



No. 2010/12

Analysis of Binary Trading Patterns in Xetra

Kai-Oliver Maurer and Carsten Schäfer





Center for Financial Studies

The *Center for Financial Studies* is a nonprofit research organization, supported by an association of more than 120 banks, insurance companies, industrial corporations and public institutions. Established in 1968 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia.

The CFS Working Paper Series presents the result of scientific research on selected topics in the field of money, banking and finance. The authors were either participants in the Center's Research Fellow Program or members of one of the Center's Research Projects.

If you would like to know more about the *Center for Financial Studies*, please let us know of your interest.

Prof. Michalis Haliassos, Ph.D.

Prof. Dr. Jan Pieter Krahen

Prof. Dr. Uwe Walz



CFS Working Paper No. 2010/12

Analysis of Binary Trading Patterns in Xetra*

Kai-Oliver Maurer¹ and Carsten Schäfer²

This Version: June 2010

Abstract:

This paper proposes the Shannon entropy as an appropriate one-dimensional measure of behavioural trading patterns in financial markets. The concept is applied to the illustrative example of algorithmic vs. non-algorithmic trading and empirical data from Deutsche Börse's electronic cash equity trading system, Xetra. The results reveal pronounced differences between algorithmic and non-algorithmic traders. In particular, trading patterns of algorithmic traders exhibit a medium degree of regularity while non-algorithmic trading tends towards either very regular or very irregular trading patterns.

JEL Classification: C40, D0, G14, G15, G20

Keywords: Financial Markets, Electronic Markets, Algorithmic Trading, Order Entry, Equity Trading, Information Theory, Entropy Measure

* The views expressed in this article are those of the authors and do not necessarily reflect the views of Deutsche Börse AG, any of its subsidiaries, or anyone else. The authors would like to thank Martin Reck, Séverine Vandelanoite, Erik Theissen, and Peter Gomber as well as participants of the regular Deutsche Börse/Center for Financial Studies seminar for support and helpful comments.

1 Head of Market & Business Research, Deutsche Börse AG, E-mail: kai-oliver.maurer@deutsche-boerse.com

2 Project Manager, Market & Business Research, Deutsche Börse AG, E-mail: carsten.schaefer@deutsche-boerse.com

1 Introduction

Financial markets are subject to a continuous change process driven by innovation in terms of products, services and technology. These innovatory activities are not independent of each other but are actually mutually reinforcing. Technological progress enables realization of new financial instruments or new trading strategies, for example. Thus, existing business models and strategies are adapting and new ones are created. A few examples: At the turn of the century automation of exchange trading in Europe and in the US released substantial efficiency gains and facilitated realization of global investment strategies and more advanced, automated trading models. Consequently, algorithmic trading has been one major driver of trading activity in financial markets. Ease and effectiveness of market access also promoted the importance of specific investor groups as direct market participants diminishing the role of intermediaries in certain areas of the market.

For service providers such as operators of regulated markets it is key to identify changes in the market place as early as possible in order to enhance respective services and to continue to meet the requirements of existing and new market participants. From a quantitative, empirical perspective it would be interesting to see if changes in the behaviour of market participants can be identified from pure trading data, thus creating another source for respective intelligence and contributing to service and product enhancements for particular groups of market participants or the overall market.

Generally, in a deductive approach to study relevant trading behaviour, market participants would be allocated into groups along a variety of characteristics deemed relevant, e.g. trading motivation, kind of market access (direct/indirect), or degree of technical assistance involved in trading decisions. However, such characteristics are usually not completely known, even not for the market operator. In particular, it is a typical problem for academic research, since accessible data sets are often thinned out in one way or another.

As a solution, this paper proposes an inductive approach where a single measure is derived from trading data as a representation of selected behavioural trading patterns of market participants. For this purpose, the analysed measure is built directly from the traders' trading decisions. An appropriate indicator should be easy to understand and easy to operationally implement. Therefore, we propose to use the Shannon (1948) entropy as known from information theory for construction of such an indicator. The entropy assesses the degree of structure in a given message or – in our application – a sequence of trading decisions. The underlying assumption of this approach is that “structural” differences between individual trading strategies (although each strategy is unknown to the market operator) will be reflected in the “degree of structure” of the respective sequences of trading decisions.

In a first, inductive step we apply the Shannon concept to empirical data from Deutsche Börse's electronic trading system for cash equities, Xetra, and then, in a second, deductive step try to verify the degree of its variation across different groups of market participants. For illustration purposes we select two groups of market participants by a self-explanatory, readily available characteristic, namely the fact whether a given trader has registered for Deutsche Börse's Automated Trading

Program (ATP) or not. Thus, in essence we analyze the variation in entropy across algorithmic and non-algorithmic traders on Xetra¹. To our knowledge, this approach is the first to combine entropy calculation of trading decisions with grouping of market participants.

The results reveal pronounced differences between algorithmic and non-algorithmic traders. In particular, trading patterns of algorithmic traders exhibit a medium degree of regularity while non-algorithmic trading tends towards either very regular or very irregular trading patterns.

The specific empirical example chosen automatically links this paper with the substantial amount of literature on algorithmic trading in financial markets. Thus, section 2 briefly discusses the phenomenon of algorithmic trading in financial markets and in academic literature. Section 3 provides details on the data set and the methodology applied, including a short review of the entropy measure applied in this paper. Section 4 presents and discusses the empirical results. Section 5 concludes the discussion by summarizing and outlining potential next steps. The appendix gives some information on Deutsche Börse's Algorithmic Trading Program.

2 Algorithmic trading in financial markets

Algorithmic trading by no means defines a homogeneous set of trading activities. On the contrary, algorithmic procedures vary significantly, when controlling different targets like e.g. *optimal execution* of a given order (brokerage) as opposed to detection and realization of *optimal investment* opportunities. For the purpose of this discussion we propose the following, very broad definition²: *Algorithmic trading covers all trading activities where a computer algorithm autonomously decides on certain characteristics of an order*. These characteristics include but are not limited to the instrument(s) to be traded, order limit, order volume, timing of order insertion, and choice of execution venue. However, an algorithm does not need to decide on *all* of these characteristics in order to represent algorithmic trading. Key to algorithmic trading, however, is the usage of strategies implemented on computer systems and the fact that the computer represents a driving element in order placement. This perception differentiates algorithmic trading from computer assisted trading and has been considered in the requirements laid out for Xetra customers to participate in Deutsche Börse's Automated Trading Program (Deutsche Börse (2008), cf. Appendix).

The above definition encompasses – but is not limited to – the following practices:

- Automated identification of investment opportunities (e.g. through statistical arbitrage);

¹ We choose the differentiation between algorithmic and non-algorithmic trading behaviour as application in this paper because this topic has been of high public and academic interest for several years now, cf. section 2.

² Gomber, Gsell (2009) discuss a variety of alternative, broad definitions.

- Activities targeting optimal placement of orders created outside of the algorithm, e.g. placed with the trading desk by an external customer, with respect to time and order volume (e.g. volume weighted average-strategies);
- Deciding optimal order placement with respect to execution venue taking into account potential execution price including implicit and explicit transaction costs (smart order routing)³.

