{-w\nu} \quad (4.8)$$

where $\Gamma(\nu) = \int_0^\infty u^{\nu-1}e^{-u}du$ denotes the Gamma function. Note, that this is a special case of the two parameter Gamma distribution $w \sim G(\nu, \beta)$ where $\beta = \nu$ to ensure that $E(w) = 1$.

[4] For a comprehensive understanding of the function and useful hints for a convenient computation, see Abramowitz and Stegun (1972).

Then the joint distribution of buys and sells according to equation (4.4) can be written as:

$$f_{BNB}(B, S; \lambda_B, \lambda_S, \nu) = \frac{\Gamma(B + S + \nu^{-1})}{B!S!\Gamma(\nu^{-1})} \left(\frac{\nu^{-1}}{\zeta}\right)^{\nu^{-1}} \left(\frac{\lambda_B}{\zeta}\right)^{B} \left(\frac{\lambda_S}{\zeta}\right)^{S} \tag{4.9}$$

where $\zeta = \nu^{-1} + \lambda_B + \lambda_S$. We follow Marshall and Olkin (1990) who derived this model first and refer to (4.9) as the bivariate Negbin model. This model has already been applied in labor market research (see Bauer et al. (1998))[5] and in marketing research to model joint purchasing decisions (see Miles (2001)).

The marginals of buys (B) and sells (S) are univariate Negbin with density:

$$f(y|j) = \frac{\Gamma(y + \nu^{-1})}{\Gamma(y + 1)\Gamma(\nu^{-1})} \left(\frac{\nu^{-1}}{\nu^{-1} + \lambda_{j,y}}\right)^{\nu^{-1}} \left(\frac{\lambda_{j,y}}{\nu^{-1} + \lambda_{j,y}}\right)^{y} \tag{4.10}$$

where $y \in \{B, S\}$ and j denotes the information regime. As for the BPIG, the expectation of buys and sells is equal to $E_j(B) = \lambda_{j,B}$ and $E_j(S) = \lambda_{j,S}$. The variance results as

$$Var_j(B) = \lambda_{j,B} + \nu\lambda_{j,B}^2 \qquad Var_j(S) = \lambda_{j,S} + \nu\lambda_{j,S}^2 \tag{4.11}$$

such that the first two moments of the BPIG and the BNB are equal if $\nu = 1/\psi^2$.

Note, that the variance of buys (sells) in the good (bad) news regime is substantially higher than in the no news regime since the mean is increased by the trading intensity of the informed traders. As the BPIG, the BNB nests the Poisson model if $\nu \to 0$.

4.3 Simulation of PIN Under Different Distributions

Which consequence has a possible misspecification of the joint distribution of buys and sells for the estimation of the PIN? To answer this question, we conduct a simulation study. First, we simulate trade data assuming that the true data generating process (DGP) comes from an independent Poisson, a BNB and a BPIG. Then, we estimate the EKOP parameters for the simulated data, again with different distributional assumptions (Poisson, BNB, BPIG).

The parameter calibration for the simulation of buy and sell data is as follows: for all three

[5] Guo (1996) refers to the distribution given in equation (4.9) as the negative multinomial and Bauer et al. (1998) call it a bivariate Poisson although the marginals are Negbin. In their paper, the bivariate Negbin consists of two Negbin distributions mixed with a Beta distribution.

specifications, we set $\varepsilon = 100$, $\mu = 150$, $\delta = 0.5$ and let 1.) α and 2.) μ vary such that the true PIN ranges from 0.02 to 0.4 with an increment of 0.02.[6] For the simulation of BNB data we let the additional distribution parameter ν vary from 0.05 (close to being Poisson) to 1 (far from being Poisson) with an increment of 0.05. Remember that a higher ν allows for a higher variance of the buy and sell data. For the BPIG data we choose the additional distributional parameter such that the variance structure is equal to the BNB data. Therefore, we use the relationship $\psi = 1/\sqrt{\nu}$. For each combination of ν-PIN (ψ-PIN) we simulate 100 data series, estimate the three models and compute means of the estimated parameters.[7] The results are shown in table 4.3.1. As one can see for both mixture distributions, if the true PIN is small the Poisson model systematically delivers upward biased PIN estimates. This bias vanishes or is even slightly reversed if the true PIN gets large. As expected, the bias is more severe, the larger the deviation of the generating mixture distribution from the Poisson. The larger (smaller) the dispersion parameter for the BNB (BPIG), i.e. a larger deviation from the Poisson distribution, the heavier is the bias of the PIN estimates. A good impression of the inaccurate estimation of the PIN can be gained by looking at figure 4.3.1. If the true PIN gets smaller and the dispersion parameter larger, the gap between true and estimated PIN widens. In contrast, the lower panel of figure 4.3.1 shows that, both the BPIG and the BNB, can accurately estimate the PIN resulting from independent Poisson distributed buys and sells regardless of the true PIN value. This is no big surprise since the Poisson is nested in each of the two mixture distributions. The corresponding table 4.3.2 shows that the dispersion parameter is estimated near zero for the BNB case and relatively large for the BPIG case. Hence, the two mixture distributions approach the bivariate Poisson.

To get a more detailed impression of the origin of the PIN bias, we take a closer look at the estimated parameter values of the EKOP model.[8] Figure C.2.1 shows surface plots of the estimated structural parameters α (the probability that an information event happens) and μ (the trading intensity of informed traders). In case 1 (upper panel), we let the true α vary to obtain a range of PIN values. We can see that both parameters are not accurately estimated by the Poisson-EKOP model.

[6] I refer to section 3.2.1 and figure 3.2.1 to recall the structural interpretation of α, δ, ε and μ.

[7] In Appendix C.1, I provide the derivation of a numerically stable and "easy-to-compute" likelihood function for the BNB-EKOP.

[8] To conserve space, we only report results when the true data generating process is BNB. For the BPIG, the results are qualitatively the same.

Table 4.3.1: **True model BNB\BPIG - estimated model Poisson.** In the left panel we report the estimates of the Poisson-EKOP averaged across 100 replications when the true data generating process comes from a BNB. The data were generated with ν ranging from 0.05 (upper panel) to 1 (lower panel) and the true PIN ranging from 0.02 (left column) to 0.4 (right column). To vary the PIN, we fix μ and vary α. Instead of a BNB, we also used the BPIG for data generation. The results are reported in the right panel.

