The rise of Algorithmic Trading and its effects on
Return Dispersion and Market Predictability

Master Thesis
Finance

Name: Rick Verheggen
Student Number: u1257435
Student E-mail:
Personal E-mail:
Date: November, 2017
Supervisor: dr. R.G.P. Frehen
Second Reader: dr. F. Castiglionesi

Abstract

A revolution is happening within the financial markets as trading algorithms are executing the grand majority of all trades. Moreover, computers are substituting human traders as well as the emotion involved in their trading. Since trading algorithms are not subject to emotion which is known to cause market inefficiencies, markets are thought to have become more efficient. Additionally, as fewer human traders are active within the market fewer predictable biases apply that are known within behavioral finance and thus is expected that the market has become less predictable. This study was designed to determine the effects of algorithmic trading on dispersion and forecast accuracy. Dispersion is measured through idiosyncratic volatility and tested against algorithmic trading, and by measuring the prediction error of the remaining human traders on the market it is tested to see if analysts' predictions have indeed become less accurate with the rise of algorithmic trading. Instead, this research finds that increased algorithmic trading has led analysts to make more accurate forecasts conjointly with a reduction in dispersion. Moreover, these findings contribute to the limited knowledge on the effects that algorithmic trading and automation pose on the financial markets.

Keywords: *Algorithmic Trading, Dispersion, Prediction Error, Forecast Accuracy, Automated Trading, Fintech*

Table of Contents

Introduction

Technology is becoming more and more of an influential factor in the lives of ordinary people around the globe and the internet has expanded in such a way that living without it, has in some countries even become impossible. Currently, Artificial Intelligence and Quantum Computing are on the verge of breakthrough and could potentially become as influential in society as the internet has become in our current daily lives. Correspondingly, in the world of finance, the rise of the internet and its subsequent technological developments are greatly impacting financial markets. For instance, transactions have become electronic and the time that it takes to execute a trade has decreased to milliseconds, and even nanoseconds. In addition, a new custom-built chip which is able to execute trades within 740 nanoseconds is being launched by Fitnetix, a UK based company. According to Johnson et al. (2012), this technological race is likely to be pushed further until the physical limits of the speed of light are met.

Amongst these technological developments in the financial markets, automated trading might be the most present-day and prominent revolution. An algorithm can be defined as a precise plan of steps that uses computations to transform the input values into an output value (Leshik & Cralle, 2011). Supply and demand on the stock markets are increasingly in the hands of these computational algorithms that fully autonomously decide to buy or sell a stock on the behalf of its "owner". As presented in *Figure* 1 by Glantz & Kissel (2013, p. 258), the percentage of market volume that can be attributed to algorithmic trading has risen greatly in the past twenty years with asset managers, high frequency traders and hedge funds accounting for most of the volume (Glantz & Kissel, 2013). Our proxy for algorithmic trading based on CRSP data support findings and also shows a clear rise in algorithmic trading activity as can be observed in *Figure 2*.

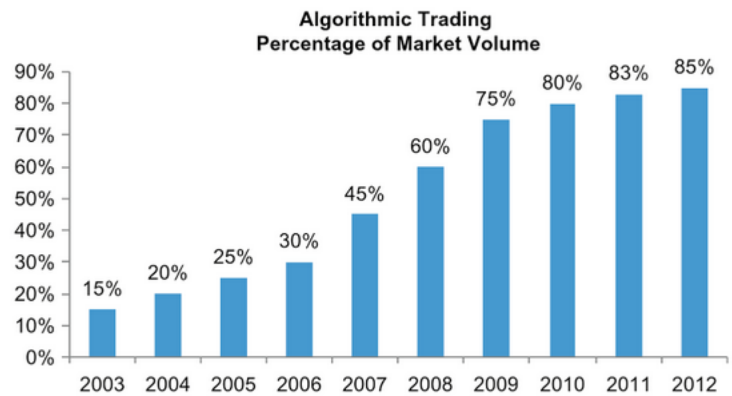

Figure 1. Algorithmic trading as a percentage of market volume. Reprinted from: Multi-asset risk modeling: techniques for a global economy in an electronic and algorithmic trading era, by M. Glantz, & R. Kissel, 2013, p. 258, Copyright by Academic Press.

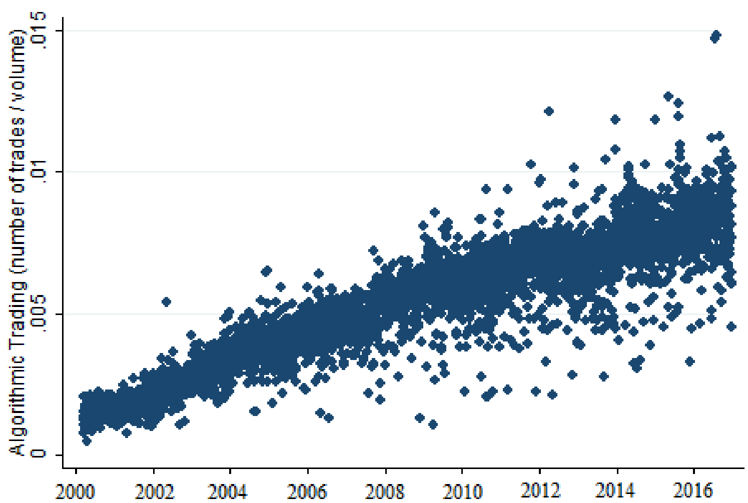

Figure 2. Proxy for Algorithmic Trading based on CRSP data

Nevertheless, algorithmic trading is still a new topic and even though its foundation can be traced back to 1949 it has only become widely spread in the last two decades (Leshik & Cralle, 2011). To give an example, if one searches algorithmic trading on Google Scholar (Date: 18/7/2017), only 500 results will appear that contain “algorithmic trading” in its title of which most are working papers and only 20 of these were written before 2005. When put into context, these 500 papers and books amount to only 0,08% of the 67000 articles which hold “financial crisis” in its name.

For this reason, many of the used sources remain books and working papers as information on algorithmic trading is still limited.

However, according to Kaya (2016), in 2014 high frequency trading already accounted for 49 percent of all the volume in U.S. equity markets, where one must keep in mind that high frequency trading is merely a subgroup of algorithmic trading. The connection between algorithmic trading and its effects on the human aspects are barely touched upon within existing financial literature.

It is likely that algorithmic trading in combination with improved artificial intelligence and quantum computing will completely change the financial markets as they are known to us now. Its relevance is undeniable and yet still so little is known about how the automation revolution impacts financial markets. Quantum computing and artificial intelligence still lie in the future, nevertheless human traders are already being substituted by computers on a great scale and its effects should be measurable using quantitative data. Measuring the effects of algorithmic trading is likely to give insights into how financial markets will behave in the future.

The rise of algorithmic trading imposes that a decline in direct human influence has manifested itself within the financial markets. Therefore, it can be reasoned that trading algorithms differ in trading behavior from human investors in the sense that trading algorithms are assumed to never deviate from their set of predefined rules unless stated in their rules. In other words, a trading algorithm will always behave within its programmed boundaries but account for all the information that is delivered to it. On the other hand, human traders are more likely to act based on their intuition and what is happening in their environment, with the tendency to value certain information above others.

These influences can be identified as behavioral biases which are recurring patterns in human behavior that simplify the predictability of their behavior (Heiner, 1983).

Humans are rational but only boundedly so and often are attracted to a majority opinion (Kahneman, 2003). In the world of finance, this pull of social gravity to the majority opinion, together with bounded rationality, cause the amplification of inefficiencies in the stock market as investors consistently keep overpricing popular stocks and underpricing less favored equities (Deman & Lufkin, 2000). Furthermore, Kim and Kim (2014) state that investor sentiment is affected by historical share price performance, which further strengthens the market inefficiencies. Considering that the stock market is already to a certain extent inefficient, it is likely that investor sentiment is often biased because of unrepresentative share prices which then again could lead to more inaccurate forecasts. Additionally, Chaboud, Chiquoine, Hjalmarsson & Vega (2014) find evidence that “algorithmic trading contributes to a more efficient price discovery process via the elimination of triangular arbitrage opportunities”. All in all, it can be assumed that the market is becoming more efficient with the increased influence of algorithms. Furthermore, according to the efficient market hypothesis developed by Fama (1995), this development should reinforce the random walk of stock prices and consequently its unpredictability.

Research on price dispersion related to algorithmic trading has not been performed previously and the most connected literature is on transaction costs dispersion by Enge, Russel & Ferstenberg (2007) where only Morgan Stanly data instead of complete stock market data is used. Furthermore, the link between algorithmic trading and market predictability also knows no predecessors and will explore new terrain in the field of algorithmic trading using the fundamental relationships between algorithmic trading, market quality, and information previously researched by Hendershott, Jones & Menkveld (2011) and Lyle & Naughton (2015).

