Lund University

Department of Economics

Master Essay – 31 August 2018

NEKN01

Does High-Frequency Trading Affect Stock Market Predictability?

A study of the GARCH variances of the forecasting errors and how they have been affected by the introduction of High-Frequency Trading.

Authors:
Gustav Egesten †
David Hasselström ‡

Supervisor:
Joakim Westerlund

† sta13geg@student.lu.se

‡ eko13dha@student.lu.se

Abstract

In this paper, it is investigated whether High-Frequency Trading has an impact on the stock market predictability or not, using nine different *Autoregressive moving average* models forecasts are generated. Thenceforth, *ordinary least squares* are used to regress the variance of the forecasting errors with High-Frequency Trading as an explanatory variable in order to see if it has any form of impact. It was found that in four out of nine cases, High-Frequency trading had a significant impact on how adequately the stock market was predicted. Thus, the interpretation of the results is that High-Frequency Trading has a predictive ability and does affect the accuracy of forecasting.

Keywords: High-Frequency Trading, Predictability, Forecasting, ARMA, GARCH, OLS, Stock Market, S&P500

‘

Acknowledgements

We would like to express our most profound gratitude to our supervisor Joakim Westerlund whom always been providing us with sensible feedback and guidance throughout the progress of this paper.

1 Introduction

1.1 Background

During the last few decades, stock markets have seen significant advancements. The development has gone from trades being executed through so-called "floor trading" where buyers and sellers met in person with the support of market intermediaries, to trades in a more considerable extent being performed through automated systems (Jain and Johnson, 2008). This development started more than 40 years ago when the National Association of Securities Dealers (NASD) created their computer-assisted market making system for automated quotation (AQ) in the United States. This created what we today know as NASDAQ (Black, 1971a; Black, 1971b). This automation can be said to be the most drastic change to the stock markets and have resulted in significant effects on trading volumes and the efficiency of the markets. The implications of the automation have been increased liquidity, an increased number of investors and higher availability of information (Jain and Johnson, 2008). By the 1990s practically all securities trading was organized in wholly automated exchanges (Gomber et al., 2011). The next revolutionary development for the stock markets was the internet which has given investors substantial information advantages, previously only available to the largest banks and investment firms (Jain and Johnson, 2008).

Algorithmic trading developed at the beginning of the 21st century and quickly increased to become the most extensive form of trading in 2008 (Financial Times, 2018; Pole, 2007). When stocks are being traded through the use of algorithms, pre-programmed computers are utilized which determine the size of the order and its price, all while monitoring the conditions for different markets and securities (Hendershott et al., 2011). The computers have the ability to react to information exceptionally quickly and hence are able to execute orders at speed much higher than that of humans, giving them a great advantage (Economist, 2007). Using algorithms that generate orders automatically has decreased the trading costs drastically for investors since no expensive human traders are needed.

High-frequency trading (HFT) is often seen as a subgroup of algorithmic trading. While algorithmic trading is mostly associated with the execution of orders for clients, HFT refers to market participants implementing of trading strategies through access to highly technological equipment (Gomber et al., 2011).

HFT is not to be defined as a trading strategy in itself but a technical means to implement already existing trading strategies. As mentioned above, HFT can be seen as a part of the natural evolution of the securities markets. Like other technical advancements, the use of algorithms along with HFT have enabled actors with access to these technologies to achieve higher returns on their investments (Gomber et al., 2011).

The increasing popularity of Algorithmic trading in general and HFT, in particular, has gone hand in hand with other structural developments in the securities markets. In most markets, only registered members have direct access, which leads to them to act as intermediaries for other investors to gain market access. The market members that perform this type of function are referred to as brokers. However, as cost awareness has increased in the buy side, brokers have created different market access models, such as *Direct Market Access* (DMA). DMA allows the orders to go immediately to the markets through the broker's network, instead of brokers physically handling them (Gomber et al., 2011).

Another driver for the vast adoption of algorithmic trading and HFT has been the issue of *latency,* i.e. the time it takes data to travel between its source and destination. Although latency has always been a relevant factor in securities trading, its importance has been dramatically increased by market participants since HFT has become more prevalent. Historically, when trades were executed physically on trading floors, traders could benefit from their physical abilities, for example, if they could run faster across the trading floor beating their competition. Another advantage could be if the trader could scream louder than his competition, getting him more attention from the market maker, leading to his orders being executed rather than some other traders. In today's markets where the trading is executed through the use of algorithms, these physical advantages of the traders are no longer relevant. However, when trading in markets at high speed, the ability to send and receive orders at the lowest possible latency is of crucial importance (Harris 2003; Liu 2009). In order to minimize latency, traders engaging in automated trading use *co-location* or proximity services, which are provided by a high number of market operators. Through the co-location of their servers, market participants can locate their trading equipment in direct adjacent to the market operator's network, thus reducing latency significantly (Gomber et al., 2011). These technological advancements have made the technology popular among investors as well as researchers interested in the potential effects on the markets.

1.2 Motivation

In media, HFT has received a large amount of attention with some claiming potential advantages of the technology but most pointing at negative aspects that could be effects of HFT. The critics often point at the fact that HFT leads to increased volatility in the markets (Gangahar, 2008). Another concern is that it leads to *phantom liquidity,* which refers to the market liquidity being fleeting because of the practically instant posting and cancellation of orders. Other critics point to the fact that some HFT firms may engage in strategies which manipulate the market through the use of quote cancellations. Furthermore, some observers have claimed that firms which engage in HFT often are involved in so-called "front-running" whereby the firms trade before a large order to buy or sell stocks based on non-public market information about a forthcoming trade. Other issues regarding HFT is that it potentially makes the market systematic risk greater, in the sense that shocks to a small number of active HFT traders could damage the entire market. All of these concerns have gained lots of attention in media since the Flash Crash of May 6, 2010 (Congressional Research Service, 2014).