	Simulated: BNB - Estimated: Poisson					Simulated: BPIG - Estimated: Poisson				
	True PIN					True PIN				
	0.02	0.1	0.2	0.3	0.4	0.02	0.1	0.2	0.3	0.4
	$\nu = 0.05$					$\psi = 1/\sqrt{0.05}$				
α	0.122	0.164	0.311	0.506	0.784	0.118	0.168	0.313	0.508	0.793
	(0.05)	(0.03)	(0.03)	(0.03)	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.02)
ε	96.29	98.45	99.67	101.03	103.00	96.04	99.02	99.77	101.21	103.15
	(2.15)	(1.89)	(1.79)	(1.77)	(1.91)	(2.04)	(1.91)	(1.75)	(1.85)	(2.00)
δ	0.499	0.502	0.498	0.495	0.497	0.495	0.506	0.500	0.499	0.504
	(0.11)	(0.06)	(0.05)	(0.04)	(0.03)	(0.10)	(0.06)	(0.04)	(0.03)	(0.03)
μ	109.73	155.94	162.14	165.35	162.28	110.76	155.21	162.52	164.08	161.20
	(33.18)	(13.50)	(7.95)	(5.91)	(4.07)	(31.29)	(13.16)	(8.75)	(6.54)	(4.08)
PIN	**0.058**	**0.114**	**0.201**	**0.293**	**0.382**	**0.057**	**0.115**	**0.203**	**0.292**	**0.382**
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
	$\nu = 0.5$					$\psi = 1/\sqrt{0.5}$				
α	0.242	0.237	0.260	0.322	0.405	0.189	0.199	0.230	0.283	0.362
	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
ε	82.11	86.75	91.84	97.93	106.39	85.22	89.19	95.12	101.86	111.65
	(2.97)	(3.23)	(3.96)	(3.59)	(3.83)	(3.13)	(3.22)	(3.51)	(4.18)	(4.36)
δ	0.500	0.502	0.504	0.502	0.501	0.506	0.503	0.502	0.493	0.498
	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.07)	(0.06)	(0.05)	(0.04)	(0.03)
μ	161.09	207.22	254.91	283.82	301.43	177.98	224.05	268.06	294.48	305.89
	(13.79)	(21.32)	(25.44)	(18.18)	(16.00)	(19.18)	(28.96)	(30.75)	(27.81)	(23.47)
PIN	**0.191**	**0.219**	**0.263**	**0.317**	**0.364**	**0.163**	**0.198**	**0.242**	**0.288**	**0.330**
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
	$\nu = 1$					$\psi = 1$				
α	0.224	0.220	0.234	0.265	0.329	0.178	0.177	0.189	0.223	0.273
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
ε	76.93	81.41	86.61	91.76	100.75	80.51	85.72	91.47	99.06	108.01
	(4.24)	(3.63)	(4.11)	(5.05)	(5.37)	(4.01)	(3.65)	(4.08)	(4.88)	(4.84)
δ	0.499	0.499	0.503	0.506	0.497	0.501	0.499	0.505	0.503	0.504
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)
μ	221.61	276.19	330.68	381.66	413.10	236.92	300.42	361.52	404.02	424.58
	(20.87)	(29.73)	(33.80)	(36.11)	(31.09)	(28.97)	(44.42)	(49.86)	(52.58)	(42.05)
PIN	**0.243**	**0.270**	**0.306**	**0.354**	**0.401**	**0.205**	**0.234**	**0.268**	**0.310**	**0.347**
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)

The probability α (upper left panel) is overestimated when the true PIN and thus, the true α, is small but heavily underestimated when the true PIN gets large. In contrast, the trading intensity μ (upper right panel) is overestimated for most of the different simulation

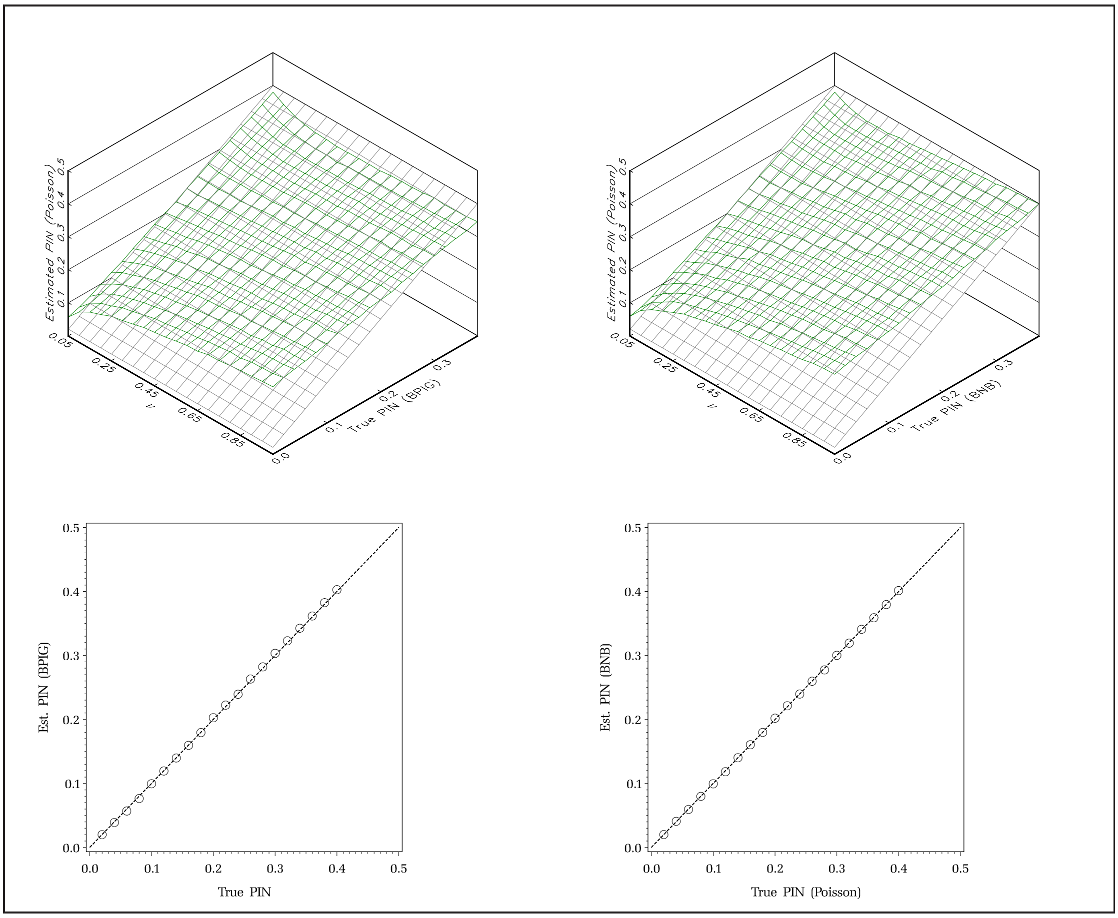

Figure 4.3.1: **PIN bias when the data generating process is BPIG\BNB\Poisson.** In the upper left panel, we plot the PIN estimated with the Poisson model when the true data generating process is a BPIG. In the upper right panel, we plot the PIN estimated with the Poisson model when the true data generating process is a BNB. The x-axis shows the range of values for the additional distributional parameter ν (BNB) and ψ (BPIG). For convenience, we use the ν-notation in both graphs since we make use of the relationship $\psi = 1/\sqrt{\nu}$. In the lower panel we plot the PIN estimated with the BPIG (left) and the BNB (right) when the true data generating process comes from a Poisson.

calibrations. Overall, this results in an overestimated PIN especially if the true PIN is small. When the true PIN gets large, the downward bias of α and the upward bias of μ seem to compensate each other. In case 2 (lower panel), we vary the true μ in order to receive the desired range of PIN values. α (lower left panel) is now biased downward for every calibration except for large values of the true PIN together with small values of the distribution parameter ν. The parameter μ (lower right panel) is biased upward for every calibration getting worse when ν is large. The upward bias of μ overcompensates the downward bias of α such that for the PIN the same picture evolves as in case 1. In both cases, δ is estimated accurately and not affected by the misspecification. The trading intensity of the uninformed traders ε

Table 4.3.2: **True model Poisson - estimated model BNB\BPIG.** In the left panel we report the estimates of the BNB-EKOP (left panel) and the BPIG-EKOP (right panel) averaged across 100 replications when the true data generating process comes from a Poisson. The data were generated with the true PIN ranging from 0.02 (left column) to 0.4 (right column). To vary the PIN, we fix μ and vary α.