For this reason, the main theme of this study is to evaluate *how increased algorithmic trading has affected analysts' capabilities to predict future market movements*. Removing emotional entities from the market is expected to improve the efficiency of the market and hence decrease the market predictability. Moreover, another sub-question is used to develop an empirical foundation for answering the main question which sums up to: *Does algorithmic trading lead to less price dispersion within the stock market?* Chaboud *et al.* (2014) show that automated trading strategies are less diverse than strategies used by human investors and that humans are responsible for a larger part of the variance in returns than their algorithmic counterparts. It follows that as algorithms possess more similarities than human traders it leads to suspect that the size of the range of returns also known as dispersion has decreased with increased algorithmic trading. Moreover, when looking at our data graphically it can be observed that return dispersion shows a clear downtrend over time, except for some extreme values during the financial crisis in 2008/2009, see *Figure 3*. Additionally, regressing dispersion against time confirms the downward slope resulting in a negative statistically significant coefficient on time with a p-value of 0.001. Considering that algorithmic trading increased over time it could imply a relation with dispersion.

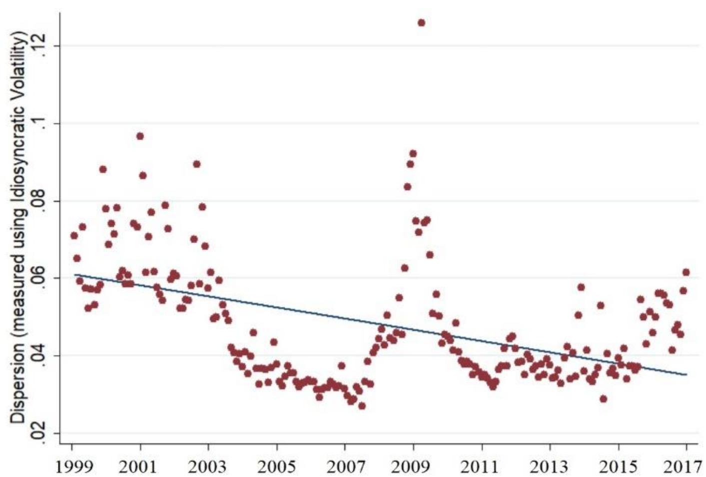

Figure 3. Dispersion against time

The current study investigates the effects of algorithmic trading in more detail, by systematically performing fixed effects panel data regressions. This might enable us to see how increased algorithmic trading has affected return dispersion and market predictability.

The regression findings lead to the conclusion that dispersion is indeed reduced through increased algorithmic trading. Furthermore, it is found that more algorithmic trading led to smaller prediction errors and hence improved market predictability.

In the next chapter, the theoretical framework that was used to establish this research will be discussed, built on the following research questions:

- *Does increased algorithmic trading within the market affect analysts' capabilities to predict future market movements?*

Sub-question:

- *Does algorithmic trading lead to less price dispersion in the stock market?*

Theoretical Background

1.1 Current State of Literature

First of all, to determine the influence of algorithmic trading on dispersion and market predictability, the origins of trading algorithms and the use of automated trading systems must be investigated. Additionally, to find how fewer human traders impact market predictability and dispersion, financial behavioral biases and market predictability should be examined as well.

1.1.1 Algorithmic Trading and Automated Trading Systems (ATS)

Leshik & Cralle (2011) explain that algorithms used for trading can be traced back to 1949 when Alfred Winslow Jones used an algorithm to balance between long and short positions on a hedge fund. An algorithm can be defined as a precise plan of steps that use computations to transform the input values into an output value. Fundamental to computer software and computations, algorithms have become a mainstream aid to the daily trader. It was not until the 1980's when algorithmic or black box trading became hugely profitable due to the invention of Pair Trading. Decreased costs, improved control mechanisms with self-documenting trade record and speed of execution are some of the advantages that algorithmic trading can offer to increase the likelihood of a trade to turn out successful.

First of all, in order to understand how exactly financial markets are affected by algorithmic trading, it is of need to get to the very basis of how a trading algorithm works. For that reason, an example algorithm for a coke vending machine is introduced. The algorithm can be constructed as simple as:

1. if sum of COINS INSERTED > $1 then RETURN(sum of COINS INSERTED – 1)
2. if sum of COINS INSERTED = $1 then DROP CAN
3. if sum of COINS INSERTED < $1 then SHOW MESSAGE(Insufficient Amount)
4. if ABORTED then RETURN(COINS INSERTED)

In this example the amount of coins inserted is the main input, its total value instructs the vending machine to drop the coke can and return any change if necessary. The algorithm will simply follow the set of rules to transform input into output and never deviates from these rules during the process. Similarly, to the example algorithm, trading algorithms are merely the set of predefined rules that convert input into output. Hence, trading algorithms are implemented within Automated Trading Systems that facilitate data collection to obtain input values and to transform output values into an actual action. Automated Trading Systems, also known as ATS, are a combination of both hardware and software that, by using trading algorithms, manages orders and positions within a stock portfolio on a basis of real-time data feeds and historical data that is stored in a database. The data input usually is a combination of factors such as the share price, volume, number of trades, technical indicators, and even news events can serve as an input value for the more advanced learning algorithms (van Vliet, 2007). It follows that the Automated Trading System autonomously creates orders based on its input values and implements these on the exchange, all within milliseconds competing with human investors (van Vliet, 2007). Hence it can be argued that an ATS is to a trading algorithm what a physical coke vending machine can be considered to be to a coke vending algorithm.

To construct an ATS one has to be familiar with computer science, quantitative finance, trading strategy and quality management. As "data is the lifeblood of electronic markets" the basis of ATS lies in the underlying data which can be managed using Microsoft Visual C++ or .NET applications. Technological superiority through ATS can offer an enormous advantage against competitors but still does not imply profitability (van Vliet, 2007).

Leshik and Cralle (2011) consider the most popular and widely used algorithms to be: Volume Weighted Average Price (VWAP), Time Weighted Average Price (TWAP), Percentage of Volume (POV), Search for Liquidity (Black Lance), Stay Parallel with the Market (The PEG), Large Order Hiding (Iceberg), Pair Trading Strategy, Leshik-Cralle, Recursive, Serial, Parallel and Iterative. Whereas Izumi, Toriumi & Matsui (2009) evaluated a distinct set of automated trading strategies. Izumi *et al.* compare the risk and return of all strategies within their sample set and concluded the strategies to provide better information than conventional methods. Moreover, the research showed that the impact of automated trading strategies on markets does not merely depend on their code. Additionally, the way they are combined and influence each other can impact the market more so.

The common factor amongst almost all popular trading algorithms seems to lie in technical analysis as the most popular trading algorithms are largely based on technical analysis related indicators such as moving average and the relative strength index as main indicators to create the buy or sell decision. Technical analysis pertains to predicting future stock prices by studying past stock price performance and several other trading statistics like trading volume and number of trades (Brock, Lakonishok & LeBaron, 1992).

Technical analysis is often considered as non-scientific due to its non-fundamental nature, nonetheless, a survey study by Menkhoff (2010) proves that the vast majority of all fund managers rely on technical analysis. Additionally, Bessembinder & Chan (1997) demonstrate that even rather simple technical analysis holds statistically significant forecasting power within financial markets. Technical analysis is more related to psychology than fundamentals and the more inductive technical analysis is used, the more it reinforces its own predictive powers almost like a self-fulfilling prophecy.

In *Figure 4* the risk and return outcome of the by Izumi *et al*. (2009, p. 3474) tested automated trading strategies agents are displayed. Partially to illustrate some available strategies other than the ones mentioned by Leshik & Cralle (2011). The results were achieved using backtesting on several stock markets. For these trading strategies to work, several parameters for the input variables can be used, it is elementary that the parameters take on values that reflect the price level of fundamental information to the firm and economic conditions and preferably use adaptive agents. The parameters and code as used by Izumi *et al.* (2009) can be found in Appendix B. Moreover, from the parameters can be derived that actual trading algorithms are very similar to the coke vending machine example algorithm illustrated above. For most of these algorithms, technical indicators based on price or volume information such as moving averages or upper and lower bands are used as input values.