Computer-implemented strategies as such are not a new phenomenon. Pole (2007), for example, dates the computer implementation of statistical arbitrage as a specific strategy incarnation back to the 1980s. Program trading based on portfolio insurance or index arbitrage strategies was already beyond the status of a niche market practice at the end of the same decade, and the Brady report discusses its involvement in the financial crisis of 1987 (Brady et al. (1988)).

Modern algorithmic trading, however, would not have been possible without the huge advancements in computing technology and the automation of financial markets. The progress in processing power and storage capacities has allowed the near or real time realization of trading strategies that had previously required runtimes of hours or even days. Likewise, equity and derivative markets have seen full automation, and respective technical interfaces and messaging protocols now allow direct connection of machines to the market place. System capacities are sufficient to process substantial amounts of incoming messages/orders and network technology and co-location services reduce latencies to new minimums. The technical progress of algorithmic trading and the technical progress in trading platforms form an interdependent, symbiotic relationship.

Over the course of recent years algorithmic trading has been a hot topic in academic literature and among practitioners and regulators. The discussion mainly focuses on the impact of algorithmic trading on market quality and market stability. Exemplary questions are: Is the additional order flow generated by algorithmic trading beneficial for market liquidity or does it have detrimental effects? Does algorithmic trading affect average order or trade sizes (e.g. Hatrick and Deliya (2008)). Does it have adverse effects on the stability of trading environments, for example in case of a technical failure of a single algorithm (Grant (2010)) or where similar strategies might enforce a specific market trend (e.g. Brady et. al. (1988), Khandani and Lo (2007)). Does algorithmic trading require special regulatory attention (e.g. Mackenzie and Chung (2010), Mehta (2010)).

However, inside-out empirical investigation into such issues is severely limited as algorithmic trading can seldom be identified. Of course, the ideal situation for empirical research would be if a market operator would require traders to register as algorithmic traders, and, until recently, Deutsche Börse indeed required such registration to allow respective traders to benefit from a special fee schedule, the Automated Trader Program (ATP) (Deutsche Börse (2009)). While registration was

³ It might be argued whether a smart order router should be considered a trading algorithm or not. While the choice of *when* and *how* to trade is typically considered as an algorithm's competence, a smart order router is originally designed to decide on *where* to execute an order. However, merging of the *when*, *how*, and *where* competences into one entity is an obvious step of evolution, so smart order routers and algorithms cannot strictly be distinguished anymore.

not compulsory, the fee rebate is considered as a strong incentive for it. Gomber and Gsell (2009) and Hendershott and Riordan (2009) are able to base their studies on this information. The appendix gives an overview of Deutsche Börse's ATP.

Without an appropriate indicator for algorithmic trading, empirical studies have to rely on more or less adequate proxies. Hendershott et al. (2008), for example, use electronic message traffic at the NYSE as a proxy for algorithmic trading, while Chaboud et al. (2009) have access to the information whether FX orders have been placed through EBS' automated interface or not.

Taking a different perspective, i.e. an outside-in view, Prialé et al. (2007) use order lifetimes to analyse order patterns and find traces of algorithmic trading arguing that certain order activities are more likely to result from algorithmic trading.

In a certain way our paper complements this approach: We propose a generalized measure for order patterns and ex post verify its variation across algorithmic and non-algorithmic traders using the classification obtained through Deutsche Börse's ATP. Thus, our paper also represents an important link between the strands of literature mentioned.

3 Data and method

3.1 Data source and market environment

The data source used in this analysis contains intraday, ultra-high frequency data in DAX30 constituents on Deutsche Börse's electronic trading system for cash equities, Xetra. The period of analysis is the entire week from October 8th to 12th, 2007. This week was chosen because it represents a quite "normal" week without any significant market events. The period lies before any of the subsequent crisis periods (starting in January 2008) and DAX index levels were still around 8.000 points as depicted in Figure 1. The ratio between the number of buy and sell order insertions ("Buy-sell ratio") on instrument level outlined in Table 1 further support this notion of a "normal" week as the ratios exhibit a rather normal fluctuation around 1.

On the level of the individual order the data set contains information on the instrument traded, the member and specific trader, order type (e.g. buy or sell order, limit or market order), order limit, order quantity, the account (i.e. whether the order has been placed in an agent or proprietary capacity), the order event (entry, deletion, modification, execution), the exact time of the event (in 1/100th of a second) and execution price (if any).

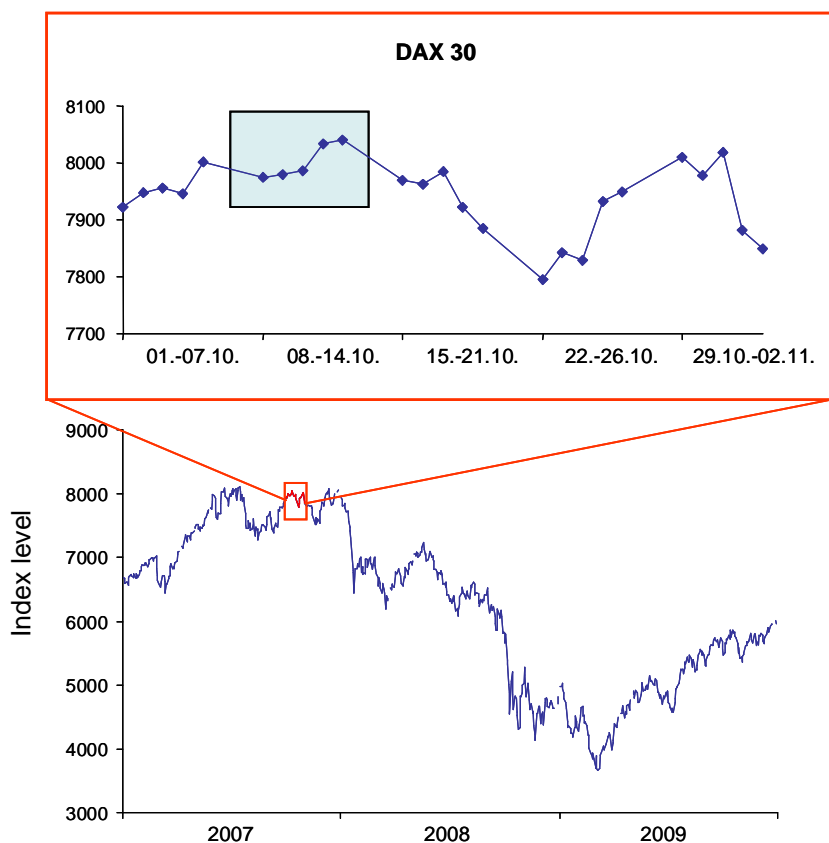


Figure 1: DAX during analysed week of Oct. 08-12, 2007 (box in top figure) and embedding into long-range market dynamics (bottom): the chosen period is relatively “calm” without significant market events.