The Flash Crash was one of the most substantial point drops ever on the very broad stock index Dow Jones Industrial Average (DJIA). The index fell by nearly 1000 points in just a couple of minutes and then rebounded, the total crash lasted for less than 40 minutes. The event led to global regulatory discussions and massive media attention (Gomber et al., 2011).

Predicting the future development of the market is in the interest of every investor. This makes investigating how HFT affects the predictability of the stock market uttermost relevant. Since HFT has been prevalent in the securities markets for a number of years, it is possible to examine the predictability over a longer period.

With this in mind, this paper intends to investigate several predictive models including a number of well-known predictors to see if the predictiveness has changed. This, of course, could be due to other changes in market conditions during the period. One such thing could be the remarkable decrease in the interest rate during the sample period and the financial crisis of 2008. However, previous research has shown that during the period ranging from the 1930s to the 1950s interest rates at similar levels was also observed, all while predictors such as the dividend-price ratio and the earnings-price ratio have been proven to be effective (Goyal and Welch, 2008). Several factors could affect the predictive abilities of these measures since there have been many factors changing the markets in recent years. However, with HFT representing between 13% up to numbers as high as 70% of the total trades executed depending on the report and market investigated, it is safe to say that it affects the world's financial markets significantly (Nasdaq OMX, 2010; Chi-X, 2009; Tradeworx 2010; European Parliament, 2010).

1.3 Thesis structure

In the first chapter, the background and motivation are presented. In the second chapter, there is a literature review based on previous research in the field of HFT. The third chapter describes the data set that is the basis for the forecast and the regression. The fourth chapter explains the methodology. In the fifth chapter, the results are presented. The sixth chapter contains the discussion. Lastly, in the seventh chapter, the conclusion is presented.

2 Literature review

2.1 Predicting stock market returns

As investors have always been interested in predicting the development of securities markets to know where to invest, investors have used an almost infinite amount of predictors to know if the market is heading upwards, downwards or perhaps will see increased volatility or any other type of pattern. These predictors vary greatly in regard to what they examine and range from concrete measures like economic growth to more obscure ones like, for example, the weather as proposed by Daskalakis, et al. (2010). However, this paper will focus on the more traditional measures.

The most common stock market predictors found in the literature are the dividend-price ratio, the dividend-yield ratio, the earnings-price ratio, the dividend-earnings (payout) ratio, a number of interest rates and spreads, the inflation rate, the book-to-market ratio, volatility and the investment-capital ratio (Goyal and Welch, 2008). The use of the book-to-market ratio can be motivated by the findings of Fama and French (1992). Lettau and Ludvigson (2001) build their work on the present-value identity suggested by Campbell and Mankiw (1989), this is called CAY which is the relationship between aggregate consumption, asset holdings, and labor income. It has proven strong predictive power for intermediate horizon market returns. The aggregate consumption may be viewed as the dividend paid by aggregate wealth.

Since the literature, in general, have used different approaches and tested different periods in different markets, fair comparisons between predictive measures are not possible in general. Goyal and Welch (2008) tested the above-mentioned predictors and found variations in predictive abilities for different measures in different periods. Their evidence suggests that these measures perform poorly during the latter half of the 20th century. However, they do point out that the oil shock of 1973-1975 had a tremendous effect on the predictors' performance. They conclude that there is a vast difference in the in-sample and out-of-sample forecasts.

Goyal and Welch (2008) used some of the most common predictors that can be found in the literature in isolation, in order to forecast the Equity premium. They used the simple model:

$$Equity\ Premium_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t$$

Where x represents the variable that is supposed to predict the equity premium. β_0 and β_1 are constants and ε_t is the error term. They found that several of the predictors had insignificant results regarding their predictive abilities when tested in isolation. Among the models with insignificant results the dividend-price ratio, earnings-price ratio, dividend-earnings ratio, default-yield spread and inflation can be found. The variables that proved significant in their paper were: The book-to-market ratio, the investment to capital ratio, the net equity expansion and the percent equity issuing.

Furthermore, Goyal and Welch (2008) used a “kitchen sink” regression in which they tested all variables together, this model proved to have significant predictive abilities, and this approach forms the basis for this paper.

2.2 Previous empirical evidence on the effect of HFT on securities markets

Part of the critique against HFT is that it causes more volatility in the securities markets. This issue has received much media attention as a consequence of the Flash Crash. In the report by America's Securities and Exchange Commission and its Commodity Futures Trading Commission, the cause of the rapid fall in stock prices was HFT. The report found that prior to the substantial drop in the market, the volatility was unusually high and the liquidity very low. The primary trigger for the sudden decline was, according to the report, a large sell order in e-mini futures on the S&P 500. The algorithm which executed the trade was programmed to take trading volume into account and not price or time. The order was executed exceptionally quickly: In about 20 minutes, instead of being executed during a couple of hours as is usual (Economist, 2010).

However, as stated above, there is a major gap between the findings in empirical academic research and the picture of HFT provided in media and regulatory discussions (Gomber et al., 2011). For example, the research of Groth (2011) found no evidence for increased volatility when algorithms, rather than humans executed trades in his empirical study. These results were in line with those of Brogaard (2010) who in his research found that HFT not only does not increase volatility but even seems to dampen intraday volatility. These findings are in direct conflict with those in the report regarding the Flash Crash by America's Securities and Exchange Commission and its Commodity Futures Trading Commission. Regarding the Flash Crash of May 6, 2010, Kirilenko et. al. (2017) found that HFT probably did not cause the crash, although it did increase the volatility. Furthermore, their findings state that even though HFT did not cause the crash, it probably contributed by pulling out of the market as the conditions got worse.