	Simulated: Poisson - Estimated: BNB					Simulated: Poisson - Estimated: BPIG				
	True PIN					True PIN				
	0.02	0.1	0.2	0.3	0.4	0.02	0.1	0.2	0.3	0.4
α	0.027	0.149	0.339	0.572	0.891	0.027	0.149	0.345	0.587	0.901
	(0.01)	(0.02)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)	(0.03)	(0.03)	(0.01)
ε	99.84	99.99	99.92	99.77	99.79	100.31	100.22	99.92	99.37	99.50
	(1.10)	(1.15)	(1.33)	(1.53)	(1.65)	(1.09)	(1.21)	(1.41)	(1.52)	(1.51)
δ	0.49	0.50	0.50	0.50	0.49	0.50	0.49	0.50	0.50	0.50
	(0.20)	(0.07)	(0.04)	(0.03)	(0.02)	(0.17)	(0.06)	(0.04)	(0.03)	(0.02)
μ	153.69	148.61	149.70	149.70	150.19	155.56	149.40	147.74	147.60	148.90
	(22.93)	(9.06)	(5.76)	(3.83)	(3.18)	(19.70)	(8.54)	(5.58)	(4.23)	(3.27)
$\nu\backslash\psi$	0.00	0.00	0.00	0.00	0.00	7.98	8.09	8.46	9.37	11.70
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.55)	(0.72)	(0.70)	(0.75)	(1.02)
PIN	**0.020**	**0.100**	**0.202**	**0.300**	**0.401**	**0.020**	**0.100**	**0.203**	**0.303**	**0.403**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

is slightly underestimated for most simulation designs but the PIN bias is clearly steered by the failure to estimate α and μ.

4.4 Empirical Application

In order to verify the simulation results, I apply the standard Poisson EKOP, the BNB-EKOP and the BPIG-EKOP to stocks of the German blue chip index DAX30 traded on the Xetra trading platform. Xetra operates as a fully electronic open order book system as described in section 2.2. For the analysis, I use daily aggregated numbers of buys and sells from the 1*st* quarter of 2004 corresponding to 64 trading days. Selected descriptive statistics of the 30 stocks can be found in table 2.2.1. Before we compare the estimated PIN values for the three specifications, let us first assess the empirical fit by comparing the original data with simulated data using the estimated parameter values. In the left panel of figure 4.4.1 we plot the original buy and sell data for a representative stock[9] (SAP) in a scatter plot.

To classify the days in no news, good news and bad news days, we use the estimation

[9]The choice of the stock is irrelevant since the figures look qualitatively the same for every stock.

results of the respective model and compute posterior probabilities in the following fashion:

$$P(j_t|B_t, S_t) = p(j) f_i(B_t, S_t; \boldsymbol{\theta}_{i|j})/L_i \tag{4.12}$$

where $i \in \{\text{EKOP}, \text{BNB}, \text{BPIG}\}$, $p(j)$ is the unconditional probability for being in information regime j and $\boldsymbol{\theta}$ is a parameter vector depending on the model specification i and the news regime j. Note, that the standard model (upper left panel) tends to classify days with larger combinations of buys and sells as information days even if the number of buys and sells is relatively balanced, i.e. the absolute difference $|B - S|$ is small. This is inconsistent with the notion that two sided trade clustering, i.e. an increase in trading activity on both sides of the market is *not* associated with asymmetric information but rather with public information and a heterogeneous trading crowd with a strong divergence in beliefs. In contrast, using estimates of any of the two mixture distributions, only days with strong order imbalances are classified as information days, thereby confirming the notion of Sarkar and Schwartz (2006) that only heavier order imbalances are evidence for informed trading. Avramov, Chordia, and Goyal (2006) found that enhanced volatility stems from uninformed traders who herd together while informed traders reduce volatility. Computing the daily realized volatility[10] for my sample, we find that it is strongly positively correlated with the total number of trades ($\rho = 0.65$) but only weakly correlated with absolute order imbalance and uncorrelated with relative order imbalance. Hence, also from that perspective the new classification makes more sense than the classification obtained when estimating the standard model.

In the right panel, we use the respective estimated model parameters to simulate 64 buy-sell combinations. As can be easily seen, the standard EKOP model performs worst by far in generating data that resembles the true data. This result holds for every stock in the sample. Venter and de Jongh (2004) provide very similar results concerning the lack of empirical fit. The generated data coming from the other two specifications can hardly be distinguished by visual inspection and look very similar to what we observe on real markets. The multivariate distributions are able to capture the positive dependence between buys and sells in each information regime while the standard model builds separate clusters.

The simulation results can be confirmed when we compare the estimated PIN's and the

[10] I use the VARHAC estimator for daily realized volatility proposed in Bollen and Inder (2002).