Code	08.10.2007			09.10.2007			10.10.2007			11.10.2007			12.10.2007		
	Buy	Sell	Buy-Sell Ratio	Buy	Sell	Buy-Sell Ratio	Buy	Sell	Buy-Sell Ratio	Buy	Sell	Buy-Sell Ratio	Buy	Sell	Buy-Sell Ratio
ADS	53%	47%	1.12	52%	48%	1.08	46%	54%	0.85	49%	51%	0.96	51%	49%	1.06
ALV	51%	49%	1.06	50%	50%	0.99	46%	54%	0.84	48%	52%	0.94	48%	52%	0.94
BAS	48%	52%	0.93	50%	50%	0.99	51%	49%	1.05	49%	51%	0.96	48%	52%	0.94
BAY	52%	48%	1.09	48%	52%	0.93	53%	47%	1.11	49%	51%	0.97	48%	52%	0.92
BMW	50%	50%	1.00	50%	50%	0.99	48%	52%	0.92	50%	50%	1.02	50%	50%	0.98
CBK	55%	45%	1.23	47%	53%	0.90	48%	52%	0.91	49%	51%	0.98	47%	53%	0.89
CON	55%	45%	1.21	53%	47%	1.12	43%	57%	0.76	45%	55%	0.82	44%	56%	0.78
DAI	55%	45%	1.21	50%	50%	0.99	49%	51%	0.97	48%	52%	0.92	46%	54%	0.87
DB1	53%	47%	1.14	52%	48%	1.10	52%	48%	1.10	52%	48%	1.10	46%	54%	0.86
DBK	49%	51%	0.98	49%	51%	0.97	50%	50%	0.98	50%	50%	1.00	47%	53%	0.88
DPB	50%	50%	0.99	51%	49%	1.04	48%	52%	0.93	49%	51%	0.98	46%	54%	0.87
DPW	44%	56%	0.80	52%	48%	1.07	47%	53%	0.90	50%	50%	1.01	46%	54%	0.84
DTE	51%	49%	1.05	50%	50%	1.00	49%	51%	0.96	50%	50%	1.00	54%	46%	1.18
EOA	46%	54%	0.85	52%	48%	1.09	50%	50%	1.01	52%	48%	1.06	49%	51%	0.97
FME	53%	47%	1.12	51%	49%	1.03	40%	60%	0.67	46%	54%	0.86	49%	51%	0.98
HEN3	48%	52%	0.91	54%	46%	1.16	45%	55%	0.83	47%	53%	0.89	40%	60%	0.66
HRX	55%	45%	1.22	50%	50%	0.98	45%	55%	0.83	46%	54%	0.86	49%	51%	0.95
IFX	42%	58%	0.73	45%	55%	0.81	46%	54%	0.85	51%	49%	1.02	52%	48%	1.09
LHA	49%	51%	0.98	49%	51%	0.95	46%	54%	0.84	48%	52%	0.92	49%	51%	0.95
LIN	54%	46%	1.17	47%	53%	0.90	44%	56%	0.80	46%	54%	0.84	47%	53%	0.88
MAN	53%	47%	1.14	53%	47%	1.12	49%	51%	0.98	45%	55%	0.82	43%	57%	0.74
MEQ	54%	46%	1.17	51%	49%	1.06	47%	53%	0.89	48%	52%	0.91	41%	59%	0.69
MRK	46%	54%	0.87	56%	44%	1.25	47%	53%	0.87	46%	54%	0.84	46%	54%	0.86
MUV2	48%	52%	0.93	46%	54%	0.86	48%	52%	0.94	49%	51%	0.96	45%	55%	0.83
RWE	50%	50%	1.00	50%	50%	0.98	45%	55%	0.81	52%	48%	1.07	48%	52%	0.94
SAP	55%	45%	1.20	50%	50%	1.02	49%	51%	0.95	50%	50%	0.99	50%	50%	0.98
SIE	51%	49%	1.04	52%	48%	1.08	48%	52%	0.94	51%	49%	1.04	48%	52%	0.93
TKA	50%	50%	0.98	49%	51%	0.96	48%	52%	0.92	49%	51%	0.97	47%	53%	0.88
TUI1	46%	54%	0.87	51%	49%	1.04	42%	58%	0.74	44%	56%	0.78	43%	57%	0.74
VOW	48%	52%	0.93	45%	55%	0.83	47%	53%	0.90	50%	50%	0.98	47%	53%	0.90
Total	51%	49%	1.03	50%	50%	0.99	48%	52%	0.91	49%	51%	0.96	47%	53%	0.90

Table 1: Buy-Sell ratio of number of order insertions per instrument in DAX30, October 8th -12th, 2007

In addition, we have information on the ATP status of each trader, i.e. whether a given trader has registered for Deutsche Börse’s Automated Trading Program (ATP) or not. This registration is used in the following to distinguish between algorithmic and non-algorithmic traders⁴.

The analysis focuses on order insertions. Order insertions are our preferred mirror of a trader’s market assessment. They are under full control of a trader and therefore a direct expression of a trader’s strategy. While the same holds for order modifications and deletions, these occur less frequently. Order executions are beyond direct control of a trader.

Overall descriptives of our sample are shown in Table 2. In the trading week analysed we observe 4.2 million order insertions. 52% of these order insertions result from algorithmic (ATP) traders. By Euro volume algorithmic traders had a market share of 39% in 2007 which was up from only 15% in 2003. The higher share of order insertions reflects the higher activity of algorithmic traders. Proprietary trading was responsible for 61% of order insertions in our sample.

Date	Account		ATP Flag		Number of order insertions
	Prop.	Agent	ATP	non-ATP	
08.10.2007	59%	41%	49%	51%	703.411
09.10.2007	60%	40%	51%	49%	785.878
10.10.2007	59%	41%	52%	48%	802.595
11.10.2007	61%	39%	52%	48%	1.042.142
12.10.2007	65%	35%	57%	43%	877.868
Grand Total	61%	39%	52%	48%	4.211.894

Table 2: Overall descriptives of the analysed sample (order insertions in DAX30 instruments)

3.2 Data processing and analysis

As an easy but still authoritative representation of a trader’s (unknown!) trading strategy we reduce the flow of his order insertions to a binary sequence of buy (“1”) or sell (“0”) bits, i.e. we strip down all complexity of an order flow down to its most fundamental characteristic – that is: *whether market expectation is up or down*. Per instrument and trader we analyse intraday sequences, i.e. individual sequences do not contain after-trading hours. It is safe to assume that trading strategies are reset overnight e.g. because of news arrival.

⁴ Since one particular species of algorithmic trading, namely “high-frequency trading” (HFT) has become very prominent in recent years, one might suggest to use the order insertion frequency of a trader as straight-forward proxy for identification of algorithmic trading. However, HFT is just one component of algorithmic trading among others. A significant part of order executions however, is triggered by another kind of algorithms that follow a “slicing & dicing” strategy, trying to lower market impact by splitting huge orders into small ones that are “melted” into the market gradually over time. Such activity does not show up by extreme order insertion frequency (actually it is set to avoid any attention at all). On the other hand, simple order routing activity that collects orders of many clients and routes them to various execution venues might exhibit high frequency without being algorithmic.