Chaboud et. al. (2014) found in their empirical results that there is evidence that algorithmic trades show higher levels of correlation than their non-algorithmic counterparts. This implies that the trading strategies used in algorithmic trading are not as diversified as those used when humans execute trading strategies. This paper, like the research mentioned above by Groth (2011) and Brogaard (2010), did not find any relationship between algorithmic trading and increased market volatility. However, the research findings do suggest that algorithmic trading provide liquidity in periods of market stress. Hendershott et. al. (2011) result are in line with these findings and prove that, especially for large stocks, algorithmic trading improves liquidity and informativeness of quotes and prices.

According to the findings of Gsell (2008), algorithmic trading does affect the market prices and the market volatility. His results imply that low latency can contribute to significantly lower market volatility, however, large trading volumes may affect the prices of stocks or securities negatively. The results also showed that high latency had a positive effect on volatility.

Other academic findings point towards the positive effects of HFT. Jovanovic and Menkvelds (2010) findings in their paper state that HFT actually could improve welfare by 30%. On the negative side, HFT might cause or aggravate already existing problems related to adverse selection, which is an effect of asymmetric information between buyers and sellers. This could lead to bid-ask spreads increasing and trade volume decreasing. Jovanovic and Menkvelds (2010) conclude that the net effect is uncertain. However, Hendershott et. al. (2011) found the opposite when examining large stocks, that is bid-ask spreads reduce as a consequence of reduced adverse selection issues. Brogaard (2010) suggests that HFT is not harmful to the other market participants who engage in non-HFT activities and that HFT tends to improve market quality as the price discovery process is improved.

These academic results can be compared to the tone found in media:

> "High-frequency traders, who on the whole have maintained a low profile, say that because their frenzied trading provides liquidity, they help markets run smoother, improving the environment for all investors. However, combine the speed at which they operate, the outsourcing of decision making to computer codes, and an almost complete lack of regulation, and this shadow market can fuel and exaggerate volatility. Speed traders argue they tamp it down. Nonetheless, politicians and regulators are starting to get nervous. 'I'm afraid that we're sowing the seeds of the next financial crash,' says Sen. Ted Kaufman (D-Dela.), arguably D.C.'s most vociferous critic of high-frequency trading, or HFT" (Philips, 2010).

Another article refers to high-frequency trading as a new type of fraud:

> "While the SEC is busy investigating Goldman Sachs, it might want to look into another Goldman-dominated fraud: computerized front running using high-frequency trading programs" (Brown, 2010).

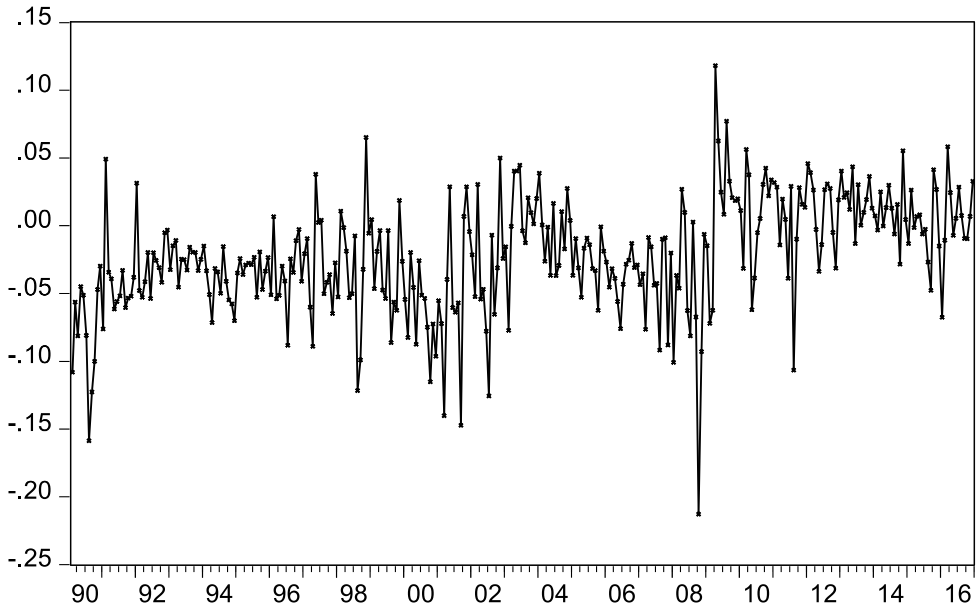
From this, we can conclude that just as Gomber et al. (2011) stated there is a large gap between the finding of academic research and the picture provided in media and regulatory discussions regarding algorithmic trading and HFT. We can also see that even though the results of the academic research vary, there seems to be a consensus that HFT has changed the trading environment on the securities markets in some way. This means that the research topic is relevant and that it is likely that the predictability also has been affected.

3. Data

3.1 Variable definitions and sources

Throughout this paper, the dependent variable will be the *equity premium.* This is the total rate of return on the S&P 500 minus the current short-term interest rate. The price data of S&P 500 is retrieved from Thomson Reuters Datastream. The equity premium is shown in figure 1.

Figure 1: Equity Premium (S&P500) 1990-2016



The dividend data used is the 12-month moving sum of dividends paid on the S&P 500 index. The data is provided by Standard & Poor's. The *dividend-price ratio* is calculated by taking the logarithm of dividends minus the logarithm of prices. The *dividend-yield ratio* is calculated by taking the logarithm of dividends minus the logarithm of *lagged* prices. This is consistent with Campbell (1987), Campbell and Shiller (1988), Fama and French (1988), Goyal and Welch (2008) and Shiller (1984).