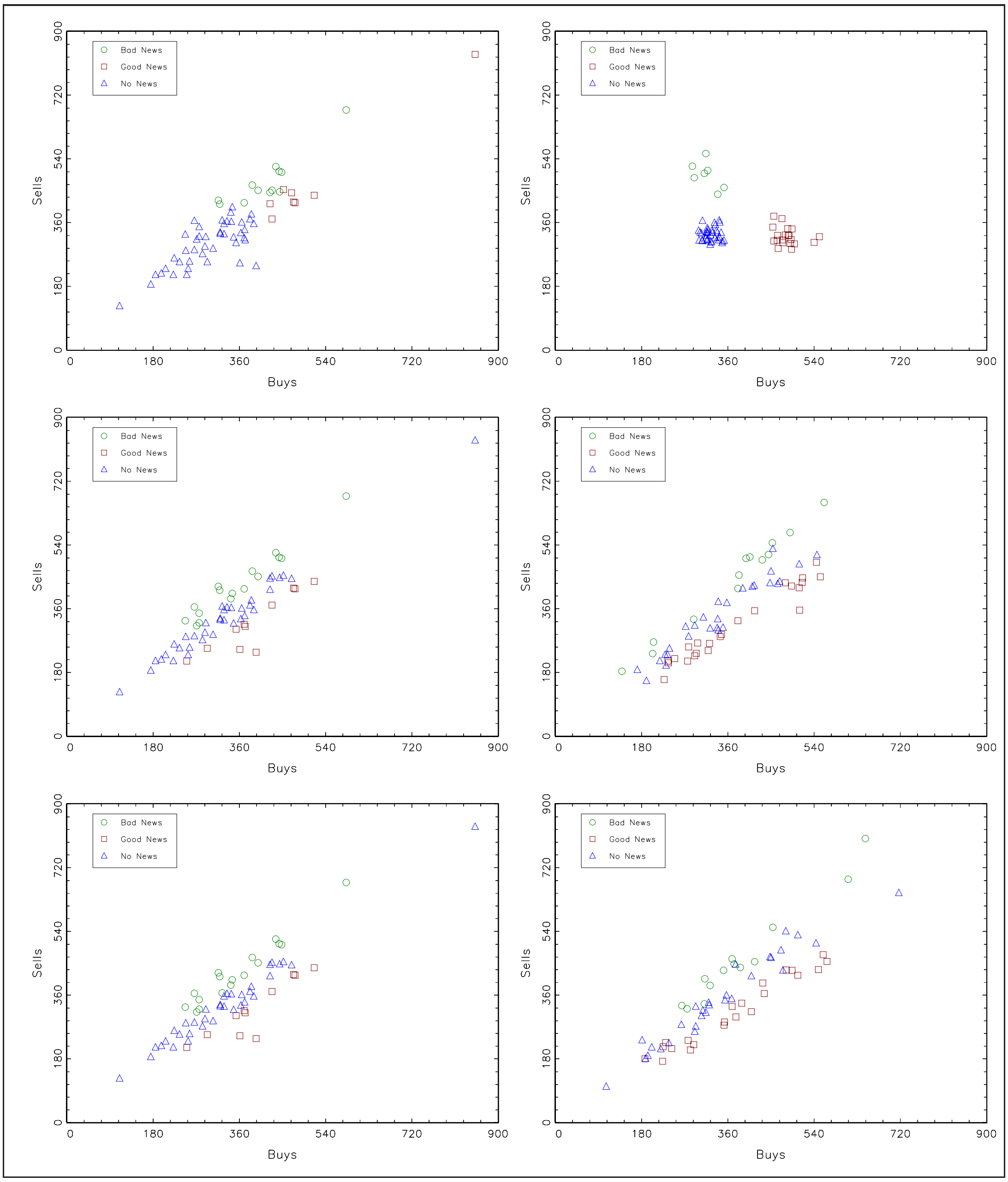

Figure 4.4.1: **Original vs. simulated data.** In the left panel, we plot the original buy and sell combinations of SAP. The days were classified as no news (triangle), good news (square) and bad news (circle) days. In the upper left days were classified using estimates of the standard EKOP, in the middle, days were classified using estimates of the BNB-EKOP and on the lower left, days were classified using estimates of the BPIG-EKOP. The right panel shows the corresponding simulated buy and sell data.

estimated structural parameters under different distributional assumptions.[11] For the BNB-EKOP, the estimated PIN's are smaller for each stock (except Henkel) when compared to the

[11] To conserve space, we only report the results of the BNB-EKOP. The results for the BPIG model are qualitatively the same.

Poisson-EKOP (see the upper left panel of figure 4.4.2).

Of special economic importance is that not only the value of the PIN changes systematically but also the PIN ranking of the 30 stocks is shuffled. In most applications of the EKOP model it is of minor importance if the PIN of stock A has a value of 0.05 or 0.07. It is rather important if the PIN of stock A is higher than the PIN of stock B. Computing rank correlation coefficients between the PINs and other information indicators such as market capitalization, Beltran, Grammig and Menkveld's α (see Beltran, Grammig, and Menkveld (2005)) and Madhavan, Richardson, Roomans θ[12] (see Madhavan, Richardson, and Roomans (1997)) reveals that the ranking of the BNB-EKOP might be more reasonable. The BNB-PIN is stronger negatively correlated with market capitalization (Poisson: -0.29 vs. BNB: -0.35) and stronger positively correlated with Beltran et al's α (Poisson: 0.16 vs. BNB: 0.20) than the Poisson-EKOP PIN across stocks. The correlation coefficient with Madhavan et al's θ, however, is almost equal.

Further, in the upper right panel of figure 4.4.2 we can see that the Poisson-EKOP tends to estimate a higher intensity of informed trading μ than the BNB-EKOP. The simulation results would predict that if the true statistical specification is BNB, the Poisson model overestimates μ for all *true* values of μ (compare with the lower right panel in figure C.2.1). In contrast, the unconditional probability that an information event happens, α, is predominantly lower in the Poisson model than in the BNB model as can be seen in the lower left panel of figure 4.4.2. Again, the simulation results can be confirmed. According to the upper right panel of figure C.2.1, we would predict a higher α using the Poisson model when the *true* α is very small and a much lower α when the *true* α gets large. Assuming the BNB specification to be the correct specification, this is exactly what we observe when using empirical data. The estimates for ε (lower right panel of figure 4.4.2) are quite similar in both models and aligned along the bisecting line. This is no big surprise since the simulation results did not reveal any severe deviation from the true value using the Poisson model. The same is true for the parameter δ.

[12] Instead of the raw θ estimate I use a standardized $\tilde{\theta} = \theta/\bar{P}$ where $\bar{P}$ is the average stock price.