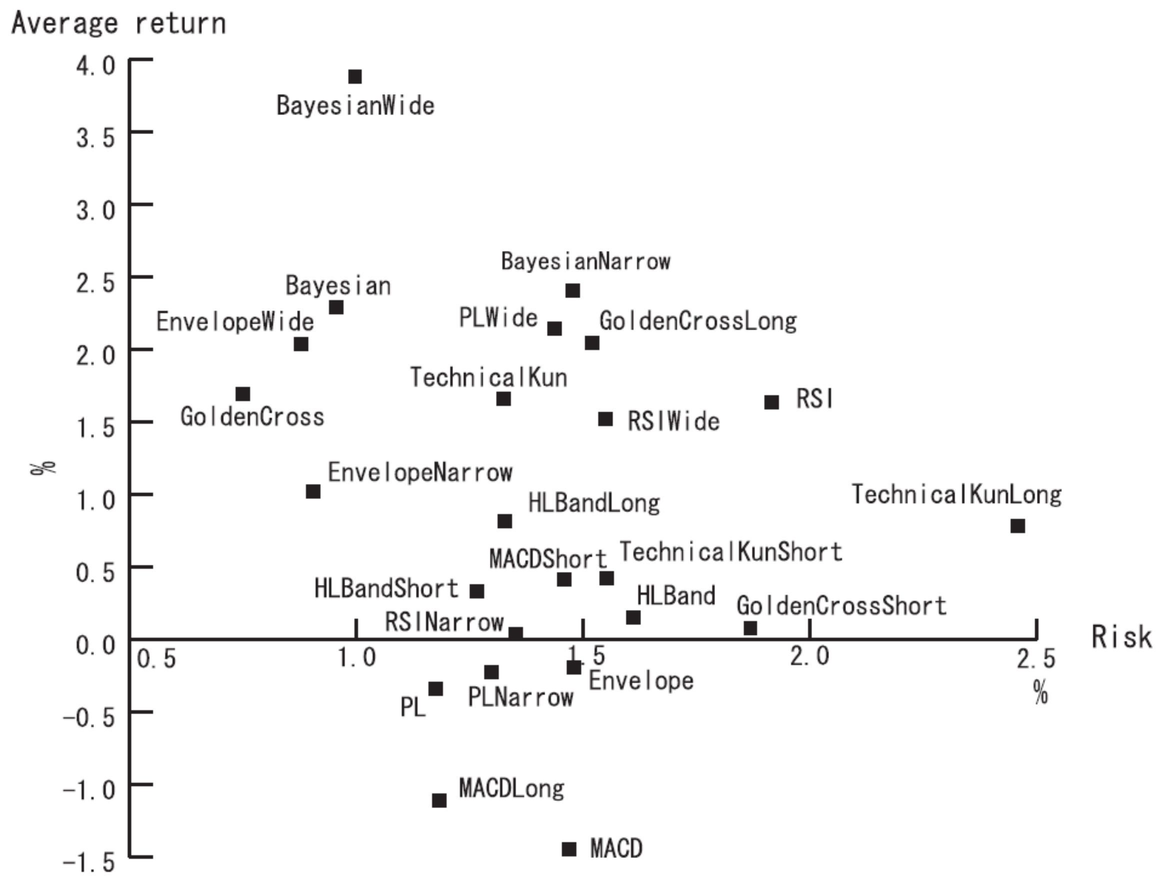

Figure 4. *Standard deviations versus Returns of ATS. Reprinted from "Evaluation of automated-trading strategies using an artificial market." By K. Izumi, F. Toriumi & H. Matsui, 2009, 72(16), 3474.*

Not only can ATS use price and volume information or technical indicators as input values. The algorithms can be integrated with machine learning to automatically read news feed and turn these into input values for the algorithm. According to Nuij *et al.* (2014) automating the incorporation of news feed into stock trading strategies can boost the returns of individual technical indicators compared to those without the incorporation of news messages. By means of extracting an event from a news feed text and pairing these with an impact based on historical stock price deviations for a specific event, this news variable can be used in addition to existing technical indicators.

Subsequently, the rules that are created through news associated events can be mutated within the trading algorithm by improved versions of the rules which have led to higher returns. Such automatic reprogramming on the basis of previous return outcomes is one example of how machine learning can be implemented in ATS.

1.1.2 Predictability & Biases in Behavioral Finance

Algorithmic trading is connected to behavioral finance in the sense that algorithms many times are programmed to trade on investor biases that exist because of individual or group behavior. The technical indicators incorporated in trading algorithms function through behavioral finance. Therefore, it could even be argued that technical economic indicators are actually socio-economic indicators. Behavioral finance often is contradictive to the efficient market theory suggesting that stock prices are actually to a certain extent predictable because of psychological and social concepts that cause inefficiencies on the stock market (Shiller, 2003).

There is polarity in human behavior that reflects how stocks oscillate between up and down trends similar to the state of mind and mood that a human or group of humans are in. All forms of emotion seem to exert forces on the stock market in one way or another. To name an example, even reaching physical new highs in the form a tall building reverbs on the stock market by leaving a peak in the graph followed by a fall. The Dubai stock market rose significantly after finishing the Burj Khalifa, world’s tallest building (Mitroi, 2014). Moreover, there are recursive patterns for some financial anomalies such as the day-of-the-week effect which are not yet understood. Evidence seems to suggest that these anomalies happen because of mass psychology (Shiller, 2003).

Vasiliou, Eriotis & Papathanasiou (2008) mention that moving averages stress where a trend is headed and flatten out fluctuations caused by the noise of irrational investors also known as noise traders. Additionally, Vasiliou *et al.* find that the utility of the technical trading rules used in their research improved over time.

1.1.3 Market Efficiency and Predictability

Litzenberger, Castura & Gorelick (2012) stated that market quality has improved in the past decades. A clear cause for this trend is increased competition through more automation and high frequency trading in the market which leads to decreases in the bid and ask spreads and improved liquidity. This improved liquidity causes the orders in limit order books to be exercised at a faster pace. Moreover, when relating market quality to algorithmic trading, Lyle, Naughton, and Weller (2015) discovered that algorithmic trading strategies which provide liquidity such as market making strategies increase market quality. Whereas liquidity taking, non-market maker algorithmic trading activity harms market quality.

Bouchaud, Farmer & Lillo (2008) conclude prices in markets to sustain a close to perfect unpredictability in the short run. Firstly, considering that outstanding liquidity is always small meaning that prices do not immediately mirror all information available to the market. Secondly, on electronic markets, there is no possibility to distinguish informed and uninformed trades for all trades have the same impact. It follows that all informative aspects of a trade should be internal to the market meaning that trades, order flow, and cancellations carry information.

Beja and Goldman (1980) rightfully state that a market constructed by humans can impossibly be so mechanically perfect and efficient that all information would directly be integrated into the prices before it can be observed.

Implying that price anomalies will always be present, leaving room for predictability. Moreover, Pesaran (2003) reinforces predictability by stating that “stanA large number of studies in the finance literature have confirmed that stock returns can be predicted to some degree by means of interest rates, dividend yields and a variety of macroeconomic variables exhibiting clear business cycle variations.” According to Pesaran, market-efficiency should be distanced from predictability.

Methodology

2.1 Data Collection & Processing

Most of the data and queries used for the research have been obtained through Wharton University of Pennsylvania's WRDS database & query tool (Wharton Research Data Services). In this research, three different datasets are used that exist within the WRDS database, named: *CRSP – Daily Stock, IBES – Price Target* and *Federal Reserve Bank – Interest Rates*. These sub-datasets eventually will be merged before the hypotheses can be tested and will be elaborated on in the following section. Further details on the datasets can be obtained from *Table A1* where all query extraction specifications are denoted.

The chosen data period from 1999 to 2017 is a trade-off between covering a period as extensive as possible while at the same time trying to keep the data editable within Stata using the limited computing power that the research has to its disposal. Moreover, since IBES data is only available from 1999 onwards, this will automatically be the start of the period. Furthermore, it can be argued using Glantz & Kissel's (2013, p. 258) *Figure 1* that algorithmic trading before 1999 would have amounted to such a small percentage of the market volume that it is not of critical value in answering the research question.

Additionally, only NASDAQ and NYSE equity price data is used as the U.S. based stock exchanges were first in establishing facilities to support the development of algorithmic trading. Consequently, high frequency trading gained volume share in the US more rapidly than in Europe, as shown in *Figure* 5 (Kaya, 2016, p. 2). Given these arguments and considering the limited computing power, U.S. data on algorithmic trading follows as the more established choice.

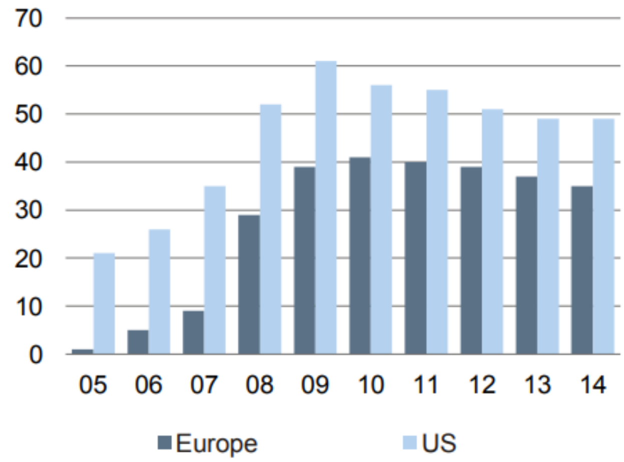

Figure 5. % Share of High Frequency Trading in total equity trading per year. Reprinted from "High-frequency trading: reaching the limits." By O. Kaya, 2016, 2. *Copyright by Deutsche Bank Research.*

2.1.1 CRSP – Daily Stock

First of all, the daily prices and trading data such as the daily number of trades and daily volume are extracted from the CRSP U.S. Stock database within WRDS. The previous mentioned CRSP query will function as the master dataset within the Stata environment and contains end-of-day prices for equity securities on the NYSE and NASDAQ exchanges. Additionally, CRSP also contains quote data, holding period returns, shares outstanding and trading volume information. Initially, the entire database is extracted for the period from 1999 to 2017 containing over 34 million observations. To start, only common stock observations are maintained within the query to improve the post-merger data compatibility with the IBES Price Target dataset. For common stock the variable share code amounts to either 10 or 11, hence only these share codes are kept within the sample. Moreover, tickers with multiple different shares are dropped as those are not properly comparable to the IBES identifiers which will be elaborated on later.