As the following examples show, a certain strategy can potentially be associated with its binary sequence:

- A “slicing & dicing” execution algorithm that seeks to acquire 100.000 stocks by submitting 500 orders with an average order size of 200 over a longer period of time will correspond to a simple sequence of 500 ones: 11111.....111111.
- A “day trader” following a certain (unknown) algorithm that tells him when to buy and when to sell might produce a “semi-rhythmic” sequence like this: 111001100011110000...
- An “order router” that mixes individual client orders (probably with partial internalisation) and submits some kind of “summary” orders to the exchange might produce a more or less disordered sequence like this: 1110100010001001100111110010111000011...

In order to detect any ordered or disordered patterns within a given sequence, we analyse the distribution of all subsets of *pattern length* N that occur within a sliding window which is moved along the intraday sequence. For a specific choice of N there are then 2^N possible combinations of buy-sell-insertions or patterns and the histogram of *realised* patterns among these *possible* combinations will represent some kind of “fingerprint” of the underlying strategy – even though the latter may be completely unknown. For instance, for pattern length $N=4$ the histogram of the above example of a “slicing & dicing” algorithm would contain only one pattern spike at “1111” while the histogram of the depicted “day trader” would show a preference of patterns like “0000”, “1111”, followed by “1110”, “0111”, “1100”, “1000”, “0001”, etc. but less frequently patterns like “0101” or “1010”. The histogram of the depicted “order router” would be “flat”, i.e. no preferred patterns.

Thus, perceptibility of trading patterns significantly depends on the optimal value of N . On the one hand, if the pattern length is too small, longer patterns might not be recognized. On the other hand, case numbers are exponentially shrinking with increasing N , again inhibiting proper identification of patterns. Therefore, our analysis is conducted for a range of different values of N .

3.3 The Shannon entropy

So far, we have reduced the (potentially) complex – and unknown – trading strategy of a trader within an instrument to a binary sequence that is still representative due to reduction to its core, i.e. its “buy/sell rhythm”. Now, we want to find a quantitative measure for it that expresses how structured the sequence and therefore its underlying trading activity is. For this purpose, we use the Shannon entropy.

Shannon (1948) created a concept now known as “information theory” and introduced a measure for the uncertainty in a message which he called “entropy”. Given an alphabet Z of m symbols and a sequence of these symbols, then each symbol occurs with a certain probability p_i . The *information* that each symbol contains is defined as

$I(p_i) = -\log_2(p_i)$. Then, the expectation value of information is the weighted sum across all m symbols and defined as *entropy*:

$$H_1 = -\sum_{i=1}^m p_i \log_2 p_i \quad (1).$$

This is in analogy to the physical meaning of “entropy” in thermodynamics.

The minimum possible entropy of a sequence is 0 and occurs if the analysed sequence is constant and consists only of one symbol, leading to one $p_i = 1$ and all other $p_{i \neq j} = 0$. With $\lim_{p_i \rightarrow 0} p_i \cdot \log_2 p_i = 0$ it follows that $\min(H_1) = 0$.

On the other hand, the maximum possible entropy occurs in case of a uniform distribution, i.e. if all m sequences are of equal probability $p_i = 1/m$. Then,

$$H_{\max} = -\sum_{i=1}^m \frac{1}{m} \log_2 \frac{1}{m} = \log_2 m \quad (2).$$

A normalised entropy follows⁵:

$$H_{\text{norm}} = \frac{H_1}{H_{\max}} = -\sum_{i=1}^m p_i \cdot \frac{\log_2 p_i}{\log_2 m} = -\sum_{i=1}^m \log_m p_i \quad (3),$$

which can vary between 0 (constant, fully predictable sequence) and 1 (completely irregular, unpredictable sequence).

In our case, the alphabet Z consists of the 2^N possible binary patterns of length N . For each sequence, we calculate the histogram of all patterns as they occur in the complete sequence. If the sequence is sufficiently long⁶, we estimate the probability of each pattern from its absolute frequency, derive the respective Shannon entropy, and normalise it finally. In the following, we use the notation “entropy” in the meaning of “normalised entropy”.

4 Empirical results

4.1 Entropy distributions for ATP and non-ATP traders

The following figures show the distribution of normalized entropy for ATP traders (bottom row in each figure) and non-ATP traders (top row in each figure) across pattern lengths 2 to 10 for an exemplary day, the Tuesday of the analysed week, i.e.

⁵ Shannon (1948) defines this as “relative entropy”.

⁶ Arbitrarily, we define “sufficiently” in our case as a sequence length long enough to expect 2 counts per possible binary pattern, leading to a minimum of 2^{N+1} order insertions. By this convention, we have to be aware that the set of analysed sequences varies with the choice of the pattern length N .

October 9th, 2007. The line in each graph corresponds to the right axes and shows the cumulative absolute frequency of trader-instrument combinations analysed for the generation of the histogram. The left axes show the absolute count for each bin of the histogram.

Entropy distributions for pattern lengths N=2...4, Oct.9th, 2007

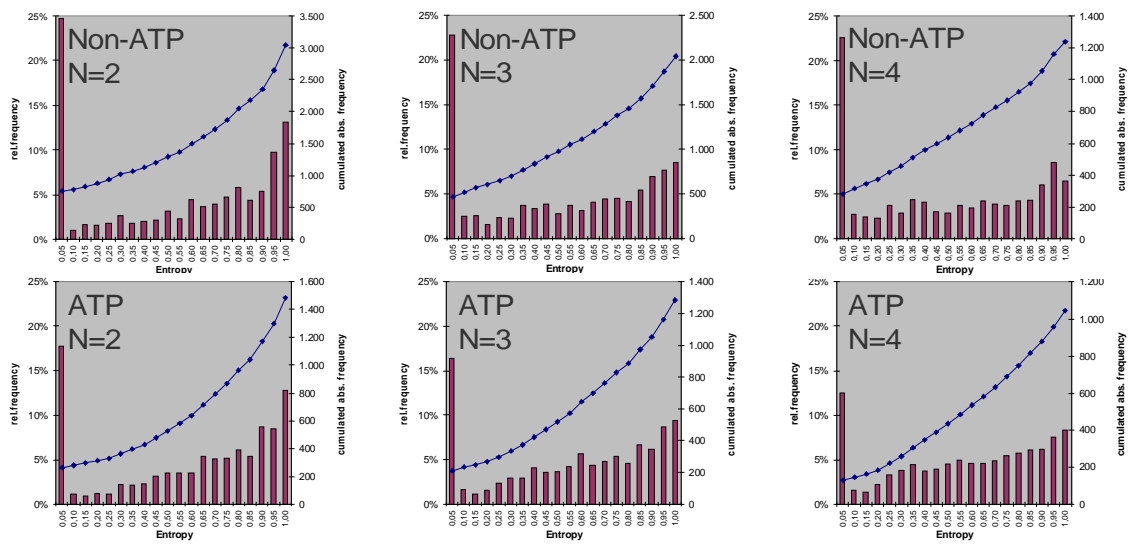


Figure 2: Distribution of normalized entropy for pattern lengths 2 to 4. ATP and non-ATP trading behaviour shows a peak at entropy close to 0 which corresponds to a constant sequence of only zeros (only sell orders) or only ones (only buy orders). This effect is more pronounced for non-ATP traders.