The earnings data used is the 12-month moving sum of earnings on the S&P 500 index. The data used here are Goyal and Welch (2008) estimates, which are based on interpolations of quarterly earnings provided by Standard & Poor. The *earnings-price ratio* is calculated by taking the logarithm of earnings minus the logarithm of prices.

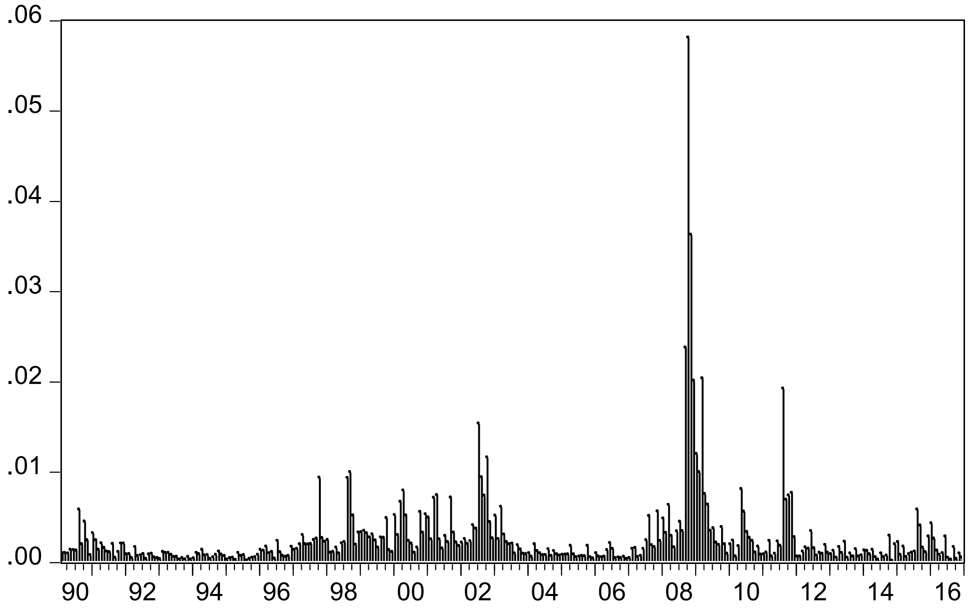
The *dividend-payout ratio* is calculated by taking the logarithm of dividends minus the logarithm of earnings. This is consistent with Campbell and Shiller (1988, 1998), Goyal and Welch (2008) and Lamont (1998).

The *book-to-market ratio* is the ratio of book value to market value for the Dow Jones Industrial Average. This ratio has been retrieved from Goyal and Welch (2008) dataset. They base their calculations on ValueLine's website due to limited access to data on book value for the entire market. This is consistent with Kothari and Shanken (1997) and Pontiff and Schall (1998).

For the *risk-free rate*, the 3-month Treasury Bill-rate, retrieved from Thomson Reuters Datastream, is used.

The *stock variance* (SVAR) is computed as the sum of squared daily returns on the S&P 500. The data is provided by The Center for Research in Security Prices (CRSP). Including a variable for the variance should help the forecasting model take volatility clustering into account when conducting our forecasts. The variance of S&P500 is illustrated in figure 2.

Figure 2: Stock Variance (S&P500) 1990-2016



Financial Times (2018) published an estimate, initially created by the TABB Group, which represents the market share of HFT in the period 2005-2016 which will be used as an approximation of the prevalence of HFT, in the securities markets, throughout this paper.

A dummy variable has been created in order to capture some of the market fluctuations during the years of the financial crisis. The variable takes the value zero until the point of the financial crisis and the value one during the financial crisis and again zero from the crisis has ended onwards. This variable is supposed to absorb the extreme fluctuations that are not otherwise captured by the models.

3.2 Data Limitations

The data of the prevalence of algorithmic trading, and HFT, in particular, is in a large extent limited. The markets and the regulators are the only sources, and they are often not willing to release this information. However, there are recognized estimates of the percentage that HFT constitutes to the total of trading (Kaya, 2016).

The book value data for an entire index is somewhat limited. Hence, we use the book-to-market data Goyal and Welch (2008) used in their examination of predictors. This data as mentioned above is retrieved from ValueLine's website and represents book value data from Dow Jones Industrial average. This makes our estimates of the predictive abilities somewhat different from the other predictors which are constructed of data on the S&P 500. However, the two indices are highly correlated during the sample period. In fact, the indices moved the same direction 95% of the times in the last 50 years (Prestbo, 2011). Furthermore, Goyal and Welch (2008) use this data in their kitchen sink forecast on the S&P 500 equity premium in which this measure proved to have significant predictive abilities.

4. Method

In order to absorb the effect of HFT on the stock market predictability, there will be several in sample forecasts, using different model specifications, generated in the period 1990-2016. The specifications have their basis in the approach with a kitchen sink regression, used by Goyal and Welch (2008). In order to generate the forecast, different ARMA specifications are used.

If an OLS-regression would be run directly on the forecasting errors, we would only be able to absorb changes in the mean. Since the assumption in this paper is not that HFT would generate a different level of returns compared to the forecasts but rather cause other changes in the predictability, one more step is included. The variance of the forecasting errors is estimated by a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. Lastly, an OLS regression is executed on the variances of the forecasting errors with HFT, SVAR and a dummy variable for the financial crisis as explanatory variables in order to find if there is a significant effect of HFT on the variance of the forecasting errors. If the variance of the forecasting errors has changed this implies that the predictive ability of the models has changed due to HFT.

