A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

Investor Heuristics Measurement and Return Predictability – a Behavioural Finance Study

FÁBIO RAFAEL MORAIS ZEFERINO

Work project carried out under the supervision of:

Professor Gonçalo Sommer Ribeiro

January 2021

Abstract

This paper presents literature-based investor heuristic measurement variables to explain and predict excess returns in the U.S market. These variables are part of a behavioural model that aims to measure the anchoring, availability, confirmation, overconfidence, and representativeness heuristics. Empirical evidence, based on the NASDAQ100 index, suggests that the behavioural model is able to explain excess returns and that it can be incorporated in the Fama-French Three-Factor Model (hybrid model) to enhance the traditional models' explanatory capabilities of regular stock returns. Lastly, this paper presents several in- and out-of-sample forecasts that support the return predictability of the behavioural and hybrid models.

***Keywords**: Behavioural Finance, Anchoring bias, Availability bias, Confirmation bias, Overconfidence bias, Representativeness bias, Heuristics Measurement, Return Predictability*

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

Modern portfolio theory was propelled by the stock market crash in 1929, the subsequent financial disclosure requirements, and the publications of *Security Analysis* (Graham and Dodd, 1934) and *The Theory of Investment Value* (Williams, 1938) (Kahn, 2018). Portfolio Selection (Markowitz, 1952) marked the beginning of efficient markets, rational investors, and available information era, later accompanied by other ground-breaking financial theories such as, The Capital Asset Pricing Model (Sharpe, 1964; Lintner, 1965; Mossin, 1966; Black, 1972), Efficient Market Hypothesis (Fama, 1970), Option Pricing Theory (Black and Scholes, 1973), and Arbitrage Pricing Theory (Ross, 1976). These asset-pricing models are part of the finance neoclassical approach, which dominated the academy until the 1970s. However, in the 1980s, it was noticed that efficient market models' predictions were not able to explain the observed stock excess volatility, which, if true, could question the foundations of efficient market theory (Shiller, 2003).

Financial theorists initiated empirical investigations on market efficiency and unveiled market anomalies (Banz, 1981; LeRoy and Porter, 1981; Shiller, 1981; Basu, 1983; Roll, 1983; De Bondt and Thaler, 1985; 1987; Stein, 1989; Bates, 1991) and market puzzles (Mehra and Prescott, 1985; Lee, Shleifer, and Thaler, 1991; Fama and French, 1992). *Prospect Theory: An Analysis of Decision under Risk* (Kahneman and Tversky, 1979) consolidated some behavioural finance concepts, which will be further scrutinized in the Literature Review, with empirical evidence on systematic irrational behaviour.

Efficient market hypothesis, or, neoclassical finance, believes that securities' prices fully reflect their intrinsic value and investors are, therefore, perfectly rational agents. Behavioural finance, on the other hand, proposes that investors suffer from heuristics[1] and

[1] Heuristics can be defined as cognitive shortcuts to help individuals process information by simplifying it, even if it includes disregarding it (Marewski and Gigerenzer, 2010).

cognitive bias[2], which means that sometimes securities are mispriced due to investors' irrational behaviour (Samson, 2020). Neoclassical theorists are reluctant in accepting behavioural finance mainly because of its disjoint collection of studies (Dimson and Mussavian, 1999) and, thus, lack of coherent theories, and because of the use of *ex-post* justifications for effects that had already been uncovered by traditional finance (Kahn, 2018).

Consequently, this paper has three main aims: to gather enough literature on investors' irrational behaviour to create a coherent theory based on a joint collection of studies; to use those studies to build a behavioural model, based on heuristic measurement, that is potentially capable of explaining current excess returns and predicting future ones, using ex-ante data; and, lastly, to assess if this behavioural model can be complemented with the traditional Fama-French Three-Factor Model (FFTFM) in explaining and predicting regular stock returns.

This paper is structured as follow: Section 2 provides a literature review for each of the main heuristics, presenting initial insights on past uncovered measurements, followed by Section 3, which deepens the heuristics' methodology. Section 4 presents and discusses the results of the robust linear regression and the in- and out-of-sample forecasts in line with previous studies. Section 5 introduces the main findings, identified limitations and future research recommendations.

2. Literature Review

Kahneman and Tversky are unarguably the founders of behavioural finance. The studies conducted through observations of human behaviour in the military and psychology experiments on undergraduates conceptualized most of the existing behavioural concepts. Even

[2] Cognitive biases are defined as systematic cognitive deviations from standard formal logic (Samson, 2020).

though the research was not directed at real investors, the implications of human behaviour are universal and can, therefore, be extrapolated to other fields (Kahn, 2018).