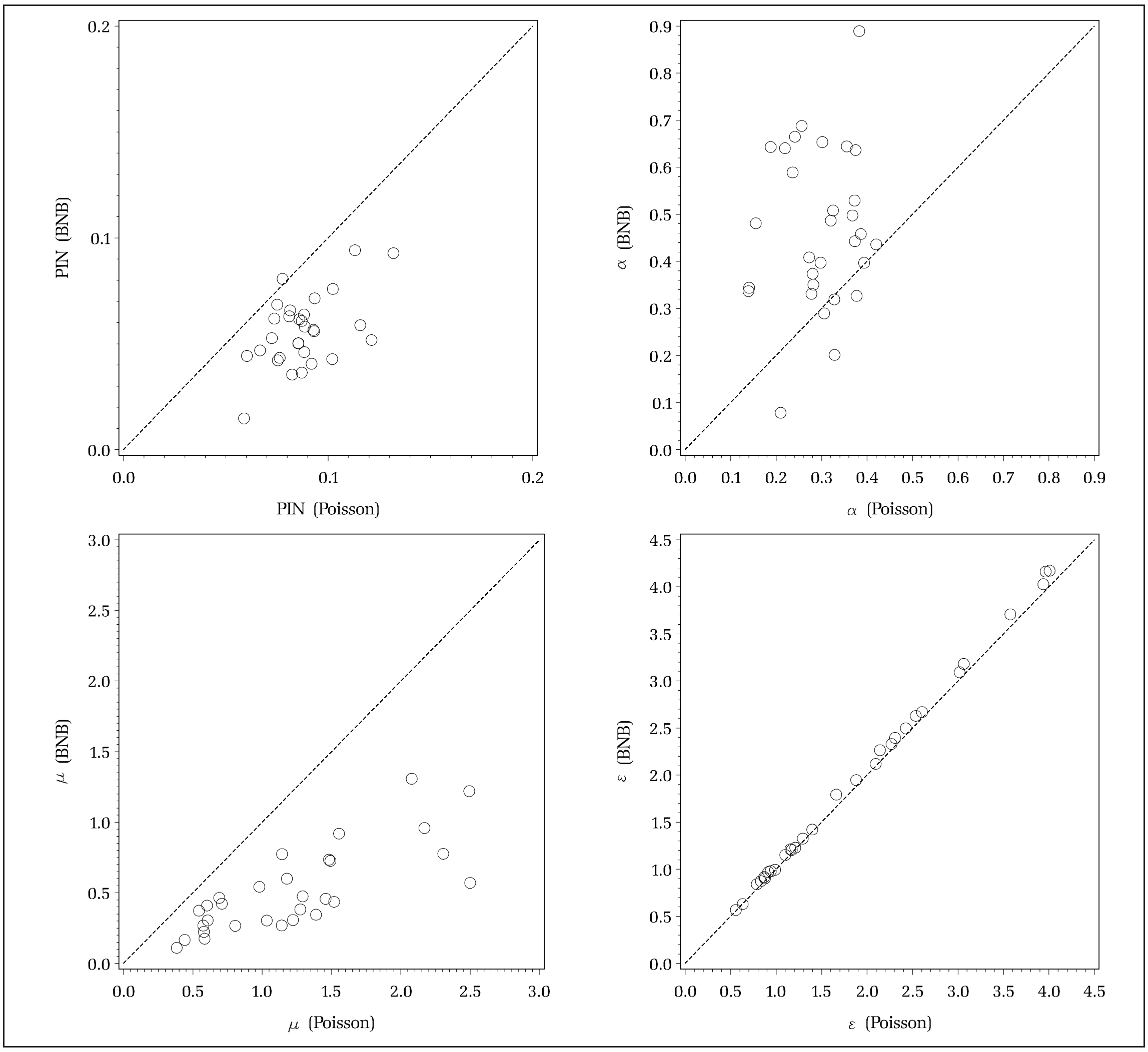

Figure 4.4.2: **BNB-EKOP Estimates vs. Poisson-EKOP Estimates.** The estimated parameters of the BNB-EKOP are contrasted with the estimated parameters of the Poisson-EKOP in a scatter plot. The upper left panel shows the estimated PIN, the upper right panel the α, the lower left panel the μ and the lower right panel the ε.

4.5 Conclusion

This chapter provides a new distributional specification for the EKOP model. It has been shown that the commonly used Poisson distribution lacks empirical fit. Therefore, we introduce a bivariate negative binomial distribution which takes into account not only heterogeneity between the two groups of informed and uninformed traders but also within the two trader groups. The specification is similar to the computationally time consuming bivariate Poisson inverse Gaussian model proposed by Venter and de Jongh (2004) but much more tractable. From a statistical point of view, we take into account the positive correlation between the number of buys and sells in a certain time interval. This correlation is a well documented stylized fact but unaccounted for in the standard Poisson-EKOP model where buys and sells are by construction negatively correlated. We have further shown in a simulation experiment that if the data generating process comes from such a mixture distribution, using the Poisson yields biased parameter estimates. Especially the PIN, a widely used measure for the degree of informed trading, is systematically upward biased.

Eventually, we applied the new model to the DAX30 stocks traded on the Xetra trading platform and compared the results with the classic Poisson-EKOP. We find that the empirical estimation results closely resemble the simulation results and provide further evidence that the Poisson assumption yields misleading results. Not only the parameter estimates but also the posterior classification of days in news or no news days is quite different. While the Poisson-EKOP tends to classify *all* days with a large number of trades as information days, the BNB-EKOP only reacts to a large order imbalance. This is in accord with the notion that high volatility caused by a high trading intensity on both sides of the market is mainly due to public information and a divergence of beliefs rather than private information.

Appendix C

C.1 Derivation of Stable Bivariate Negbin Likelihood

The likelihood function of the EKOP model with a bivariate Negative Binomial distribution for buys and sells is given by:

$$\begin{aligned}
\mathfrak{L}_{BNB}(\boldsymbol{\theta}) &= \prod_{t=1}^{T}(1-\alpha)\frac{\Gamma(B_t+S_t+\nu^{-1})}{B_t!S_t!\Gamma(\nu^{-1})}\left(\frac{\nu^{-1}}{2\varepsilon+\nu^{-1}}\right)^{\nu^{-1}}\left(\frac{\varepsilon}{2\varepsilon+\nu^{-1}}\right)^{B_t}\left(\frac{\varepsilon}{2\varepsilon+\nu^{-1}}\right)^{S_t} \\
&+\alpha\delta\frac{\Gamma(B_t+S_t+\nu^{-1})}{B_t!S_t!\Gamma(\nu^{-1})}\left(\frac{\nu^{-1}}{2\varepsilon+\nu^{-1}+\mu}\right)^{\nu^{-1}}\left(\frac{\varepsilon}{2\varepsilon+\nu^{-1}+\mu}\right)^{B_t}\left(\frac{\varepsilon+\mu}{2\varepsilon+\nu^{-1}+\mu}\right)^{S_t} \\
&+\alpha(1-\delta)\frac{\Gamma(B_t+S_t+\nu^{-1})}{B_t!S_t!\Gamma(\nu^{-1})}\left(\frac{\nu^{-1}}{2\varepsilon+\nu^{-1}+\mu}\right)^{\nu^{-1}}\left(\frac{\varepsilon+\mu}{2\varepsilon+\nu^{-1}+\mu}\right)^{B_t}\left(\frac{\varepsilon}{2\varepsilon+\nu^{-1}+\mu}\right)^{S_t}
\end{aligned}$$

Defining $a = \mu/\varepsilon$ and $b = \mu/(2\varepsilon + \nu^{-1})$ and and factoring out yields:

$$\begin{aligned}
\mathfrak{L}_{BNB}(\boldsymbol{\theta}) &= \prod_{t=1}^{T}\frac{\Gamma(B_t+S_t+\nu^{-1})}{B_t!S_t!\Gamma(\nu^{-1})}\left(\frac{\nu^{-1}}{2\varepsilon+\nu^{-1}}\right)^{\nu^{-1}}\left(\frac{\varepsilon}{2\varepsilon+\nu^{-1}}\right)^{B_t}\left(\frac{\varepsilon}{2\varepsilon+\nu^{-1}}\right)^{S_t} \\
&\quad\times\Bigg\{(1-\alpha)+\alpha\delta\left(\frac{1}{1+b}\right)^{\nu^{-1}}\left(\frac{1}{1+b}\right)^{B_t}\left(\frac{1+a}{1+b}\right)^{S_t} \\
&\quad+\alpha(1-\delta)\left(\frac{1}{1+b}\right)^{\nu^{-1}}\left(\frac{1+a}{1+b}\right)^{B_t}\left(\frac{1}{1+b}\right)^{S_t}\Bigg\}
\end{aligned}$$

Rearranging terms yields:

$$\mathfrak{L}_{BNB}(\boldsymbol{\theta}) = \prod_{t=1}^{T} \frac{\Gamma(B_t + S_t + \nu^{-1})}{B_t!S_t!\Gamma(\nu^{-1})} \varepsilon^{(B_t+S_t)} \nu^{-1\nu^{-1}} \left(\frac{1}{2\varepsilon + \nu^{-1}}\right)^{(B_t+S_t+\nu^{-1})}$$
$$\times \left\{ (1-\alpha) + \left[\alpha\delta(1+a)_t^S + \alpha(1-\delta)(1+a)_t^B\right] \left(\frac{1}{1+b}\right)^{(B_t+S_t+\nu^{-1})} \right\}$$