4.1 Estimation period

It is not entirely clear how the estimation period should be chosen and over which period the model should be evaluated. However, this choice has to be ad-hoc. Furthermore, it is of vital importance that there is enough initial number of observations to get a regression estimate that is reliable when the evaluation period begins, all while having an evaluation period that contains enough observations to be representative as to if the forecasting model works (Goyal and Welch, 2008). In the case of this paper, data is one limitation regarding the choice of both estimation period and evaluation period. However, with access to approximately 16 years during the evaluation period and 11 years in the estimation period, this should render a sufficient sample size. However, the number of observations in the estimation period and evaluation period vary between the model specifications as is illustrated in table 1. This variation is needed as some variables might be affected by specific events which make the forecasts inaccurate and makes the further research less interesting.

Table 1: Forecast observations & periods

Specification	Number of observations in estimation period (rolling)	Evaluation period	Number of observations in evaluation period
1	130	2001M01-2016M12	192
2	124	2000M07-2016M12	198
3	124	2000M07-2016M12	198
4	130	2001M01-2016M12	192
5	130	2001M01-2016M12	192
6	124	2000M07-2016M12	198
7	130	2001M01-2016M12	192
8	124	2000M07-2016M12	198
9	130	2001M01-2016M12	192

4.2 Model specification

In previous research within this area, the models used are in most cases different. As Goyal and Welch (2008) found, the kitchen sink regression, in which they tested all predictors at once proved to have significant predictive abilities. Hence, this approach is used. However, some of our financial ratios cannot be tested together due to perfect collinearity. These include the dividend-price ratio together with the dividend yield ratio, the earnings-price ratio with the payout ratio, etc. This means the model specifications shown in table 2 are used and compared to the actual *equity premium* in order to retrieve our forecasting errors.

Table 2: Variables included in the different model specifications

Variables/Specification	1	2	3	4	5	6	7	8	9
Equity Premium	x	x	x	x	x	x	x	x	x
Inflation	x	x	x	x	x	x		x	
Earnings Price Ratio				x					
Risk Free Rate	x	x	x	x	x	x	x		x
Dividend Yield Ratio	x			x	x	x	x	x	x
SVAR	x	x	x	x	x	x	x	x	x
Book to Market Ratio	x	x			x		x	x	x
Cay					x	x			x
Payout Ratio	x	x	x		x	x		x	x
Dividend Price Ratio		x	x						

4.3 Heteroscedasticity

Throughout this paper, it is assumed that the data used is heteroscedastic due to the phenomena of *volatility clustering* which often occur in financial data. Volatility clustering implies that the volatility tends to depend on the variance in the previous period. The volatility clustering can be seen in Figure 2 the volatility tends to gather around specific periods. Due to this, a variable for the stock variance is included in the forecasts and the OLS regression.

4.4 Models

In order to understand the impact of HFT on the stock market, there will be an IS forecast generated by different models for the period 1990-2016. A recurring feature in similar research is that the forecasting model includes some form of lags or/and shocks, consistent with the previously mentioned research by Goyal and Welch (2008). Hence, the ARMA will be useful, which include both of these.

To obtain the optimum number of lags that should be included in the model, the Akaike Information Criterion (AIC) will be used. The maximum amount of lags and shocks is set to four since the coefficients of higher order lags tend to be very small and the effect of an additional lag on the forecast would be negligible.

4.5 Forecasting Procedure

The forecasts performed in this paper are of an in-sample character. This means that the forecast will be made of observations of equity premium that already are in the data set and then compared to the actual observations. The idea behind the approach is to see how the accuracy of the forecast has evolved. From the forecasted values, the residuals can be derived by subtracting the actual values of the equity premium.

There are different ways to evaluate the forecasts performed in this case. Firstly, to get an intuition about the accuracy, it could be appropriate to plot the actual and predicted values in a graph to see if they in some way follow the same pattern. In essence, this paper investigates a quite problematic period where the crisis changes the financial climate significantly for a period in the set. Hence, it is of great interest to, in some way, absorb this effect. Furthermore, the AIC is used to more formally get information about how the different forecasts perform relative to one another. Although, all the forecasts performed will be the basis for the regressions later.

4.6 GARCH and OLS regression

The GARCH-regression will be run on the forecasting errors obtained in order to estimate the volatility of the residuals during the forecasted period.

The estimated variances from the GARCH-model will later be regressed through an OLS-regression where HFT, SVAR and a dummy variable for the financial crisis is included as the independent variables. The result from the OLS will tell whether the HFT have an impact on the way the stock market can be forecasted.

5. Results and Analysis

The most prominent result of this paper is that HFT has a significant effect on the predictive abilities of some of the specifications tested. As can be seen in table 3, in two out of nine OLS-regressions, HFT proves to be significant at the 5%-level. The model specifications that showed these results were *Specification 2* and *Specification 4.* In two out of nine OLS-regressions, HFT proves to be significant at the 10%-level. The model specifications that showed these results were S*pecification 6* and *Specification 8.* In the cases where HFT had significance, all model specifications but *Specification 8* showed a positive coefficient for HFT. This is to be interpreted as the variable contributing to the variance of the forecasting errors which means that HFT has made the forecasting models less effective at predicting to the stock market. In one specification, the effect was the opposite. However, these results are only weakly significant as it was only possible to observe significant effects in four out of the nine specifications. Not having consistent results across all specifications makes it hard to, with certainty, state what the effect is. These findings seem to be consistent with the findings of previous research which, in many cases, has concluded that HFT has some effect on securities markets, but most studies do not agree on what the effect is.