The comprehension and construction of this papers' model are dependent on the understanding of Kahneman and Tversky's central contributions to behavioural finance, which can be divided into two fundamental theories. Initially, with the inception of the Prospect Theory, they empirically demonstrated that individuals tend to underweight events that are possible in comparison to those that are certain, thereby criticizing the expected utility theory. Hereafter expected utility would be a function of a valuation function, which derives from a reference point and not from final outcomes, and a weighting function, which derives from decisions weights rather than actual probabilities or degrees of belief. Contrary to prior beliefs, individuals applied alternative weights to gains and losses, favouring the probabilistic terms of the upside, alternatively known as loss-aversion. The Prospect Theory also implied that individuals' decisions were determined by several psychological factors. Therefore, framing was conceptualized, even though it was previously discussed in the Prospect Theory. Framing implies that the perception of a problem and the probabilistic evaluation process can be altered depending on the connotations with which the options are presented. Consequently, quantitative studies based on questionnaires or other similar-type approaches will always have some degree of framing, whether intentionally or not, which makes it less viable to draw unbiased results. Therefore, heuristics quantitative data indicators, such as the ones further presented in Methodology, are, besides innovative, more reliable, and replicable.

Nevertheless, such influential and innovative theories are always prone to criticism. According to McDermott's article (2001), there are two main criticism worth mentioning. Tversky's experiments do not replicate real-life problems, rather, these experiments are cleverly constructed, and thus, the subjects "cognitive biases" reflect the experiments' difficulty and selectivity (Gigerenzer and Murray, 1987; Gigerenzer, 1991, 1996). Tversky is accused of miss-utilizing the subjects adaptiveness capabilities in decision making. By focusing on single events or unfamiliar tasks, the subjects were misled to worse results (Brade et al., 1998; Rose et al., 1999). Despite the relevance of these objections, the theories still seem to hold, which means that only the study’s precision is being questioned, not the consequences. Thus, it is still pertinent to study the heuristics and cognitive bias that result from the Prospect Theory.

Due to the dominant role that the psychological factors play in the decision-making process of investments selection, several biases that affect human behaviour have been unravelled. The conjunction of existing literature revealed the predominance of five behavioural biases, namely: anchoring bias (Kahneman and Tversky, 1974; George and Hwang, 2004; Shefrin, 2010; Samson, 2020), availability bias (Kahneman and Tversky, 1973; 1974; Shefrin, 2010; Samson, 2020), confirmation bias (Nickerson, 1998; Shefrin, 2010; Samson, 2020), overconfidence bias (Belsky and Gilovich, 1999; Barber and Odean, 2000; Shefrin, 2010; Samson, 2020) and representativeness bias (Kahneman and Tversky, 1974; Shefrin, 2010; Samson, 2020).

Anchoring bias emerges from the attachment of a reference point in an individual’s judgement (Kahneman and Tversky, 1974). Studies have proven that even when the subjects are alerted to their anchoring’s dissociation with the problem, its effects are still persistent. Anchoring causes investors to be more reluctant to buy (sell) assets (Colin and George, 2004; Li and Yu, 2012; Chang et al., 2017) or to incorporate good (bad) news (George and Hwang, 2004) when current prices are close to historical highs (lows), thereby influencing investors trading patterns (Grinblatt and Keloharju, 2001). The 52-week high ratio, further developed in Methodology, has been consistently used to proxy for the anchoring bias (George and Hwang, 2004; Driessen et al., 2010; Hur and Singh, 2019). Empirical findings have shown that investors tend to buy stocks when their prices are continuously and rapidly decreasing or continuously surpassing historical highs (Abdin et al., 2020).

Availability bias is a form of probabilistic distortion caused by the facility with which the information is available. Kahneman and Tversky (1974) revealed that investors will evaluate the quality of a stock based on recent or easily recallable information, disregarding other relevant facts and past information. Barber and Odean (2008) revealed that investors will buy the most recent attention-grabbing stocks, thus, stock-picking correlates with the stock's recent attention. Even though there is no direct relationship between the bias and these two measures, Kristoufek (2013), and Bollen at al. (2012), Ranco et al. (2015), and Nisar and Yeung (2018) have used Google Trends and Twitter, respectively, as potential capable tools for measuring attention-grabbing stocks, later scrutinized in Methodology.