Applying exp and log to the power terms inside the curly brackets and taking the logarithm of the whole function, we receive receive the log-likelihood function (denoting $\ln \mathfrak{L}_{BNB} := \mathfrak{l}_{BNB}$):

$$\begin{aligned}
\mathfrak{l}_{BNB}(\boldsymbol{\theta}) = & \sum_{t=1}^{T} \ln \Gamma(B_t + S_t + \nu^{-1}) - \ln B_t! - \ln S_t! - \ln \Gamma(\nu^{-1}) \\
& + (B_t + S_t) \ln \varepsilon + \nu^{-1} \ln \nu^{-1} - (B_t + S_t + \nu^{-1}) \ln(2\varepsilon + \nu^{-1}) \\
& \times \ln \{(1+\alpha) + [\alpha\delta \exp(S_t \ln(1+a)) \\
& + \alpha(1-\delta) \exp(B_t \ln(1+a)) \exp(-(B_t + S_t + \nu^{-1})) \ln b
\end{aligned} \tag{C.14}$$

Note, that the terms including the factorials can be dropped if necessary, since they only shift the level of the likelihood value but not the location of the maximum. ■

C.2 Parameter Bias

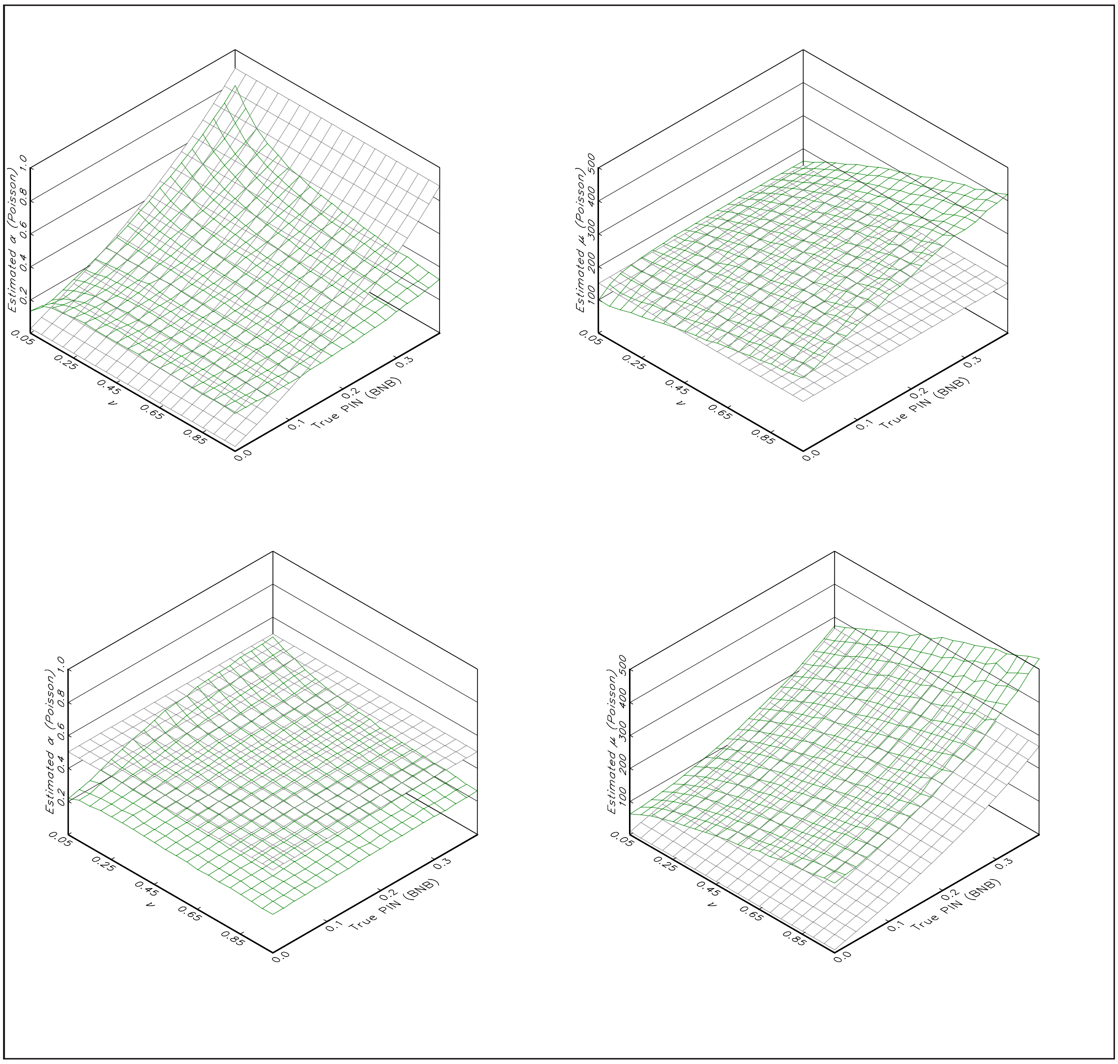

Figure C.2.1: **Parameter bias when μ or α are varied.** In the left panel, we plot the estimated α versus the true α. In the right panel, the estimated μ is plotted against the true μ. The x-axis shows the range of values for the additional distributional parameter ν (BNB). In the upper panel, we vary the true α to obtain the range of PIN values shown on the y-axis and in the lower panel we vary the true μ to obtain the range of PIN values shown on the y-axis. The true data generating process comes from a BNB-EKOP and a Poisson-EKOP is estimated.

Chapter 5

Summary and Outlook

In this thesis I present extensions of two famous models measuring information in the context of financial market microstructure. In the first part I present an extension of the spread decomposition model of Madhavan, Richardson, and Roomans (1997) taking into account the role of time for the informational content of a trade. The chapter is based on an article of Grammig, Theissen, and Wünsche (2007). The basic model splits the bid-ask spread in a non-information related share due to institutional order processing costs and an information related share depending on the surprise in order flow. Our extension consists of letting the information related part depend on the waiting time between the current and the last transaction. This extension is inspired by the finding of Dufour and Engle (2000) that high trading activity is associated with more information in the market. While challenging their results we find that for the DAX30 stocks traded on the FSE and a matched sample of NYSE stocks a negative relationship exists between trading activity measured as durations between transactions and the informational content of a trade. Further, we argue that the different results might be due to the modeling approach. While we measure information as the adverse selection share of the bid-ask spread, Dufour and Engle (2000) use the impact of a trade on the midquote revision following the trade. Specific trade types such as partially filled market-to-limit orders or partially filled limit orders with a limit price equal or better than the current best quote change the midquote substantially by shifting the bid and ask price in the same direction. Hence, they are by construction very informative in the VAR approach of Dufour/Engle. In our framework, however, this is not necessarily the case. It would be desirable for future research to further analyze *different* order types with respect to their informational content in greater detail.