Furthermore, SVAR had a significant effect in six out of nine OLS-regressions at the 5%-level. The coefficients of the statistical significant SVAR variables were all positive. The dummy variable which takes the financial crisis in shows significance in three of the nine regressions at the 5%-level. The coefficient of the dummy variable is negative in some regressions, and positive in some.

When it comes to forecasting, it is always about getting “as close” as possible. There are probably a thousand variables that could explain the progress of equity premium. Due to the scope of this paper, the course of action was somewhat limited. However, the nine models that were forecasted find endorsement in previous research within the subject and the fact that multiple models are examined gives strength to the results presented in table 3.

Table 3: Results of OLS regression

Specification	Variable	Coefficient	Std. Error	t-Statistic	Prob.
1					
	C	0.002343	0.000413	5.673494	0.0000***
	HFT	2.59E-06	6.20E-06	0.417178	0.6770
	SVAR	0.110876	0.044374	2.498663	0.0133**
	DUMMY	0.000163	0.000855	0.190428	0.8492
2					
	C	0.001216	0.000992	1.226166	0.2217
	HFT	3.67E-05	1.49E-05	2.463250	0.0147**
	SVAR	0.123721	0.106574	1.160897	0.2472
	DUMMY	0.002926	0.002054	1.424901	0.1558
3					
	C	0.001284	0.000202	6.362734	0.0000***
	HFT	-8.86E-07	3.03E-06	-0.292712	0.7701
	SVAR	0.090923	0.021685	4.192982	0.0000***
	DUMMY	0.001749	0.000418	4.184842	0.0000***
4					
	C	0.000864	0.000468	1.844533	0.0667*
	HFT	1.55E-05	0.050332	2.394128	0.0282**
	SVAR	0.120500	7.03E-06	2.211250	0.0176**
	DUMMY	0.001478	0.000970	1.523473	0.1293
5					
	C	0.002651	0.000402	6.601704	0.0000***
	HFT	-8.39E-07	6.05E-06	-0.138719	0.8898

	SVAR	0.110672	0.000832	-0.667594	0.0111**
	DUMMY	-0.000556	0.043168	2.563753	0.5052
6					
	C	0.002779	0.001034	2.687761	0.0078***
	HFT	2.45E-05	1.56E-05	1.570367	0.0881*
	SVAR	0.089470	0.111152	0.804934	0.4219
	DUMMY	-0.001069	0.002143	-0.498873	0.6185
7					
	C	0.002185	0.000382	5.713190	0.0000***
	HFT	4.18E-07	5.74E-06	0.072888	0.9420
	SVAR	0.110810	0.000792	0.808788	0.0076***
	DUMMY	0.000640	0.041094	2.696501	0.4197
8					
	C	0.004312	0.000682	6.326450	0.0000***
	HFT	-1.65E-05	1.02E-05	-1.617615	0.0974*
	SVAR	-0.026602	0.001411	5.079085	0.7169
	DUMMY	0.007169	0.073243	-0.363207	0.000***
9					
	C	0.002716	0.000380	7.153067	0.0000***
	HFT	-2.01E-06	5.72E-06	-0.351913	0.7253
	SVAR	0.100536	0.000787	-0.962659	0.0147**
	DUMMY	-0.000758	0.040819	2.462987	0.3370

***=signifcance at 1%, **=significance at 5%, *=significance at 10%

Specification 1 proves to have the highest adjusted R-squared (0.521739), this is illustrated in table 4. However, *Specification 5* and *Specification 9* show similar numbers. As is shown in table 2 these specifications are very similar to *Specification 5,* differing from *Specification 1* in the way that it does not include the book-to-market ratio and *Specification 9* is not including the risk-free rate.

Specification 4 show the lowest adjusted R-squared (0.289699), this specification uses the dividend-price ratio instead of the dividend-yield ratio and excludes the book-to-market ratio. Both the dividend-yield ratio and the book-to-market ratio proves to be significant in *Specification 1,* while the dividend-price ratio did not show significance in *Specification 4.* This should explain the drastically lower adjusted R-squared value.

Table 4: Selected ARMA models with corresponding adjusted R-squared and AIC values			
Specification	**Selected ARMA model**	**Adjusted R-squared**	**AIC value**
1	(3,4)	0.521739	-4.563478761
2	(2,1)	0.376669	-4.405444526
3	(1,2)	0.344954	-4.370194742
4	(1,0)	0.289699	-4.28156926
5	(3,4)	0.517307	-4.548892846
6	(4,3)	0.450878	-4.492273486
7	(4,2)	0.498151	-4.567429319
8	(4,3)	0.447174	-4.468485111
9	(3,4)	0.519501	-4.561557706

The AIC-values shows similar results where *Specification 4* have the most substantial value (-4.28156926), which implies that this forecast has the most inferior quality. *Specification 7* have the lowest AIC-value (-4.567429319), which implies that this specification has the highest quality. However, *Specification 1* and *Specification 9* have almost precisely the same AIC-values of -4.563478761 and -4.561557706.

Having equity premium as the dependent variable can, of course, be questioned. In essence, another variable that in some way measure the return on the stock market, for example, the return rate, could have generated different results, and that could be something to consider in further research. The reason why a different dependent variable could generate different results is that the coefficients of HFT in some of the OLS-regressions are very small. Although, the coefficients should be in the "same" direction (if positive coefficient here, it should be a positive coefficient with a different dependent variable).

Furthermore, there are other methods to apply to find the impact of HFT. It does not necessarily need to be the approach in this paper with the forecast, GARCH- and OLS-regression. In essence, one could argue that the question could be answered in a more simplified way, with a common OLS-regression with adequate variables (including HFT) to check significance and hence use it in future forecasts.