Entropy distributions for pattern lengths N=5...7, Oct.9th, 2007

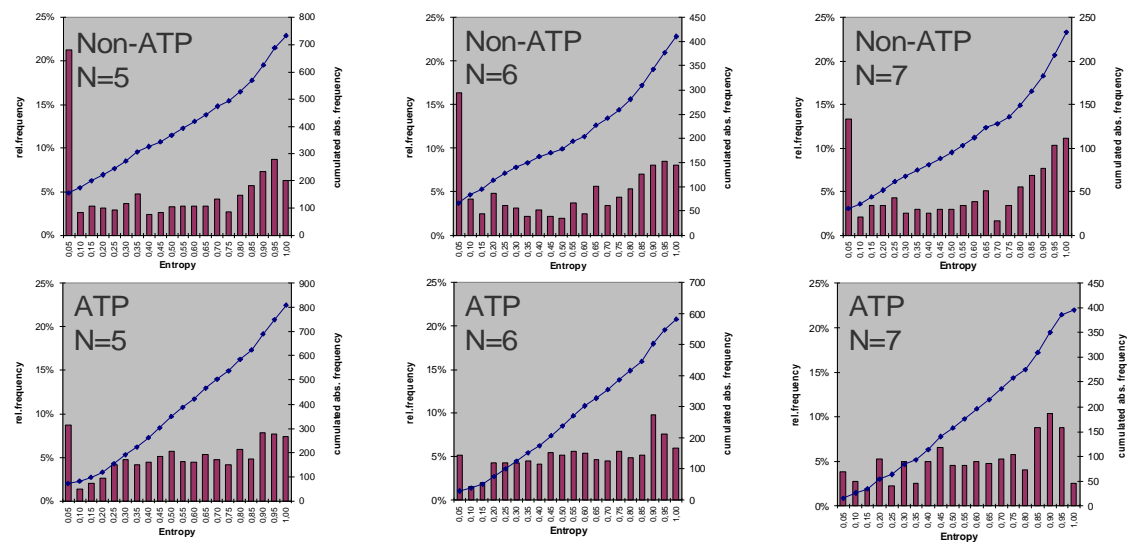


Figure 3: With increasing pattern length N the peak at low very entropies declines and has vanished at $N=7$ for ATP traders.

Entropy distributions for pattern lengths $N=8\dots 10$, Oct.9th, 2007

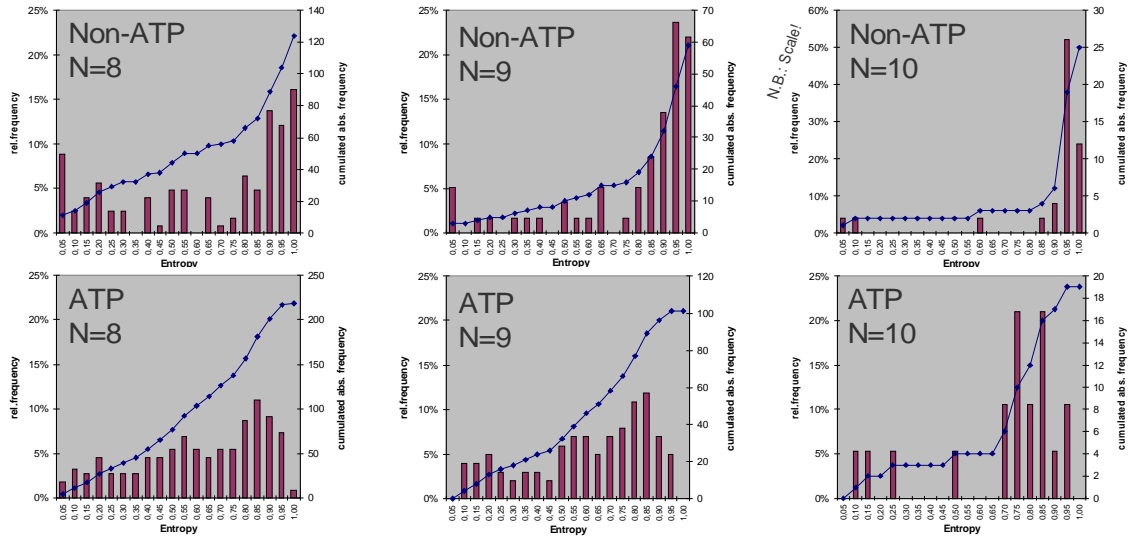


Figure 4: With even longer pattern length, the peak at very low entropies vanishes for non-ATP traders as well and tendency towards higher entropies becomes visible. For non-ATP traders this tendency is towards maximum entropy. The corresponding maximum for ATP traders seems to occur at high but not maximum entropies. From $N=10$ on, the available sample size of trading sequences long enough for a reliable estimation of the probability distribution of patterns becomes too small for a meaningful entropy calculation and therefore the distribution of entropies is thinned out.

The results exhibit a significant peak at very low entropies that declines with increasing pattern length N . The peak is generally more pronounced for non-ATP traders but, for $N < 7$, it is also clearly visible for ATP traders.

This low entropy indicates a very ordered underlying sequence with significant short-term autocorrelation. Closer inspection of the orders shows that the result is due to long periods of consecutive buy or sell order insertions within a trader’s order stream. These periods lead to a histogram of binary patterns that is far away from a uniform distribution and expresses some distinctive peaks. Such a histogram corresponds to an extremely low entropy.

The probability of observing such short time correlation is higher for shorter than for longer pattern lengths N which explains the observed decline of the respective peak with increasing N .

At pattern length $N=9$ the difference between ATP and non-ATP traders is most pronounced. In general, entropy distributions of ATP and non-ATP traders become more distinctive with increasing N . This is due to the fact that with increasing pattern length N more complex patterns can be analysed. Therefore, underlying regularities can be more clearly identified and distinguished from “irregular” patterns.

Algorithmic traders exhibit a lower entropy resulting from more pronounced patterns, i.e. a higher degree of regularity, which is the result of their stricter, rule based trading approach. In contrast, non-algorithmic traders have a less rule based trading approach, their trading behaviour exhibits more irregular patterns and, thus, higher entropy.

The sample size represents an important limitation: The number of patterns increases exponentially with N and the respective number of observations declines. In our analysis, the resulting sample of order streams long enough to be evaluated has sufficient size only up to about $N=9$. For $N=9$ we are able to observe 59 non-ATP order streams and 101 ATP order streams. For $N=10$ these numbers reduce to 25 non-ATP order streams and 19 ATP order streams, respectively.