Additionally, a .TXT file consisting of the remaining company ticker identifiers is derived from the dataset within Stata in order to simplify the extraction of successive queries within WRDS as only information on those predetermined companies will be withdrawn from WRDS thus depressing the file size. Within the daily stock price query the actual price, bid, ask and shares outstanding are adjusted using the so-called adjustment factors in order to make the mentioned variables comparable over the entire 1999-2017 period. These adjustment factors are constructed by CRSP and adjust for corporate actions such as stock splits, dividends, and rights offerings. Additionally, the effective spread variable is created similarly to Hendershott *et al.* (2011) by means of taking the difference between the closing bid and ask its midpoint and the actual transaction price of that day as well as a volatility variable that is calculated as the deviation amid the daily high and the daily low.

2.1.2 IBES – Price Target

IBES, also known as the Institutional Brokers' Estimate System is a Thomson Reuters' database which holds historical analyst estimates for more than twenty forecast measures such as earnings per share, revenue, price targets, buy-hold-sell recommendations and gross profits regarding over 60,000 companies. After completing the extraction of price target estimation data including their horizon and analyst name data from WRDS using the same 1995-2017 period as used before, it was found that the IBES data could not directly be merged with the CRSP data. Concerning IBES, it contains two ticker variables and merely the variable official ticker is compatible with the ticker variable in CRSP and should not be confused with "ticker" in the IBES dataset.Hence, "oftic" is changed to its CRSP name: ticker.

Additionally, it must be mentioned that the in IBES so-called "announcement date" should be the leading date. Finally, price target estimation values are matched with their respective future actual price by lagging the forecast with its horizon meaning that an estimation with a horizon of 6 months is lagged 6 months.

2.1.3 Federal Reserve Bank – Interest Rates

The WRDS RATES database used in this research is based upon the Federal Reserve Board's H.15 release that contains selected interest rates for U.S. Treasuries and private money market and capital market instruments. Daily rates are per business day and reported in annual terms. To include interest rates as a controlling factor within the regressions, the rates of U.S. treasury bills with a maturity of 3 months are extracted from the WRDS RATES database for the period 1995 to 2017. The rates are merged with the master dataset using date as the common variable.

2.2 Data Analysis Methodology

To shed light on the automation process that entails the shift from human traders to automated trading systems, analyst predictions and their accuracy will be elaborated on in relation to algorithmic trading. First, however, our scope will focus on how algorithmic trading is measure and how dispersion has changed through algorithmic trading.

Moreover, all independent variables that will be used in regressions, are standardized to facilitate economic interpretation. Standardization is achieved by subtracting the corresponding time series' mean from the variables and dividing this deviation by the time series' standard deviation.

By standardizing all independent variables in such fashion, the standardized regression coefficients will represent a standard deviation change of the independent variables in the dependent variable. Hence, independent variable *X* is standardized such that:

$$(1)\ X'_{tj} = \frac{X_{tj} - \mu(X)}{\sigma(X)}$$

2.2.1 Algorithmic Trading Measure

Preparatory, a proxy has been developed to measure the development of algorithmic trading over time within the available CRSP data. To quantify algorithmic trading in a variable Hendershott, Jones, and Menkveld (2011) and Boehmer, Fong & WU (2015) use the daily number of electronic messages from the TAQ database per $100 of trading volume as a proxy to measure algorithmic trading. It is the most established measure within academic research, however, the TAQ database is not at this research's disposal and hence an inferior but comparable proxy is created. Inferiority lies in the fact that electronic messaging traffic information is not available in CRSP. However, as volume data is available, the best alternative measure would be a proxy that replaces the number of electronic messages with a comparable variable. Our data shows that volume did not increase over time while the number of trades did in a comparable way to the electronic messages used in HJM's proxy, making this a simplified but functioning replacement within our proxy for algorithmic trading. Moreover, algorithmic trading is associated with improved liquidity and an increased number of trades with a smaller volume per trade (Hendershott *et al.*, 2011).

Hence the new proxy for algorithmic trading is calculated as the daily number of trades executed for ticker j per dollar trading volume of that day derived from the CRSP database.

$$(2)\ Algorithmic\ Trading_{tj} = \frac{number\ of\ trades_{tj}}{volume_{tj}}$$

For it being a much noisier proxy, it gives a very similar representation of the development of algorithmic trading over time that was established by Glantz & Kissel (2013) which can be noted in *Figure 1 & 2*.

2.2.2 Effects of Algorithmic Trading on Dispersion

It is assumed that algorithms have more similarities than its human counterparts and for this reason dispersion is expected to decrease with more algorithmic trading. As flash crashes are known to happen with algorithmic trading (Johnson *et al.*, 2012) extreme short-term dispersion might have increased instead. However, considering that this study is only able to use daily data, flash crashes are not expected to influence the results. Hence, the hypotheses are formulated as:

H0: *Dispersion does not change with increased algorithmic trading*

H1: *Dispersion changes with increased algorithmic trading*

Idiosyncratic or stock-specific volatility is used to measure dispersion. Idiosyncratic risk can be calculated in numerous ways, the various measures however all give comparable results (Malkiel & Xu, 2003). Moreover, according to Bello (2008), there are no significant differences between the Capital Asset Pricing Model, the Fama French Three Factor Model and the Carhart Model regarding their outcome. Hence, in this study, the CAPM is used to calculate idiosyncratic volatility as this suits the dataset best. The CAPM formula used is as follows:

$$(3) \quad R_{tj} - Rf_t = \alpha_j + \beta_j(Rm_t - Rf_t) + \varepsilon_{tj}$$

Where: R_{tj} is Return of Stock j, Rf_t is equal to the Risk-Free Rate, Rm_t represents Return on the Market Portfolio and ε_{tj} is the error term of returns. Moreover, ε_{tj} is the variable of interest as it represents firm specific or idiosyncratic risk, and hence, dispersion.

First, two new variables are created to simplify the alpha and beta estimation process within Stata, namely: $ERS = R_{tj} - Rf_t$ and $ERM = Rm_t - Rf_t$. These are then applied in a simple OLS regression to estimate alpha and beta per ticker over the entire period. Almost 9500 regressions similar to (4) below are performed using a loop function in Stata after which the results are then saved in the variables α and β.

(4) $Y_{ERS} = \alpha_j + \beta_j * ERM_t$

Once alpha and beta are estimated ε_{tj} is then calculated as:

(5) $\varepsilon_{tj} = ERS_{tj} - \alpha_j - \beta_j(ERM_t)$

It follows that idiosyncratic volatility and thus dispersion is the monthly standard deviation of the error term as displayed below:

(6) $Idiosyncratic\ Volatility_{t(m)j} = \sigma_{t(m)}(\varepsilon_{tj})$

Finally, idiosyncratic volatility or preferably called dispersion is regressed on the algorithmic trading measure as in line with the hypotheses to analyze if return dispersion has changed through an increase in algorithmic trading. The model is also performed while controlling for firm fixed effects and year fixed effects as it is clear from *Figure 3* that for dispersion there seems to be quite a variance amongst different years and in particular for years of financial crisis.

The reason why fixed effects are used instead of random effects is that the Hausman test for random effects versus fixed effects is significant at the 99.9% significance level for regression (7) meaning that the unique errors ε_{tj} are correlated with the regressors and hence fixed effects panel data regressions are used to analyze dispersion. In regression (8) and (9) firm fixed effects and year fixed effects are added respectively to see if and how firm and year specific effects influence our model.