The approach to look at the variance of the forecasting errors which is proposed in this paper does, however, give a more interesting perspective to the analysis. It would be possible to run an OLS-regression directly on the forecasting errors and examine if HFT has had a significant effect. The results of this approach will show if the HFT-variable has had any effect on the mean of the forecasting errors rather than any effect on the forecasting errors. This approach could be interesting to examine as well. However, it would not fully answer the research question of this paper, namely "Does High-Frequency Trading have an Effect on Stock Market Predictability?". This approach would rather answer the question "Does HFT Make Forecasting Models Consistently Over- or Underestimate The Development on the Stock Markets?". In essence, the results of the approach used in this paper do explain more when it comes to if HFT has affected predictability.

6 Conclusion

6.1 Summary of empirical findings

This paper examines the question *does HFT have any effect on Stock market predictability?* It is possible to conclude that it probably has. In four out of the nine forecasting models tested, the HFT variable showed a significant effect when analyzing the variance of the forecasting errors. However, it is hard to state in what way HFT has affected the predictability as, in one case, the coefficient was negative. The conclusion is therefore that it could be useful to include a variable that takes the presence of HFT into account when trying to predict future developments on the stock market.

This can be compared to Jovanovic and Menkvelds (2010) findings that HFT might cause or aggravate already existing problems related to adverse selection, which is an effect of asymmetric information between buyers and sellers. If problems with asymmetric information aggravate due to HFT, it is not impossible to assume that the predictability of the market changes in some way.

Chaboud et. al. (2014) found that algorithmic trades show higher levels of correlation than their non-algorithmic counterparts. These results might affect the predictability as more correlated trading might be harder to foresee, this could be an explanation of the findings in this paper.

As stated in the literature review, even though the findings of previous academic research vary concerning the effect HFT has had on securities market there seems to be a consensus that it has changed the market in some way. This finds support in the results provided in this paper as well.

6.2 Limitations and extensions

In this paper, however, there are naturally some limitations, the greatest one being access to data. First and foremost, the data on the presence of HFT is, as mentioned earlier somewhat limited and have affected the manner in which this study was possible to conduct. Furthermore, this is true for many of the predictors used as well which in the case of this paper was only available on a monthly basis. The effects of HFT might have been captured better with a higher frequency of the data.

Furthermore, a limitation that affects the forecasts but not necessarily the results of the paper is the financial crisis. The financial crisis is tough for any forecasting model to take into account and was the reason for including the variable in the examination of the forecasting errors. However, the primary goal of this paper is not to make the most accurate forecast but rather examine if HFT has an effect on predictability and this is the basis for the choice to use several forecasting models.

In future research, it would be interesting to examine the topic in a period that does not involve the financial crisis which might affect the performance of the forecasting models. However, this would require a different approach as it would not allow for the evaluation period to include observations where HFT is not present. An approach that could be useful would be, with access to better data, use a higher frequency of the observations, for instance, daily data. This would mean that it might be possible to capture smaller fluctuations in the presence of HFT during a period where the markets showed more stability, thus generating more telling results.

Furthermore, it would be interesting to examine other markets than the New York Stock Exchange or Nasdaq and, for instance, examine European markets where HFT have been less prevalent. Studies of other securities than stocks could also be interesting as there is HFT prevalent in many markets.

7 References

Black, F. (1971a). Towards a Fully Automated Exchange, Part I, *Financial Analysts Journal,* vol. 27, pp. 29–34

Black, F. (1971b). Towards a Fully Automated Exchange, Part II, *Financial Analysts Journal,* vol. 27, pp. 24–28

Brogaard, J.A. (2010). High Frequency Trading and its Impact on Market Quality, Phd Thesis, Kellogg School of Management, Northwestern University School of Law, Available Online: https://secure.fia.org/ptg-downloads/hft_trading.pdf [Accessed 10 May 2018]

Brown, E. (2010). Computerized Front Running: Another Goldman-dominated Fraud, *Huffington Post*, 26 June, Available online: https://www.huffingtonpost.com/ellen-brown/computerized-front-runnin_b_548148.html [Accessed 15 May 2018]

Campbell, J. Y. (1987). Stock returns and the Term Structure. *Journal of Financial Economics*, vol. 18, no. 2, pp.373–99

Campbell, J. Y. & Mankiw, N. G. (1989). Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence, *NBER Macroeconomics Annual 1989*, vol. 4 pp. 185-246

Campbell, J. Y. & Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, vol. 1, no. 3, pp. 195–227

Campbell, J. Y. & Shiller, R. J. (1998). Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management*, vol. 24, no. 2, pp. 11–26

Chaboud, A., Chiquoine, B., Hjalmarsson, E. and Vega, C. (2014). Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market, The Journal of Finance, vol. 69, no. 5, pp. 2045-2084

Chi-X (2009). Europe Trading Statistics – Q1 2009[pdf] Available at: http://www.chi-x.com/chi-x-pressreleases/chi-x-europe-q1-2009-trading-stats-final--2-.pdf [Accessed 12 May 2018]

Congressional Research Service (2014). High-Frequency Trading: Background, Concerns, and Regulatory Developments, Available Online: https://fas.org/sgp/crs/misc/R43608.pdf [Accessed 20 April 2018]

Daskalakis, G. and Markellos R.N., Symeonidis, L. (2010). Does Weather Affect Stock Market Volatility?, *Finance Research Letters*, vol. 7, no 4, pp. 214-223