4.2 Extension of analysis to a full week

In order to analyse the persistency of findings, the entropy analysis is performed for each day of a full week (Oct. 8th-12th, 2007): Per day and per pattern length the average entropy of order insertion sequences in an instrument is calculated and compared between ATP and non-ATP traders. The results shown in Figure 5 reveal consistent findings with section 4.1 and show no significant day-to-day variation of entropy distributions – neither for the ATP nor for the non-ATP group. Though, the strong difference between ATP- and non-ATP persists.

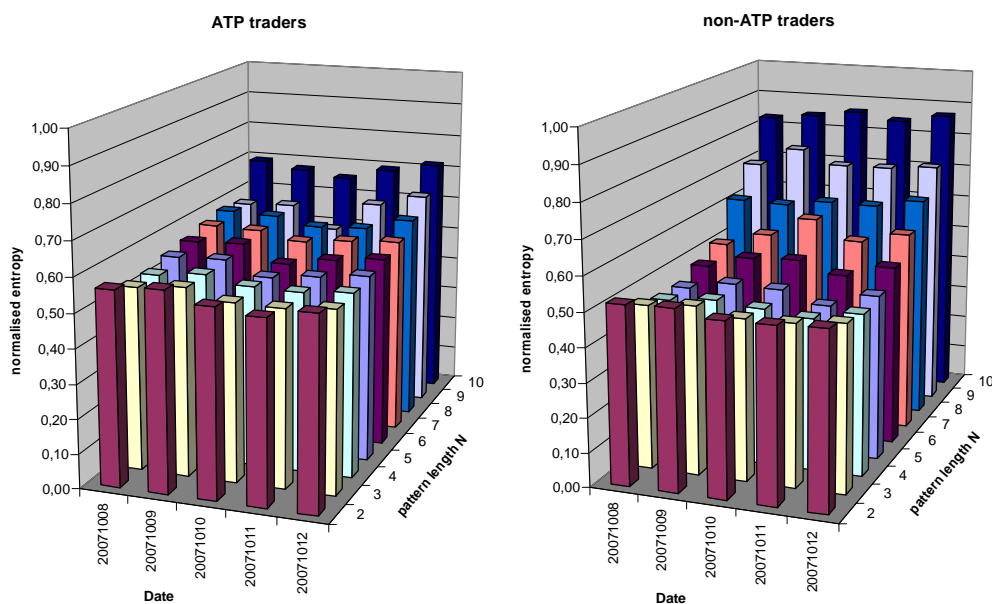


Figure 5: Comparison of the average entropy across all ATP resp. non-ATP traders for several pattern lengths across a full week. No significant day-to-day variation but strong difference between the group of ATP and that of non-ATP traders.

Since inter-day results do not vary significantly, for a detailed comparison between the entropy distributions of ATP and non-ATP trading the analysed sample is extended to the full week. This way, effects specific to a certain trading day can average out.

Figure 6 depicts the resulting entropy distributions of the ATP group (solid lines) and the non-ATP group (dashed lines) in several percentiles (0.25, 0.5, 0.75) of each distribution. While the distributions significantly overlap for pattern lengths N up to 7,

they can be much better distinguished for $N=8$ and $N=9$. For even larger N , the sample size is not sufficient for a proper estimation of entropies.

We conclude that in our sample the pattern length $N=9$ is an appropriate choice for further analysis and keep this fixed in the following.

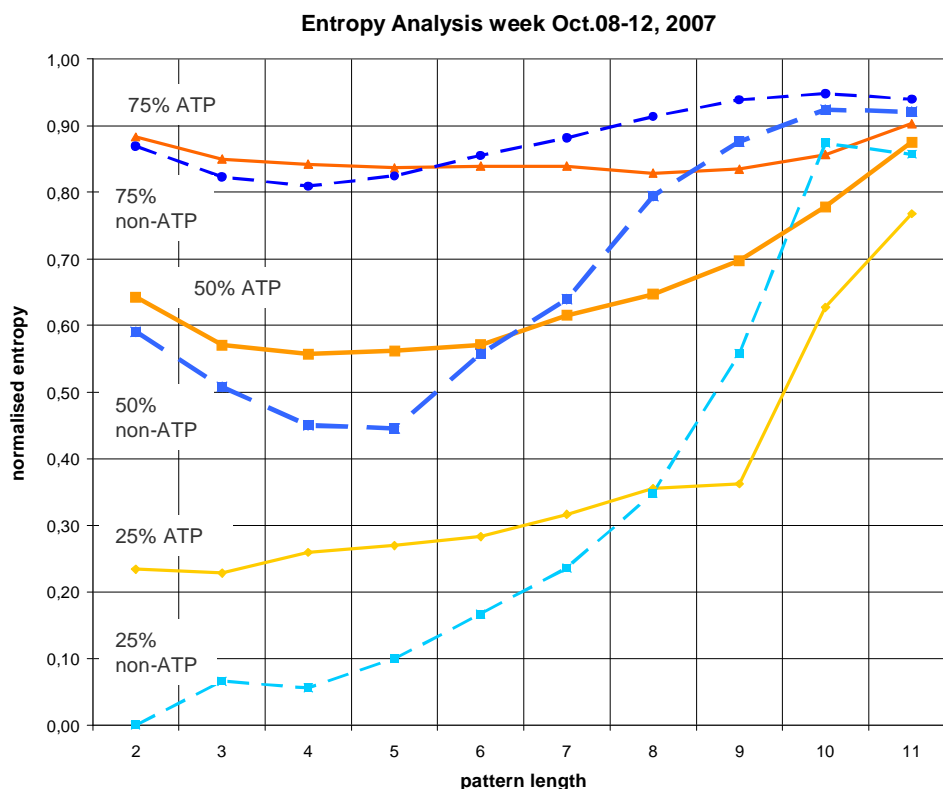


Figure 6: Comparison of entropy distributions based on data of a full week. Distributions are characterized by their quartiles. One can clearly distinguish between the distributions of ATP (percentiles in solid lines) and non-ATP trading (percentiles in dashed lines), in particular from $N=8$ on when the 75% percentile of ATP group is close to the median of the non-ATP group. For $N=10$ separation is almost complete but the underlying sample size is unreliable (compare Figure 4).

4.3 Predictive power of the ATP-flag

In the previous sections, we have analysed the entropy of a given sequence and aggregated the result to distributions that were finally related to the ATP, respectively non-ATP nature of the trader.

In this section, we switch perspective by analysing what conclusion on the ATP flag we can draw from a given entropy. In other words: for an unknown, randomly drawn order sequence, we want to estimate whether it originates from an ATP or non-ATP trader based on the entropy of that particular sequence.

For this purpose, we focus on that “active” sample of all trader-instrument combinations that consists of enough order insertions to calculate a meaningful⁷ normalised entropy for pattern length $N=9$. For each sequence we calculate its entropy and determine the absolute and relative number of ATP and non-ATP cases within entropy bins. Figure 7 shows the resulting histogram after averaging of daily findings over all trading days of October 2007.