Comparing the results of regressions (7) and (8) will show the effect of firm specific effects whereas the comparison of (8) and (9) is to display the influence of year fixed effects.

$$(7)\quad Y_{Idiosyncratic\ Volatility_{tj}} = \beta_0 + \beta_1 * Algorithmic\ Trading_{tj} + \varepsilon_{tj}$$
$$(8)\quad Y_{Idiosyncratic\ Volatility_{tj}} = \beta_0 + \beta_1 * Algorithmic\ Trading_{tj} + a_j + \varepsilon_{tj}$$
$$(9)\quad Y_{Idiosyncratic\ Volatility_{tj}} = \beta_0 + \beta_1 * Algorithmic\ Trading_{tj} + a_j + \gamma_t + \varepsilon_{tj}$$

**With a_j as firm fixed effects and γ_t as year fixed effects*

2.2.3 Effects of Algorithmic Trading on Analyst Forecast Accuracy

To analyze the prediction accuracy of the remaining human analysts within the market, historical Thomson Reuters analysts' estimations obtained from the IBES dataset are used to obtain the prediction error for a certain forecast. It follows that the difference between the estimation value at time t and the adjusted price on date t divided by the adjusted price on that date gives the prediction error of a certain estimation by analyst i for stock j. Additionally, the prediction error is squared to emphasize on the analysts that were off most in their forecasts, be it below or above. As the squared prediction error will only return positive values it lays focus on just the deviation itself for the direction of the deviation is not of concern.

$$(10)\qquad Prediction\ Error_{t,i,j} = \left(\frac{estimation\ value_{t,i,j} - adjusted\ price_{t,j}}{adjusted\ price_{t,j}}\right)^2$$

Consecutively, the analyst prediction error variable will then be tested using regression analysis within the Stata statistical analysis software to see if analysts' predictions have become statistically more accurate since the development of automation within stock markets.

The dataset can be described as an unbalanced three-dimensional panel dataset for which stock ticker, date and analyst name represent the dimensions, for every ticker there are different numbers of analyst estimations on varying dates.

The “missing” data is due to analysts specializing in specific stocks and because the date at which estimations are placed is random, however, there is no actual missing data. The ticker and analyst variable are into a new combined variable called tic_alys where each group merely represents the specific forecasts by analyst i for ticker j. This procedure removes the need to drop the third dimension in order to run a multi-dimensional fixed effects panel data regression within Stata. These dimensions are only combined for regression (14) and (16) where firm and analyst fixed effects are included conjointly. To answer the research question the following hypotheses are developed:

H0: *Analysts' prediction error is not influenced by increased algorithmic trading*

H1: *Analysts' prediction error is influenced by increased algorithmic trading*

These hypotheses lead to the regressions below of which it is expected that analyst prediction error has indeed increased in the period where automation has taken place. It seems unlikely that analysts can predict the direction of future stock prices as the analysts would have to be able to execute transactions faster than the algorithms.

Therefore, it is hard to form a definite hypothesis as algorithmic trading probably also leads to less dispersion which could facilitate analyst predictions. For this reason, the hypothesis is two-sided where time t is in date format and per day. Testing analyst prediction error versus algorithmic trading is the most direct way of examining the effects that algorithmic trading has on analyst forecast accuracy. As many other factors potentially affect the forecast accuracy, sufficient control variables are to be added and fixed or random effects will be controlled for.

Moreover, to determine whether the regressions need to be controlled for fixed or random effects the Hausman test is used again. Testing for random versus fixed effects again gives a significant outcome with a 99.99% confidence level and hence H0 is rejected meaning that fixed effects need to be applied within the panel data regressions.

It follows, that six different panel data regressions will be tested within Stata to determine how prediction error is influenced. The first regression model is a plain panel regression merely to test the effect of algorithmic trading on the analyst prediction error whereas the remaining five are fixed effects panel data regressions that each control for a certain fixed effect. Regression (11) is the plain panel data regression, then firm fixed effects are added in (12) to see how firm-specific effects affect the regression output compared to the plain model. Thirdly, year fixed effects are controlled for as well using year dummies to control for a time trend and comparing regression (13) with (12) should deliver insight in the effects that time exerts on the dependent variable. Successively, analyst fixed effects are controlled for in regression (14) and again by merely adding this factor to the model it should become clear if and how the model is influenced by analyst-specific properties. By comparing the outcomes of the four regressions it should become clear if, how and which fixed effects affect prediction error. The first four regressions amount to:

$$(11) \quad Y_{\ln(Analyst\ Prediction\ Error)_{tij}} = \beta_0 + \beta_1 * Algorithmic\ Trading_t + \varepsilon_t$$

$$(12) \quad Y_{\ln(Analyst\ Prediction\ Error)_{tij}} = \beta_0 + \beta_1 * Algorithmic\ Trading_{tj} + a_j + \varepsilon_{tij}$$

$$(13) \quad Y_{\ln(Analyst\ Prediction\ Error)_{tij}} = \beta_0 + \beta_1 * Algorithmic\ Trading_{tj} + a_j + \gamma_t + \varepsilon_{tij}$$

$$(14) \quad Y_{\ln(Analyst\ Prediction\ Error)_{tij}} = \beta_0 + \beta_1 * Algorithmic\ Trading_{tj} + a_j + \gamma_t + \tau_i + \varepsilon_{tij}$$

**With a_j as firm fixed effects, γ_t as year fixed effects and τ_i as analyst fixed effects.*

Moreover, to control for other external effects, control variables are added to the regression leading to regression (15) and (16). By comparing these two regressions it should become clear how some exogenous factors affect prediction error and if the results remain robust even when other variables are controlled for. The control variables in regression (15) and (16) follow from established academic research on algorithmic trading and are summarized in *Table 2.*

First of all, volume is added as it is related to the algorithmic trading measure and is calculated as the daily average trading volume (Hendershott *et al.*, 2011). Moreover, number of trades are also added as they are part of the algorithmic trading proxy. Therefore, it is expected that number of trades conjointly with algorithmic trading should have a positive impact on the dependent variable whereas there should be a negative correlation between volume and analyst prediction error. Moreover, as algorithmic trading increases over time, it is expected for a positive relation to exist between date and the dependent variable. Effective spread is an indicator of adverse selection and also should influence analyst prediction error as a wider effective spread permits an increase in the amount of price discovery (Hendershott *et al.*, 2011). Effective spread is calculated as the difference between midpoint ask/bid and the actual closing transaction price (Lyle *et al.*, 2015). Moreover, increased Algorithmic Trading is on average associated with more volatility, however unlikely due to more price discovery. Volatility in this context is the log intraday price range or high minus low. Additionally, considering share turnover, it is used as a proxy for investor optimism and related to stock returns and hence possibly forecast accuracy (Subrahmanyam, 2015). Share turnover is calculated by dividing the number of trades on a day by the average number of shares outstanding on that day.

Consecutively, interest rate or risk-free rate is added as a control variable as Pesaran (2005) stated that share prices can to some extent be predicted using interest rates, for this reason, it is hypothesized that interest rates might influence analysts' forecast capabilities. Moreover, the interest rates are daily and therefore are not ruled out when controlling for year fixed effects.

$$(15)\ Y_{Analyst\ Prediction\ Error_{tij}} = \beta_0 + \beta_1 * Algorithmic\ Trading_{tj} + \beta_2 * Volume_{tj} + \beta_3 * Date_{tj} + \beta_4 * Number\ of\ Trades_{tj} + \beta_5 * Effective\ Spread_{tj} + \beta_6 * Share\ Turnover_{tj} + \beta_7 * Interest\ Rate_t + \beta_8 * Volatility_{tj} + a_j + \gamma_t + \varepsilon_{tij}$$

$$(16)\ Y_{Analyst\ Prediction\ Error_{tij}} = \beta_0 + \beta_1 * Algorithmic\ Trading_{tj} + \beta_2 * Volume_{tj} + \beta_3 * Date_{tj} + \beta_4 * Number\ of\ Trades_{tj} + \beta_5 * Effective\ Spread_{tj} + \beta_6 * Share\ Turnover_{tj} + \beta_7 * Volatility_{tj} + a_j + \gamma_t + \tau_i + \varepsilon_{tij}$$

**With a_j as firm fixed effects, γ_t as year fixed effects and τ_i as analyst fixed effects.*

Table 1. Control variables in comparable research on AT that are available in WRDS data.

Control variables	Hendershott & Riordan (2009)	Hendershott *et al.* (2011)	Lyle *et al.* (2015)	Boehmer *et al.* (2015)
Share turnover		X	X	X
Volume	X	X		
Number of trades		X		
Volatility	X	X	X	X
Effective spread	X	X	X	
Lagged Dependent				X

Research Findings

3.1 Dispersion and Algorithmic Trading

The output of panel data regressions (7), (8) & (9) that are run to test the dispersion hypotheses are presented in *Table 3*. First of all, it should be noted that by adding controlling factors to the basic regression (7) the adjusted R-squared increased from 0.097 to 0.219 and 0.667 respectively when controlling for firm-fixed effects and firm- plus year fixed effects. N amounts to a consistent 13.3 million in all three regressions and the coefficient on algorithmic trading increases with every addition in control. Nevertheless, algorithmic trading is highly statistically significant with a p-value below 0.001 in every regression and therefore H0 is rejected. Dispersion indeed changes with the increase in algorithmic trading. Contrarily, it should be noted that as the coefficients are standardized, a single standard deviation increase in algorithmic trading barely has any effect on the dependent variable and its economic significance, therefore, is limited. It can also be noted that when moving from regression (7) to regression (8) findings remain quite robust when adding firm fixed effects. On the other hand, when year fixed effects are included the R-squared triples and the coefficient on algorithmic trading is lowered from -0.004 to -0.0001. Hence, in the third and most complete regression, the effect of a one standard deviation increase in algorithmic trading only decreases dispersion with 0.0001.