Economist (2007). Algorithmic Trading – Ahead of the Tape, the best newsreaders may soon be computers. *Economist*, 21 June, Available Online: https://www.economist.com/finance-and-economics/2007/06/21/ahead-of-the-tape [Accessed 19 May 2018]

Economist (2010). What caused the flash crash? One big, bad trade. 1 October. Available online: https://www.economist.com/blogs/newsbook/2010/10/what_caused_flash_crash [Accessed 14 April 2018]

European Parliament (2010). Report on regulation of trading in financial instruments – dark pools' etc., Committee on Economic and Monetary, Available Online: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-

//EP//NONSGML+REPORT+A7-2010-0326+0+DOC+PDF+V0//EN [Accessed 9 May 2018]

Fama, E.F., French, K. (1988). Dividend Yields and expected stock returns, *Journal of Financial Economics,* vol. 22 no. 1 pp. 3-25

Fama, E.F., French, K. (1992). The cross-section of expected stock returns. Journal of Finance vol. 47 no. 2 pp. 427-465

Gomber, P., Arndt, B., Lutat, M. & Uhle, T. (2011). High-Frequency Trading, Goethe Universität Available online: https://deutsche-boerse.com/blob/2532148/050b9a4c5c9a2fbd73eca72fe9c0ff73/data/high-frequency-trading_en.pdf [Accessed 14 April 2018]

Goyal, A. & Welch, I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, *Review of Financial Studies,* vol. 21 no. 4, pp. 1455-1508

Gsell, M. (2008). Assessing the Impact of Algorithmic Trading on Markets: A Simulation Approach, Center for Financial Studies (CFS), working paper, no. 2008/49, Goethe University Frankfurt

Harris, L. (2003). Trading and exchanges: Market microstructure for practitioners, *Oxford University Press*, Oxford

Hendershott, T., Jones, C.M. and Menkveld, A.J. (2011). Does Algorithmic Trading Improve Liquidity?, *The Journal of Finance*, vol. 66, no 1, pp. 1-33

Gangahar, A. (2008). Algoritmic Trading Heightens Volatility, *Financial Times*, 4 December. Available Online: https://www.ft.com/content/3f57311e-c246-11dd-a350-000077b07658 [Accessed 7 March]

Groth, S. (2011). Does Algorithmic Trading Increase Volatility? Empirical Evidence From the Fully-Electronic Trading Platform Xetra, Mimeograph, Available Online: http://aisel.aisnet.org/wi2011/112 [Accessed 14 April]

Lettau, M., & Ludvigsson, S. (2001). Consumption, Aggregate Wealth and Expected Stock Returns, *The Journal of Finance,* vol. 56 no. 3 pp. 815-849

Jain, P.K. & Johnson, W.F. (2008). Trading Technology and Stock Market Liquidity: A Global Perspective, Mimeograph, Available Online: https://ssrn.com/abstract=952322 [Accessed 16 April 2018] Mimiograph

Jovanovic, B., Menkveld, A.J. (2010). Middlemen in Limit-order Markets, Department of Economics, New York University. Available Online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.362.9469&rep=rep1&type=pdf [Accessed 14 April 2018]

Kaya, O. (2016). High-frequency Trading - Reaching the limits [pdf]. Available at: https://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000454703/Research_Briefing%3A_High-frequency_trading.pdf [Accessed 10 April 2018]

Kirilenko, A., Samadi, M., Kyle, A., Tuzun, T. (2017). The Flash Crash: High-Frequency Trading in an Electronic Market, *The Journal of Finance,* vol. 72 no. 3 pp. 967-998

Kothari, S. & Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: a time-series analysis. *Journal of Financial Economics*, vol. 44, no. 2, pp. 169–203

Lamont, O. (1998). Earnings and expected returns. *Journal of Finance*, vol. 53, no. 5, pp. 1563–1587

Liu, W.M. (2009). Monitoring and limit order submission risks, *Journal of Financial Markets* vol. 12, pp. 107–141

Myer, G., Bullock, N. and Rennison, J. (2018). How high-frequency trading hit a speed bump, *Financial Times,* 1 January, Available online: https://www.ft.com/content/d81f96ea-d43c-11e7-a303-9060cb1e5f44 [Accessed 5 March 2018]

Nasdaq OMX (2010). *2009 Annual Report*, Available Online: http://www.shareholder.com/visitors/DynamicDoc/document.cfm?DocumentID=2793&CompanyID=NDAQ&zid=bd26897d [Accessed 25 April 2018]

Philips, M. (2010). High Frequency Trading and the Roiling Markets, *Newsweek*, 31 May, Available online: http://www.newsweek.com/high-frequency-trading-and-roiling-markets-73147 [Accessed 15 May 2018].

Pole, A. (2007). Statistical Arbitrage, Algorithmic Trading Insights and Techniques, *Wiley Finance,* Vol. 1, pp. 13

Pontiff, J. & Schall, L. D. (1998). Book-to-market ratios as predictors of market returns. *Journal of Financial Economics*, Vol. 49, issue 2, pp. 141–60

Prestbo, J. (2011). S&P 500 and the Dow: Tale of two markets, *Marketwatch,* 1 Januari, Available online:https://www.marketwatch.com/story/sp-500-and-dow-tale-of-two-markets-2011-01-21 [Accessed 5 May 2018]

Shiller, R. J. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*, vol. 2, pp. 457–98

Tradeworx (2010). Tradeworx, Inc. Public Commentary on SEC Market Structure Concept Release. Comments on Concept Release on Equity Market Structure [Release No. 34-61358; File No. S7-02-10]. Available online:
www.sec.gov/comments/s7-0210/s70210-129.pdf
[Accessed 17 May 2018]