One can derive the probability to find ATP or non-ATP trading for a given entropy of the underlying sequence. For instance, in Figure 7, the rightmost bin that covers very high entropies above 0.95 is made up of 80% non-ATP and only 20% ATP trading. For entropies between 0.70 and 0.75 the situation is the other way around: such “medium entropy” indicates with more than 90% ATP trading as origin of the underlying buy-sell order insertion sequence. For comparison, the overall ratio of ATP and non-ATP trading of 71.3% is depicted by the horizontal line. It corresponds to 3378 ATP and 1359 non-ATP trader-instrument combinations in the underlying “active” data set of October 2007.

For most entropy values there exists a clear distinction between ATP and non-ATP trading – in absolute figures as well as in deviation from the overall ATP ratio. As a tendency, one can derive the observation that non-ATP trading follows either a very regular (=low entropy) or very irregular order insertion pattern while ATP trading concentrates at medium to high entropies.

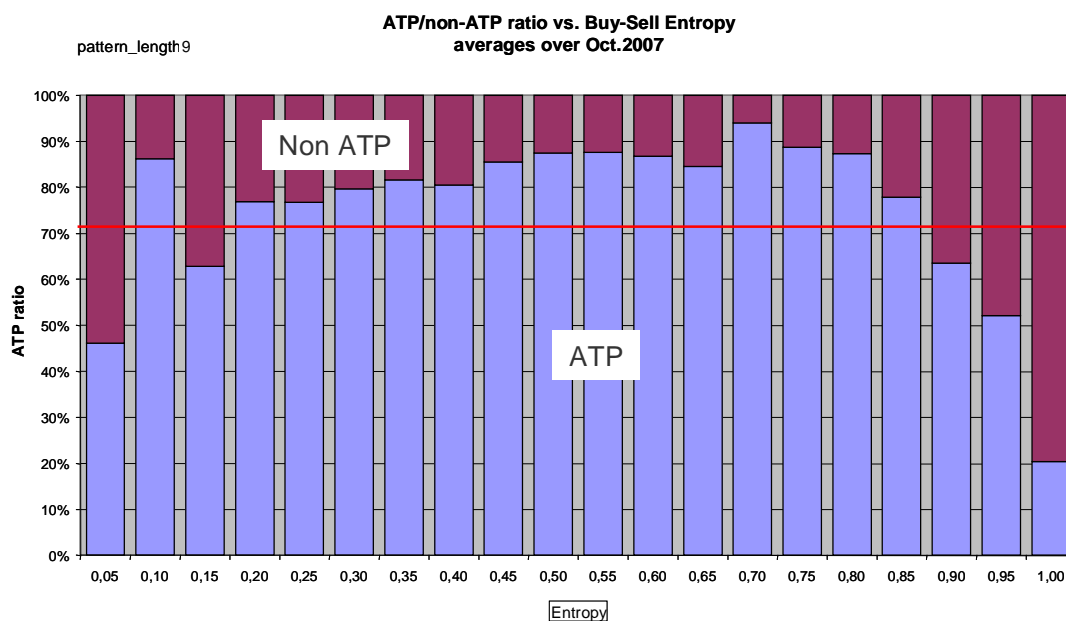


Figure 7: Dependence of the distribution of ATP resp. non-ATP share within the active sample on the Buy-Sell entropy for pattern length $N=9$. The horizontal line depicts the overall ratio between ATP and non-ATP trader-instrument combinations in the “active” data set in October 2009 without taking entropy into account (71.3% ATP). For most entropy values there exists a clear distinction between the two groups. This can potentially be used for characterisation of “unknown” trading behaviour.

⁷ In accordance with 4.1 we have defined “meaningful” as more than 2 times 2^N daily order insertions (in the displayed case with $N=9$, this are 1024)

4.4 Introducing account information

Another interesting aspect with potential influence on the trading pattern is the account information, i.e. differentiation of whether a trader acts in a proprietary or in an agency capacity.

For this purpose we have calculated separate entropy distributions of ATP and non-ATP trading for “Agent” and “Proprietary” trading accounts. Figure 8 shows the results.

Obviously, proprietary trading tends towards higher entropies, in particular do more than 50% of the non-ATP order streams cluster at entropies larger than 0.9.

Agency trading is clearly split into two regimes: algorithmic trading exhibits low entropies, non-algorithmic trading high entropies. In particular, algorithmic agency trading shows extraordinary low entropies. This is evidence for presence of execution algorithms producing structured order streams with short range bias towards “sells” or “buys”.

Non-algorithmic trading, both proprietary and agency, does show some preference for extremely low entropies below 0.1. This can potentially be explained by human traders trying to execute large orders by mostly periodic sequences of only buy or only sell orders, leading to minimal entropy.

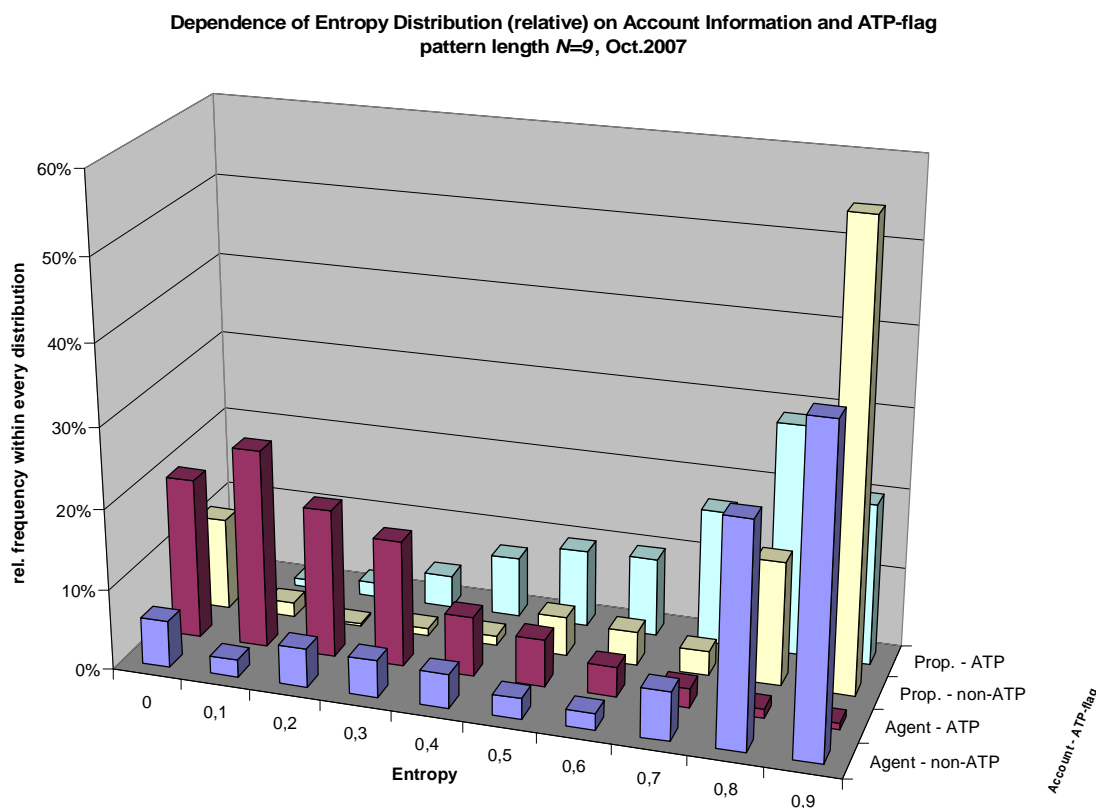


Figure 8: Entropy distribution along account and ATP information. ATP agent trading strikes out with preference of low entropies while the other 3 combinations (ATP as well as non-ATP proprietary trading and non-ATP agent trading) tend towards high entropies.