Table 3. Panel data regression output with dispersion (Idiosyncratic Volatility) as dependent variable

Dispersion	(7)	(8)	(9)
Algorithmic Trading	-0.005***	-0.004***	-0.000***
Constant	0.049***	0.049***	0.064***
Firm Fixed Effects		Yes	Yes
Year Fixed Effects			Yes
N	13362249	13362249	13362249
Adjusted R-squared	0.097	0.219	0.667

* = p < 0.05; ** = p < 0.01; *** = p < 0.001

3.2 Prediction Error and Algorithmic Trading

The pairwise correlations which can be found in *Table 4.* are observed before performing regression analysis on the data. When beholding the algorithmic trading proxy, it can be noted that date and the risk-free rate are reasonably correlated with algorithmic trading. A high positive correlation with the date variable supports evidence that algorithmic trading has increased in size over time. Moreover, the negative correlation amongst the interest rates and algorithmic trading is likely due to time also being strongly negatively correlated with the risk-free rate and does not need to imply causality.

Table 4. Pairwise correlations between dependent & independent variables

	Prediction error	*Algorithmic trading*	*Volume*	*Date*	*Number of trades*	*Spread*	*Share turnover*	*Risk-free rate*	*Vola tility*
Prediction error	1								
Algorithmic trading	-0.116	1							
Volume	-0.007	-0.227	1						
Date	-0.075	0.669	-0.007	1					
Number of trades	-0.024	-0.077	0.816	0.147	1				
Spread	-0.001	-0.003	0.001	-0.002	0.001	1			
Share turnover	0.000	-0.061	0.149	-0.021	0.226	0.000	1		
Risk free rate	0.045	-0.447	-0.003	-0.672	-0.114	0.002	0.009	1	
Volatility	-0.029	0.224	0.075	0.068	0.215	0.003	0.282	-0.007	1

Executing the steps within Stata concerning analyst prediction error as described in the analysis methodology section, gives the following first regression results as displayed in *Table 5.*

Regression (11) is the basic regression and shows a particularly low R-squared of 0.013, followed by (12) where controlling for firm fixed effects drastically increases the adjusted R-squared with 0.62 and hence the explanatory value of the model. Moreover, the coefficient on algorithmic trading decreases from -0.982 in regression (11) to -0.346 in (12). Adding the year dummies in regression (13) does not impact the R-squared or coefficient on algorithmic trading much as both are merely elevated 0.005 and -0.01 respectively. Moreover, controlling for analyst fixed effects (14) does improve the explanatory value of the model as the adjusted R-squared is increased with 0.1 whereas the coefficient on algorithmic trading is barely affected with a -0.048 change. As can be perceived in regression (5) and (6) adding the separate control variables does not affect the model's R-squared and also the findings on algorithmic trading remain robust in both cases (15) and (16). Additionally, the constant, volume, number of trades, risk-free rate and volatility are all statistically significant in the model with a 99,99% confidence level. For risk-free rate, in particular, it is in line with the expectations that interest rates influence predictability as stated by Pesaran (2005). It can be derived from the Stata output that our independent variable of interest namely algorithmic trading is highly statistically significant at a 99,99% confidence level within the all the panel data regressions. Consecutively this means that H0 is rejected and algorithmic trading is expected to influence analysts' prediction accuracy.

Furthermore, number of trades being significant adds up to the idea that algorithmic trading positively affects analyst forecast accuracy. First of all, for algorithmic trading is measured using the number of trades and meanwhile because it influences the number of trades as algorithmic trading is associated with a higher number of trades per unit of volume.

Moreover, since all independent variables within *Table 5* are standardized it can be noted that the economic magnitude of algorithmic trading and interest rate are of specific interest when looking at their corresponding values. A single standard deviation increase in algorithmic trading decreases prediction error with -0.293, which is substantial considering that prediction error is squared and calculated in percentages. Hence, the true effect of such an increase in algorithmic trading on forecast accuracy is the square root of the coefficient which attributes to -54.1%. Apart from algorithmic trading, the economic significance of interest rate should also be mentioned as its effect is rather large considering that a single standard deviation increase in interest reduces the prediction error with 69.1% ($= -\sqrt{0.478}$).

Table 5. Panel Data Regressions with Prediction error as dependent variable

Prediction error	(11)	(12)	(13)	(14)	(15)	(16)
Algorithmic trading	-0.982***	-0.346***	-0.356***	-0.308***	-0.338***	-0.293***
Volume					0.082***	0.082***
Date					0.008	-0.186
Number of trades					-0.097***	-0.098***
Spread					-0.009	-0.013
Share turnover					0.034**	0.036***
Interest rate					-0.579***	-0.478***
Volatility					-0.066***	-0.062***
Constant	1.311***	1.3106***	1.349***	1.372***	2.910***	2.317***
Firm fixed effects		Yes	Yes	Yes	Yes	Yes
Year fixed effects			Yes	Yes	Yes	Yes
Analyst fixed effects				Yes		Yes
Control variables					Yes	Yes
N	350744	350744	350744	350339	349373	348970
Adjusted R-squared	0.013	0.633	0.635	0.736	0.635	0.736

*Legend: * = p < 0.05; ** = p < 0.01; *** = p < 0.001*

Discussion, Limitations & Recommendations for Future Research

4.1 Dispersion and Algorithmic Trading

For the first set of hypotheses focused on whether dispersion has changed through increased algorithmic trading the research indeed resulted in high statistical significance. All coefficients on algorithmic trading are significant at the 99.99% confidence level throughout the model hence H0 is strongly rejected. Despite the fact the test is double sided, simple graphical evidence and academic literature had already formed expectations considering the impact on dispersion. For the algorithmic trading coefficient is negative, dispersion is found to decrease with the development of algorithmic trading. This finding is also in line with the assumption that trading algorithms share more similarities than human traders. Moreover, these results are new considering that when it comes to related academic literature on algorithmic trading only transaction cost dispersion has been researched before by Enge *et al.* (2007). With more than thirteen million observations the model and its results should be very representative for the population. The model's most possible weakness lies in that the CAPM alpha and beta per firm are estimated over the entire period. However, as estimating the alphas and betas for each firm already consumed more than 12 hours of non-stop computing, it was not possible resource-wise to create monthly estimates. Additionally, flash crashes are known to happen within algorithmic trading as previously proven by Johnson *et al.* (2012), however as daily data is used the model is not affected by these minute-long stock market crashes. Finally, the economic impact of the results is quite low as the standardized coefficient on algorithmic trading is only -0,0001 within the most complete regression (9).

4.2 Prediction Error and Algorithmic Trading

The second set of hypotheses aimed at answering the main research question, namely whether the rise of algorithmic trading affected human analysts' forecasts accuracy. Again, all coefficients on algorithmic trading turned out to be highly statistically significant with a P-value below 0.001 for every modeled regression in *Table 5*. The negative coefficient on algorithmic trading implies that an increase in algorithmic trading leads to more accurate forecasts as the coefficient on algorithmic trading is negative. Following the efficient market hypothesis, it was expected that more algorithmic trading would lead to a less predictable, more efficient market and even though efficiency increased, analysts managed to make better forecasts according to the regression results. Lyle *et al.* (2015) had found that market maker algorithmic trading activity was more likely to affect market quality than other types of algorithmic trading. Looking at the results in *Table 5* it could be the case that less voluminous algorithmic trades did affect forecast accuracy positively and are represented more widespread in our sample considering that market and non-market maker trades are averaged out as CRSP only contains daily data. Furthermore, the decrease in dispersion could also have facilitated market predictability.

Apart from the main research question, it was found that a large part of the variation in prediction error seems to be firm-specific as R-squared increases substantially when adding firm fixed effects to the model. This could be due to some firm's share prices being harder to predict than others. It should also be noted that there is a wide array of exogenous factors that are likely to affect the prediction error variable, not all however, can be included in the model. Some of these variables include other technological developments such as the internet which also developed during the sample period.

Through the internet, analysts might have gained more direct and widespread access to new events and other information that was not available beforehand, potentially leading to better forecasts. Furthermore, it should also be mentioned that the IBES dataset was found to contain errors, many of the erroneous entries could be filtered out as they noted extreme share price estimates such as $20000 for a share that had an actual value of $13 at the time. Nevertheless, knowing the data derived from WRDS is not flawless this could imply that errors with normal values might exist and these are not easily excluded from the data for such faults are hidden. The study might be affected by these errors and further research on the validity of WRDS data should be performed. Additionally, the proxy used for algorithmic trading is an inferior proxy to the most established one developed by Hendershott *et al.* (2011), however, it best represents algorithmic trading using daily data instead of 5-minute trade order quote data.