The second part of this thesis deals with another type of microstructure model which allows to measure the probability of informed trading of any exchange traded asset. The model class of sequential trade models like the EKOP model make use of transaction counts to estimate the trading intensity of informed and uninformed traders, the probability that an information event occurred as well as the probability of the signal type (good or bad news). In empirical applications the cross correlation and the serial dependence of the count data series of buys and sells is often neglected. Instead, it is assumed that aggregated buys and sells in a predefined time interval are independently Poisson distributed. In chapter 3 I make use of an extended version of the EKOP model proposed by Easley, Engle, O'Hara, and Wu (2002) which allows for time varying arrival rates. The dynamics of the trading intensities of informed and uninformed traders are modeled as a bivariate vector process. In contrast to Engle et al (2004) I use five minute intervals to count buys and sells instead of daily aggregates. First, the strategic behavior of the two trader groups can be better measured on a higher frequency and second, intra-day trading patterns can be revealed. The results indicate that informed traders try to enter the market when uninformed trading activity is high. The behavior of the uninformed traders is more ambiguous and depends on the size of the traded company. For larger stocks, uninformed traders tend to avoid informed traders while for smaller stocks they follow the informed. This is consistent with the theory that for stocks with a very fast-paced information flow (presumably large stocks) informed traders immediately exploit all their information before it becomes worthless. For smaller stocks where information is not revealed that rapidly it might be preferable to exploit superior information more slowly in order to avoid adverse price effects. A further improvement of the empirical analysis would be taking into account the intra-daily seasonal pattern of the trade intensities, though adding additional complexity to the model might severely hamper a stable convergence.

Another way of dealing with the problem of independent buys and sells could be a modification of the distributional assumption. In chapter 4 I have shown that buy and sell combinations generated with estimates from an independent bivariate Poisson model do not resemble observed data. This lack of empirical fit has already been addressed by Venter and de Jongh (2004). They propose to use a bivariate Poisson Inverse Gaussian distribution which introduces dependence between the number of buys and sells in a given time interval. In order to decrease complexity in the numerical optimization I propose the bivariate negative binomial distribution. The latter model is much less time consuming in the estimation process but fits the data as well as the Poisson Inverse Gaussian model. I further analyze how structural model parameters are affected when estimating the standard Poisson model, assuming the class of Poisson mixture distributions to be the true data generating process. It is shown in a simulation study that in this case, the probability of informed trading is systematically biased upward. The bias is more severe when the true PIN is very small. This could lead to serious problems when using the estimated PIN in cross sectional regressions. Moreover, I show in an empirical study that the PIN ranking of different stocks changes as well whenever the distributional assumption is altered. This is even more problematic since it is often the ranking which constitutes evidence in favor of or against a theoretical hypothesis. Overall, there is great potential in using the class of mixed Poisson distributions for the empirical analysis of sequential trade models since it describes the data fairly well and is not necessarily more technically demanding than the standard Poisson model.

Bibliography

Abramowitz, M., and I. Stegun (1972): *Handbook of Mathematical Functions*. Dover Publications, Inc.

Admati, A., and P. Pfleiderer (1988): "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies*, 1, 3–40.

Ahn, H., K. Bae, and K. Chan (2001): "Limit Orders, Depth and Volatility: Evidence from the Stock Exchange of Hong Kong," *Journal of Finance*, 56, 767–788.

Aït-Sahalia, Y., P. Mykland, and L. Zhang (2005): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18(2), 351–416.

Amihud, Y., and N. Mendelson (1980): "Dealership Market: Market Making with Inventory," *Journal of Financial Economics*, 8, 31–53.

Avramov, D., T. Chordia, and A. Goyal (2006): "The Impact of Trades on Daily Volatility," *Review of Financial Studies*, 19(4), 1241–1277.

Bagehot, W. (1971): "The Only Game in Town," *Financial Analysts Journal*, 27, 12–14, 22.

Bauer, T., A. Million, R. Rotte, and K. Zimmermann (1998): "Immigrant Labor And Workplace Safety," IZA Discussion Paper No. 16.

Bauwens, L., and P. Giot (2000): "The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks," *Annales d'Économie et de Statistique*, 60, 117–149.

——— (2001): *Econometric Modelling of Stock Market Intraday Activity*. Kluwer Academic Publishers.

Beltran, H., J. Grammig, and A. Menkveld (2005): "Understanding the Limit Order Book: Conditioning on Trade Informativeness," Working Paper.

Biais, B., L. Glosten, and C. Spatt (2005): "Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications," *Journal of Financial Markets*, 8, 217–264.

Boehmer, E., J. Grammig, and E. Theissen (2007): "Estimating the Probability of Informed Trading - Does Trade Misclassification Matter?," *Journal of Financial Markets*, 10, 26–47.

Bollen, B., and B. Inder (2002): "Estimating Daily Volatility in Financial Markets Utilizing Intraday Data," *Journal of Empirical Finance*, 9, 551–562.

Brock, W., and A. Kleidon (1992): "Periodic Market Closure and Trading Volume: A Model of Intraday Bids and Asks," *Journal of Economic Dynamics and Control*, 16, 451–489.

Brown, P., N. Thomson, and D. Walsh (1999): "Characteristics of the Order Flow Through An Electronic Open Limit Order Book," *Journal of International Financial Markets, Institutions and Money*, 9, 335–357.

Cameron, A., and P. Trivedi (1998): *Regression Analysis of Count Data*. Cambridge University Press.

Cao, C., O. Hansch, and X. Wang (2004): "The Informational Content of an Open Limit Order Book," EFA 2004 Working Paper, Penn State University and Southern Illinois University.

Chordia, T., R. Roll, and A. Subrahmanyam (2002): "Order Imbalance, Liquidity and Market Returns," *Journal of Financial Economics*, 65, 111–130.

——— (2005): "Evidence on the Speed of Convergence to Market Efficiency," *Journal of Financial Economics*, 76, 271–292.

Copeland, T., and D. Galai (1983): "Information Effects on the Bid-Ask Spread," *Journal of Finance*, 38(5), 1475–1469.

Diamond, D., and R. Verrecchia (1987): "Constraints on Short-Selling and Asset Price Adjustments to Private Information," *Journal of Financial Economics*, 18, 277–311.

Dufour, A., and R. Engle (2000): "Time and the Price Impact of a Trade," *Journal of Finance*, 55, 2467–2498.

Easley, D., R. F. Engle, M. O'Hara, and L. Wu (2002): "Time-Varying Arrival Rates of Informed and Uniformed Trades," Working Paper.

Easley, D., N. Kiefer, and M. O'Hara (1996): "Cream-Skimming or Profit Sharing? The Curious Role of Purchased Order Flow," *Journal of Finance*, 51, 811–833.

——— (1997a): "The Information Content of the Trading Process," *Journal of Empirical Finance*, 4, 159–186.

——— (1997b): "One Day in the Life of a Very Common Stock," *Review of Financial Studies*, 10(3), 805–835.

Easley, D., N. Kiefer, M. O'Hara, and J. Paperman (1996): "Liquidity, Information and Infrequently Traded Stocks," *Journal of Finance*, 51(4), 1405–1436.

Easley, D., and M. O'Hara (1992): "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47(2), 577–605.

Easley, D., M. O'Hara, and G. Saar (2001): "How Stock Splits Affect Trading: A Microstructure Approach," *Journal of Financial and Quantitative Analysis*, 36, 25–51.