5 Summary and conclusion

This paper focuses on the question if the trading behaviour of market participants as observed on financial markets is sufficiently distinct so that differences can be identified by a single measure. Respective variations in trading behaviour might occur over time or between arbitrary subgroups of market participants. The proposed measure is the Shannon entropy.

In the specific context chosen, i.e. variations in trading behaviour between algorithmic and non-algorithmic traders in cash equity markets, the entropy indeed exhibits pronounced differences between algorithmic and non-algorithmic traders. While the degree of the effect varies with the length of the analysed trading pattern, the results show that trading patterns of algorithmic traders exhibit a medium to high degree of regularity while trading patterns of non-algorithmic traders can be associated with either very low or very high regularity. These findings are consistent with the assumption that algorithmic traders are following a stricter, rule based trading approach than non-algorithmic traders which is reflected in the entropy measure. Including account information into the analysis reveals that agent trading is made up of two separate regimes: one low-entropy algorithmic trading group and one high-entropy non-algorithmic group. For proprietary trading, however, such distinction is less obvious, though differences between the algorithmic and the non-algorithmic group can also be clearly observed.

Further analyses based on the entropy measure will focus on more complex trading patterns and different longitudinal or cross-sectional settings. In its present form, however, the paper already represents an important link between inside-out-studies that infer differences in trading behaviour from available proxy indicators for algorithmic trading and outside-in studies that identify trading patterns in general.

Clearly, the presented approach requires the ability to assign each order to the respective trader. Further trader characteristics beyond the algorithmic or non-algorithmic nature of trading decisions would be helpful for explaining any segmentation of market participants suggested by calculation of the entropy measure.

References

Brady, N.F., Cotting, J.C., Kirby, R.G., Opel, J.R., and Stein, H.M. (1988) – “Report of the presidential task force on market mechanisms”, Washington

Chaboud, A., Chiquoine, B., Hjalmarsson, E., and Vega, C. (2009) – “Rise of the machines: algorithmic trading in the foreign exchange market”, working paper, Division of International Finance at the Federal Reserve Board of Governors

Deutsche Börse (2008) – “ATP agreement for participation in the “Automated Trading Program” (ATP) via the electronic trading system Xetra”

Deutsche Börse (2009) – “Price list for the utilization of the trading system of FWB Frankfurt Stock Exchange”

Gomber P., Gsell M. (2009) – “Algorithmic trading engines versus human traders – do they behave different in securities markets?”, working paper No. 2009/10, Center for Financial Studies

Grant, J. (2010) – “Computer-driven trading raises meltdown fears”, Financial Times, published on ft.com, January 25th, 2010

Hatrick, K., Deliya, D. (2008) – “Seasonality, microstructure and market evolution: an algorithmic perspective”, working paper, Deutsche Bank AG

Hendershott, T., Jones, C., Menkveld, A. (2009) – “Does Algorithmic Trading Improve Liquidity?”, Journal of Finance, Forthcoming; WFA 2008 Paper.

Hendershott, T., Riordan, R., (2009) – “Algorithmic Trading and Information”, NET Institute Working Paper No. 09-08.

Khandani, A.E., Lo, A.W. (2007) – “What happened to the quants in August 2007?”, working paper, MIT

Mackenzie, M., Chung, J. (2010) – “High frequency trading under SEC scrutiny”, Financial Times, published on ft.com, January 12th, 2010

Mehta, N. (2010) – “High-frequency trade rebates questioned in SEC review (Update 2)”, Business Week, published on businessweek.com, January 22nd, 2010

Pole, A. (2007) – “Statistical arbitrage: algorithmic trading insights and techniques”, Hoboken

Prix, J., Loistl, O., Huetl, M. (2007) – “Algorithmic Trading Patterns in Xetra Orders”, The European Journal of Finance, Vol. 13, No. 8, 717–739

Shannon C.E. (1948) – “A mathematical theory of communication”, Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656

Appendix: Automated Trading Program of Deutsche Börse

The Automated Trading Program (ATP) of Deutsche Börse has been a program to register particular Trader IDs, allowing respective traders to benefit from a special fee schedule (Deutsche Börse (2009)). While registration was not compulsory, the fee rebate was considered as a strong incentive for registering.

In order to qualify the trader also had to fulfil specific requirements⁸ such as:

1. ATP transactions have to be generated by an electronic system which has to determine two out of the three following order parameters: price (order type and/or order limit where applicable), timing (time of order entry) and quantity (quantity of the order in number of securities).
2. The electronic system that generates the ATP transactions must generate buy or sell orders independently, i.e. without frequent manual intervention, using a specified program and data.
3. ATP transactions must be channelled into the electronic trading system Xetra directly without further manual intervention using an ATP User-ID.
4. The exchange fees or the fees charged by the ATP member to its clients have to be directly or indirectly considered by the electronic system when determining the order parameters.

The ATP has been replaced by a new pricing model effective November 2nd, 2009.

⁸ For details, see Deutsche Börse (2008).

CFS Working Paper Series:

No.	Author(s)	Title
2010/11	Annamaria Lusardi Olivia S. Mitchell	How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy and Retirement Readiness
2010/10	Annamaria Lusardi Daniel Schneider Peter Tufano	The Economic Crisis and Medical Care Usage
2010/09	Annamaria Lusardi Olivia S. Mitchell Vilsa Curto	Financial Literacy among the Young: Evidence and Implications for Consumer Polic
2010/08	Volker Wieland Maik Wolters	The Diversity of Forecasts from Macroeconomic Models of the U.S. Economy
2010/07	Antje Brunner Jan Pieter Krahen	Hold-Up in Multiple Banking: Evidence from SME Lending
2010/06	Roman Kräussl André Lucas David R. Rijsbergen, Pieter Jelle van der Sluis Evert B. Vrugt	Washington Meets Wall Street: A Closer Examination of the Presidential Cycle Puzzle
2010/05	Sander J.J. Konijn Roman Kräussl Andre Lucas	Blockholder Dispersion and Firm Value
2010/04	Narasimhan Jegadeesh Roman Kräussl Joshua Pollet	Risk and Expected Returns of Private Equity Investments: Evidence Based on Market Prices
2010/03	Roman Kraeussl Christian Wiehenkamp	A Call on Art Investments

Copies of working papers can be downloaded at <http://www.ifk-cfs.de>