Practical implications for the results are existent for traders, banks, fund managers, pension managers, algorithmic traders, and policymakers. Considering active algorithmic trading, profit margins have diminished due to competition and extreme transaction speeds where profits barely outweigh transaction costs anymore (Hendershott & Moulton, 2010). As prediction errors decreased for human analysts it might be beneficial for investment banks and investment institutions to replace certain algorithms for human traders once again. Perhaps due to the market becoming satisfied with trading algorithms that have similar strategies, aspects that are only found within human traders such as intuition and emotion might have become more profitable. These speculations, however, require further research.

Moreover, as algorithmic trading improves market predictability and thus lowers market quality, it might be relevant for policy makers and market regulating institutions to consider regulation regarding algorithmic trading in order to improve market quality.

It is clear however that financial markets as they are known to us now are undergoing a transformation and investment strategies must adapt as well to remain profitable. Robotization continues to happen within finance and these findings should attribute to the limited knowledge that is available about the effects of algorithmic trading on market predictability, slowly revealing its impact on financial markets.

Nevertheless, considerable research is still required before the effects of automation are understood. Future research could look at market maker and non-market maker algorithmic trading activity and the separate effects of various kinds of algorithmic trading. Until now, research on algorithmic trading has focused on algorithmic trading activity its entirety whereas the different sub-groups of algorithmic trading such as High Frequency Trading, Execution Algorithms, etc. seem to affect the market in diverse ways.

Conclusion

In this paper, empirical results are provided considering the effects of algorithmic trading on dispersion and market predictability. The results have been achieved by means of panel data regressions where controlling for firm fixed effects, year fixed effects and even analyst and firm fixed effects combined in the regressions where prediction error is the dependent variable. First of all, three different datasets namely CRSP – Daily Stock, IBES – Price Target, and FRB Interest Rates are merged to perform the analysis and algorithmic trading as a measure is created.

It was hypothesized that algorithmic trading leads to less dispersion and affects market predictability. The coefficient on algorithmic trading is very statistically significant for dispersion conjointly with prediction error. Hence, the results imply that algorithmic trading indeed leads to less dispersion and greater forecast accuracy which is not fully according to expectations. However, when looking at economic relevance, it is found that for dispersion the results have little economic implications whereas for market predictability the economic magnitude is significant. Moreover, once the findings are related to previous research it becomes clear that algorithmic trading should likely be differentiated in market-maker and non-market marker algorithmic trading. Non-market marker algorithmic trading activity is expected to positively affect forecast accuracy leading to reduced prediction errors which is in line with Lyle *et al.*'s (2015) findings that non-market maker algorithmic trading depreciates market quality.

This research, however, is merely a first attempt at trying to unpuzzle some of the effects of algorithmic trading on market predictability. Further research, using more detailed data including market maker information is required to truly discover the impact of algorithmic trading on market predictability.

References

Allen, F., & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of Financial Economics, 51(2)*, 245-271.

Bello, Z. Y. (2008). A Statistical Comparison of the CAPM to the Fama-French Three Factor Model and the Carhart's Model. *Global Journal of Finance and Banking Issues,* 2(2).

Bicchieri, C. (1987). Rationality and predictability in economics. *The British journal for the philosophy of science*, 38(4), 501-513.

Bessembinder, H., & Chan, K. (1998). Market efficiency and the returns to technical analysis. *Financial management,* 5-17.

Beja, A., & Goldman, M. B. (1980). On the dynamic behavior of prices in disequilibrium. *The Journal of Finance*, *35*(2), 235-248.

Blume, L., Easley, D., & O'hara, M. (1994). Market statistics and technical analysis: The role of volume. *The Journal of Finance, 49*(1), 153-181.

Boehmer, E., Fong, K. Y., & Wu, J. J. (2015). *International evidence on algorithmic trading* (AFA 2013 San Diego Meetings Paper). Retrieved from SSRN’s website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2022034

Bouchaud, J. P., Farmer, J. D., & Lillo, F. (2008). *How markets slowly digest changes in supply and demand* (Working Paper). Retrieved from SSRN’s website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1266681

Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of finance*, *47*(5), 1731-1764.

Brown, L. D., & Rozeff, M. S. (1978). The superiority of analyst forecasts as measures of expectations: Evidence from earnings. *The Journal of Finance*, *33*(1), 1-16.

Bunn, D., & Wright, G. (1991). Interaction of judgmental and statistical forecasting methods: issues & analysis. *Management Science*, 37(5), 501-518.

Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, *69*(5), 2045-2084.

Conway, M. R. (2003). *Professional Stock Trading: System Design and Automation.* Waltham, MA: Acme Trader LCC.

Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics*, *27*(3), 285-303.

Dreman, D. N., & Lufkin, E. A. (2000). Investor overreaction: evidence that its basis is psychological. *The Journal of Psychology and Financial Markets*, *1*(1), 61-75.

Easley, D., Lopez de Prado, M. & O'Hara, M. (2012). Flow Toxicity and Liquidity in a High Frequency. *Review of Financial Studies*, *25*(5), 1457-1493.

Engle, R. F., Russell, J. R., & Ferstenberg, R. (2006). Measuring and Modeling Execution Cost and Risk. *The Journal of Portfolio Management, 38*(2), 14-28.

Fama, E. F. (1995). Random walks in stock market prices. *Financial analysts journal*, *51*(1), 75-80.

Freund, W. C., & Pagano, M. S. (2000). Market efficiency in specialist markets before and after automation. *Financial Review*, 35(3), 79-104.

Fischer, R. (1993). *Fibonacci applications and strategies for traders* (Vol. 4). New York, NY: John Wiley & Sons.

Fong, K. Y., Parwada, J. T., & Yang, J. W. (2017). *Algorithmic Trading and Mutual Fund Performance* (Working Paper). Retrieved from SSRN's website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2980774

Glantz, M., & Kissell, R. (2013). *Multi-asset risk modeling: techniques for a global economy in an electronic and algorithmic trading era*. San Diego, CA: Academic Press.

Groß-Klußmann, A., & Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, 18(2), 321-340.

Groth, S. S. (2011). Does Algorithmic Trading Increase Volatility? Empirical Evidence from the Fully-Electronic Trading Platform Xetra. *Wirtschaftsinformatik proceedings 2011,* 112.

Heiner, R. A. (1983). The origin of predictable behavior. *The American economic review*, *73*(4), 560-595.

Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *The Journal of Finance*, *66*(1), 1-33.

Hendershott, T., & Moulton, P. C. (2011). Automation, speed, and stock market quality: The NYSE's hybrid. *Journal of Financial Markets*, 14(4), 568-604.

Hendershott, T., & Riordan, R. (2009). *Algorithmic trading and information* (Manuscript, University of California, Berkeley). Retrieved from: http://faculty.haas.berkeley.edu/hender/atinformation.pdf

Izumi, K., Toriumi, F., & Matsui, H. (2009). Evaluation of automated-trading strategies using an artificial market. *Neurocomputing*, 72(16), 3469-3476.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 6). New York, NY: Springer.

John Lu, Z. Q. (2010). The elements of statistical learning: data mining, inference, and prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(3), 693-694.

Johnson, N., Zhao, G., Hunsader, E., Meng, J., Ravindar, A., Carran, S., & Tivnan, B. (2012). *Financial black swans driven by ultrafast machine ecology* (Working Paper). Retrieved from SSRN's website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2003874

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American economic review*, *93*(5), 1449-1475.

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert systems with Applications*, 38(5), 5311-5319.

Katz, J. O., & McCormick, D. L. (2000). *The encyclopedia of trading strategies*. New York, NY: McGraw-Hill.

Kaya, O. (2016*). High-frequency trading: reaching the limits* (Deutsche Bank Research Briefing). Retrieved from Deutsche Bank Research's website: http://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000454703/Research_Briefing%3A_High-frequency_trading.PDF

Kearns, M., & Ortiz, L. (2003). The Penn-Lehman automated trading project. *IEEE Intelligent Systems*, 18(6), 22-31.

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

Kirilenko, A. A., & Lo, A. W. (2013). Moore's law versus murphy's law: Algorithmic trading and its discontents. *The Journal of Economic Perspectives*, *27*(2), 51-72.

Kumar, R. (2014). Magic of Fibonacci Sequence in Prediction of Stock Behavior. *International Journal of Computer Applications*, 93(11).

Kumiega, A., & Van Vliet, B. E. (2012). Automated finance: The assumptions and behavioral aspects of algorithmic trading. *Journal of Behavioral Finance*, *13*(1), 51-55.

Lam, K. (2005). *Is the Fama-French three factor model better than the CAPM?* (Doctoral dissertation, Department of Economics-Simon Fraser University). Retrieved from http://summit.sfu.ca/system/files/iritems1/9274/etd1764.pdf

Leshik, E., & Cralle, J. (2011). *An introduction to algorithmic trading: basic to advanced strategies* (Vol. 544). Hoboken, NJ: John Wiley & Sons.

Litzenberger, R., Castura, J., & Gorelick, R. (2012). The impacts of automation and high frequency trading on market quality. *Annual Review of Financial Economics*, 4(1), 59-98.

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The journal of finance*, 55(4), 1705-1770.