Engle, R. (2000): "The Econometrics of Ultra High Frequency Data," *Econometrica*, 68(1), 1–22.

Engle, R., and J. Russell (1998): "Autoregressive Conditional Duration: A New Model For Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Eubank, R., and P. Speckman (1990): "Curve Fitting by Polynomial-Trigonometric Regression," *Biometrika*, 77(1), 1–9.

Fernandes, M., and J. Grammig (2006): "A Family of Autoregressive Conditional Duration Models," *Journal of Econometrics*, 130, 1–23.

Flood, M., R. Huisman, K. Koedijk, and R. Lyons (1998): "Search Costs: The Neglected Spread Component," Working Paper.

Foster, F., and S. Viswanathan (1990): "A Theory of the Interday Variations in Volume, Variance and Trading Costs in Securities Markets," *Review of Financial Studies*, 3, 593–624.

Foucault, T. (1999): "Order Flow Composition and Trading Costs in a Dynamic Limit Order Market," *Journal of Financial Markets*, 2, 99–134.

Froot, K., D. Scharfstein, and J. Stein (1992): "Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation," *Journal of Finance*, 47, 1461–1484.

Garman, M. (1976): "Market Microstructure," *Journal of Financial Economics*, 3, 257–275.

George, T., G. Kaul, and M. Nimalendran (1991): "Estimation of the Bid-Ask Spread and its Components: A New Approach," *Review of Financial Studies*, 4(4), 623–656.

Gerety, M., and H. Mulherin (1992): "Trading Halts and Market Activity: An Analysis of Volume at the Open and the Close," *Journal of Finance*, 47, 1765–1784.

Geweke, J. (1986): "Modelling the Persistence of Conditional Variances: A Comment," *Econometric Reviews*, 5, 57–61.

Ghysels, E., and J. Jasiak (1998): "GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model," *Studies in Nonlinear Economics & Econometrics*, 2(4), 133–149.

Glosten, L. (1987): "Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices," *Journal of Finance*, 42(5), 1293–1307.

Glosten, L., and L. Harris (1988): "Estimating the Components of the Bid-Ask Spread," *Journal of Financial Economics*, 21, 123–142.

Glosten, L., and P. Milgrom (1985): "Bid, Ask and Transaction Prices in a Specialist Maket with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71–100.

Gomber, P., U. Schweickert, and E. Theissen (2004): "Zooming in on Liquidity," Working Paper.

Grammig, J., D. Schiereck, and E. Theissen (2001): "Knowing Me, Knowing You: Trader Anonymity and Informed Trading in Parallel Markets," *Journal of Financial Markets*, 4, 385–412.

Grammig, J., E. Theissen, and O. Wünsche (2007): "Time and the Price Impact of a Trade: A Structural Approach," Working Paper.

Grammig, J., and M. Wellner (2002): "Modelling the Interdependence of Volatility and Inter-Transaction Duration Processes," *Journal of Econometrics*, 106, 369–400.

Guo, G. (1996): "Negative Multinomial Regression Models for Clustered Event Counts," *Sociological Methodology*, 26, 113–132.

Harris, L. (2003): *Trading and Exchanges*. Oxford University Press.

Hasbrouck, J. (1991a): "Measuring the Information Content of Stock Trades," *Journal of Finance*, 46, 179–207.

——— (1991b): "The Summary Informativeness of Stock Trades: An Econometric Analysis," *The Review of Financial Studies*, 4(3), 571–595.

——— (2007): *Empirical Market Microstructure*. Oxford University Press.

Henke, H. (2004): "Measuring the Probability of Informed Trading: Estimation Error and Trading Frequency," Working Paper.

Ho, T., and R. Macris (1984): "Dealer Bid-Ask Quotes and Transaction prices: An Empirical Study of Some AMEX Options," *Journal of Finance*, 39(1), 23–45.

Ho, T., and H. Stoll (1981): "Optimal Dealer Pricing under Transactions and Return Uncertainty," *Journal of Financial Economics*, 9, 47–73.

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

Irvine, P., G. Benston, and E. Kandel (2000): "Liquidity Beyond the Inside Spread: Measuring and Using Information in the Limit Order Book," Working Paper.

Johnson, N., and S. Kotz (1970): *Continuous Univariate Distributions - I*. Wiley & Sons, Inc.

Karlis, D., and E. Xekalaki (2005): "Mixed Poisson Distributions," *International Statistical Review*, 73(1), 35–58.

Katti, S. (1960): "Moments of the Absolute Difference and Absolute Deviation of Discrete Distributions," *The Annals of Mathematical Statistics*, 31(1), 78–85.

Kokot, S. (2004): *The Econometrics of Sequential Trade Models*. Springer.

Kyle, A. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1336.

Lee, C., and M. Ready (1991): "Inferring Trade Direction From Intraday Data," *Journal of Finance*, 44(2), 733–746.

Lei, Q., and G. Wu (2005): "Time-varying Informed and Uninformed Trading Activities," *Journal of Financial Markets*, 8(2), 153–181.

Long, J. (1997): *Regression Models for Categorial and Limited Dependent Variables*. Sage Publications.

Madhavan, A. (2000): "Market Microstructure: A Survey," *Journal of Financial Markets*, 3, 205–258.

Madhavan, A., M. Richardson, and M. Roomans (1997): "Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks," *Review of Financial Studies*, 10, 1035–1064.

Marshall, A., and I. Olkin (1990): "Multivariate Distribution Generated from Mixtures of Convolution and Product Families," in *Topics in Statistical Dependence*, ed. by H. Block, A. Sampson, and T. Savits, vol. 16 of *IMS Lecture Note Monograph Series*, pp. 371–393. Institute of Mathematical Statistics.

Miles, D. (2001): "Joint Purchasing Decisions: A Multivariate Negative Binomial Approach," *Applied Economics*, 33, 937–946.

Nelson, D. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Nyholm, K. (2002): "Estimating the Probability of Informed Trading," *Journal of Financial Research*, 25, 485–505.

Odders-White, E., and M. Ready (2004): "Credit Ratings and Stock Liquidity," Working Paper.

O'Hara, M. (1995): *Market Microstructure Theory*. Blackwell.

Parlour, C. (1998): "Price Dynamics in Limit Order Markets," *The Review of Financial Studies*, 11(4), 789–816.

Ranaldo, A. (2006): "Intraday Market Dynamics Around Public Information Arrivals," Swiss National Bank Working Papers.

Sarkar, A., and R. Schwartz (2006): "Two-Sided Markets and Intertemporal Trade Clustering: Insights into Trading Motives," Federal Reserve Bank of New York Staff Report no. 246.

Stoll, H. (1978a): "The Supply of Dealer Services in Securities Markets," *Journal of Finance*, 33(4), 1133–1151.

——— (1978b): "The Pricing of Security Dealer Services: An Empirical Study of NASDAQ Stocks," *Journal of Finance*, 33(4), 1153–1172.

Venter, J., and D. de Jongh (2004): "Extending the EKOP Model to Estimate the Probability of Informed Trading," Working Paper.

Zhang, M., J. Russell, and R. Tsay (2001): "A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data," *Journal of Econometrics*, 104, 179–207.