Lyle, M. R., Naughton, J. P. & Weller, B. M. (2015). *How Does Algorithmic Trading Improve Market Quality?* (Working Paper). Retrieved from SSRN's website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2587730

Malkiel, B. G., & Xu, Y. (2003). Investigating the behavior of idiosyncratic volatility. *The Journal of Business*, *76*(4), 613-645.

Menkhoff, L. (2010). The use of technical analysis by fund managers: International evidence. *Journal of Banking & Finance*, *34*(11), 2573-2586.

Mitroi, A. (2014). The Psychology Of Finance: One Decade, Two Nobel's. *Annals-Economy Series*, 344-353.

Nabil, M. (2002). Projected Fibonacci Targets. *Technical Analysis of Stocks and Commodities*, *20*(5), 46-51.

Nuij, W., Milea, V., Hogenboom, F., Frasincar, F., & Kaymak, U. (2014). An automated framework for incorporating news into stock trading strategies. *IEEE transactions on knowledge and data engineering*, 26(4), 823-835.

Nuti, G., Mirghaemi, M., Treleaven, P., & Yingsaeree, C. (2011). Algorithmic trading. *Computer*, *44*(11), 61-69.

Park, S., Jang, H. J., & Loeb, M. P. (1995). Insider trading activity surrounding annual earnings announcements. *Journal of Business Finance & Accounting*, 22(4), 587-614.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259-268.

Pesaran, M. H. (2003). Market efficiency and stock market predictability. *Mphil Subject*, *301*.

Pruitt, G., & Hill, J. R. (2003). *Building winning trading systems with TradeStation* (Vol. 167). Hoboken, NJ: John Wiley & Sons.

Rozeff, M. S., & Zaman, M. A. (1988). Market efficiency and insider trading: New evidence. *Journal of Business*, 25-44.

Stridsman, T. (2003). *Trading systems and money management: a guide to trading and profiting in any market*. McGraw-Hill Companies.

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *The Journal of Economic Perspectives*, *17*(1), 83-104.

Subrahmanyam, A. (2005). On the Stability of the Cross-Section of Expected Stock Returns in the Cross-Section: Understanding the Curious Role of Share Turnover. *European Financial Management*, *11*(5), 661-678.

Hendershott, T., & Riordan, R. (2009). *Algorithmic trading and information* (Manuscript, University of California, Berkeley). Retrieved from https://s3.amazonaws.com/academia.edu.documents/41530807/Into_the_Breech_The_Increasing_Gap_betwe20160124-6877-

Van Vliet, B. (2007). *Building Automated Trading Systems: With an Introduction to Visual C++. NET 2005*. Cambridge, MA: Academic Press.

Vapnik, V. (2000). *The nature of statistical learning theory*. Red Bank, NJ: Springer science & business media.

Vasiliou, D., Eriotis, N., & Papathanasiou, S. (2008). Incorporating technical analysis into behavioral finance: A field experiment in the large capitalization firms of the Athens stock exchange. *International Research Journal of Finance and Economics, 14*.

Weissman, R. L. (2005). *Mechanical trading systems: Pairing trader psychology with technical analysis* (Vol. 220). Hoboken, NJ: John Wiley & Sons.

Young, A. R. (2009). *Expert Advisor Programming: Creating Automated Trading Systems in MQL for MetaTrader 4.* Ormskirk: Edgehill Publishing.

Yuferova, D. (2015). *Intraday Return Predictability, Informed Limit Orders, and Algorithmic Trading* (Working Paper). Retrieved from SSRN's website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2686082

Zheng, B., Moulines, E., & Abergel, F. (2012). Price jump prediction in limit order book. *Journal of Mathematical Finance, 3*(2), 242-255.

Zubulake, P., & Lee, S. (2011). *The High frequency game changer: how automated trading strategies have revolutionized the markets* (Vol. 486). Hoboken, NJ: John Wiley & Sons.

Appendix A

WRDS Database source, extracted variables & data period overview

WRDS Library	Data Period	Variables	Description	Path	Extracted Observations
crspa / dsf	01/01/1995 – 31/12/2016	SHRCD COMNAM TICKER PRIMEXCH PRC VOL BID ASK NUMTRD RET SHROUT CFACPR CFACSHR SPRTRN	Daily Stock Prices for NASDAQ and NYSE	CRSP > Annual Update > Stock / Security Files > CRSP Daily Stock	34.6 million observations
ibes / ptgdet	01/01/1999 – 31/12/2016	OFTIC CNAME ESTIMID ALYSNAM ANNDATS ESTCUR HORIZON VALUE	Reuters Analyst Price Target Estimations	IBES > IBES Academic > Detail History > Detail History - Price Target	4.5 million observations
frb / rates	01/01/1995 – 31/12/2016	DATE TB_M3	Daily U.S. 3-month T-Bill rates	Federal Reserve Bank > Interest Rates > Interest Rates (Federal Reserve, H15 report)	5740 observations

Table A2. WRDS Query overview

Appendix B

ATS Code

Below, the code in C++ of trading strategies by Izumi *et al.* (2009, p. 3475) are presented corresponding to *Figure 4.* and their respective used parameters in *Figure B1*.

"

Variables:

p(t) is a price at day t; v(t) is a trading volume at day t; rand(n) is a random value from 0 to n-1.

> **buy()**
>
> BuyPrice:= p(t) +/- rand(p(t - 1) - p(t))/2
>
> Buy 1 unit at BuyPrice
>
> **sell()**
>
> SellPrice:=p(t) +/- rand(p(t -1) - p(t))/2
>
> Sell 1 unit at SellPrice

1. GoldenCross Agent

MA_l (t) := Moving Average from t - LMA to t -1

MA_s(t) := Moving Average from t - SMA to t -1

> **if** MA_s(t - I) < MA_l(t - 1) and MA_l(t) >MA_s(t)
>
> **then** buy()
>
> **else if** MA_s(t - 1) > MA_l(t -1) and MA_l(t) <MA_s(t)
>
> **then** sell()
>
> **end if**

2. HLBand Agent

HB :=max(p(t - 1), ... ,p(t - n))

LB :=min(p(t - 1), ... , p(t - n))

if p(t)>HB **then**

buy()

else if p(t)<LB **then**

sell() **end if**

3. MACD Agent

EMA_s(t):=Exponential Moving Average from t - SMA to t - 1

EMA_l(t):=Exponential Moving Average from t - LMA to t - I

MACD(t) = EMA_s(t)- EMA_l(t)

Signal(t) = 1 / $PS_{n=1...PS}$ MACD(t – n)

if MACD(t - 1)<Signal(t - 1) and MACD(t)>Signal(t) **then**

buy()

else if MACD(t -1)>Signal (t -1) and MACD(t) - Signal(t) **then**

sell()

end if

4. Envelope Agent

MA(t):= Moving Average from t - n to t - 1

if p(t)<MA(t) • (1 - P) **then**

buy()

else if p(t)>MA(t) • (1 + P) **then**

sell() **end if**

5. PsychologicalLine Agent

```
gain:=0
i:=0
    while i <12 do
    if p(t-i)> p(t-(i+1)) then
    gain:= gain +1
    end if
    i:=i+1 end while
    if gain/12 <PL then
    buy()
    else if gain/12 >PH then
    sell() end if
```

6. RSI Agent

```
gain:=0
1oss:=0
i:=0
    while i <12 do
    dp:=p(t - i)-  p(t - (i + 1))
    if dp>0 then
    gain:=gain + 1
    else if dp<0 then
    loss:=loss + 1 end if
    i :=i + 1 end while
    if gain/(gain +loss)<PL then
    buy()
    else if gain/(gain + loss)>PH then
    sell()
    end if”
```

ATS Parameters

Strategy	Parameter values
GoldenCross	{SMA, LMA} = {12 day, 24 day}
GoldenCrossShort	{SMA, LMA} = {5 day, 15 day}
GoldenCrossLong	{SMA, LMA} = {25 day, 75 day}
HLBand	n = 10
HLBandShort	n = 5
HLBandLong	n = 25
MACD	{SMA, LMA, PS} = {12 day, 26 day, 9 day}
MACDShort	{SMA, LMA, PS} = {6 day, 12 day, 5 day}
MACDLong	{SMA, LMA, PS} = {16 day, 40 day, 12 day}
Envelope	P = 0.05
EnvelopeNarrow	P = 0.025
EnvelopeWide	P = 0.10
PsychologicalLine	{PH, PL} = {0.75, 0.25}
PsychologicalLineNarrow	{PH, PL} = {0.60, 0.40}
PsychologicalLineWide	{PH, PL} = {0.85, 0.15}
RSI	{RH, RL} = {0.3, 0.7}
RSINarrow	{RH, RL} = {0.2, 0.8}

Figure B1. ATS Parameter values. Reprinted from "Evaluation of automated trading strategies using an artificial market." By K. Izumi, F. Toriumi & H. Matsui, 2009, 72(16), 3472.