Confirmation bias occurs when individuals overweight information that confirms their prior beliefs and preconceptions, and underweight or completely disregard potential disconfirming information (Nickerson, 1998; Hart et al., 2009; Shefrin, 2010; Charness and Daves, 2017). Unfortunately, in line with Cafferata and Tramontana (2019), the literature on empirical demonstrations of confirmation bias is scarce. Nonetheless, Park et al. (2013) proved the existence of confirmation bias in investors through stock message boards[3]. Pouget et al. (2016), besides empirical evidence on this bias, presented novel predictions regarding analysts' opinion dispersion and behaviour, also deepened in Methodology, and a unified rationale to explain excess volume, excess volatility, and momentum.

Overconfidence results from a significant divergence between ones' perceived capabilities and actual performance, which bestows an illusion of control and information

[3] Message boards are online discussion websites where users and readers can interact with one another in a specific topic of interest, which, in this case, are stocks.

precision causing incredulity to others' beliefs (Belsky and Gilovich, 1999; Barber and Odean, 2000). The pioneering Daniel, Hirshleifer and Subrahmanyam (DHS) model (Daniel et. al, 2002) empirically demonstrated that overreaction is continuous as investors overreact to their initial private information, and, due to self-attribution bias, their overconfidence is further enhanced on the subsequent arrival of information, which generates a momentum factor. Despite several attempts to explain market excess trading volume[4], overconfidence seems to be the most predominant explanation for the active investing puzzle (Daniel and Hirshleifer, 2015). Consequently, excess trading volume (Campbell et al., 1993; Hirshleifer, 2001; Agarwal et al., 2008; Hsu and Shiu, 2010; Boussaidi, 2013; Byun et al., 2016; Raharja et al., 2017; Yang et al., 2018) and momentum (Daniel et al., 2002; Hou et al., 2009; Asem and Tian, 2010) comprise the main indicators for investor overconfidence proxy.

Representativeness arises when individuals over rely on stereotypes and thus misjudge an events' probability of occurrence by comparing it with a similar event. This process can result from base rate neglect or sample size bias. In the former, individuals neglect the probabilities of an event and consequently have underconfident or overconfident assumptions relative to their initial events' probability. In the latter, individuals, convinced that the sample is representative and significant, fundament their information and beliefs on inadequate data samples (Kahneman and Tversky, 1974; Khan et al., 2017). Ganzach (2001) found that the analysts' perceived risk and return of unfamiliar stocks depend on their global attitude: high (low) returns and low (high) risk will be attributable to stocks that are perceived as good (bad). Empirical studies (Shefrin, 2010; Kelly et al., 2011; Bordalo et al., 2019) indicate that analysts' recommendations could be utilized as a proxy for investors' representativeness.

[4] Excess trading volume, or active investing puzzle, corresponds to the phenomenon of aggressive trading. This strategy consists of high trading volume, which results in higher risk and lower net returns due to increased trading costs. (Daniel and Hirshleifer, 2015)

Besides these predominant heuristics, the joint body of literature consistently addressed some influential departing factors such as, idiosyncratic volatility (Ang et al., 2006; Peterson and Smedema, 2011), retail investors' sentiment (Black, 1986) and retail investors' presence in the financial markets (Odean, 1999). Idiosyncratic volatility measures the variations of returns that are not captured by asset-pricing models. With the Prospect Theory it is possible to explain the market anomaly, since investors value stocks differently from the traditional Bayesian rules. Similarly, Black (1986) discovered a noise factor that caused markets to be inefficient, but, simultaneously, it disabled the market from taking advantage of arbitrage opportunities due to market constraints. This noise was driven by the market expectations (retail investors' sentiment) of noise traders (retail traders). Later, it was found that retail investors have a high degree of overconfidence, which leads to excess trading (retail investors' presence in the financial markets) (Odean, 1999). Investors are reluctant to sell short and quickly sell after small profits (loss-aversion), this irrational behaviour, more predominant in retail investors, has an important influence in the market (Odean, 1999).

Despite the topics' controversy, prior literature suggests that retail investors are especially susceptible to heuristics when compared to institutional investors (Odean, 1999; Grinblatt and Keloharju, 2001; Goetzmann and Massa, 2002; Schmeling, 2007; Barber and Odean, 2008; Franzinni and Lamont, 2008; Hur et al., 2010; Vrouenraets, 2011; Khan et al., 2017), which makes it a crucial indicator, as high retail investor presence can potentially exacerbate the effect of heuristics. Some of the distinct approaches to measure retail investors' presence (1) and sentiment (2) are, (1) New York Stock Exchange Trade and Quotes (TAQ) (Campbell et al., 2005; Barber et al., 2008; Kaniel et al., 2008), institutional ownership (Hwang et al., 2016) and 13(f) reports (Sias and Whidbee, 2010; Vrouenraets, 2011), and (2) Sentix (Schemling, 2007), American Association of Individual Investors (AAII) (Verma et al., 2008; Burghardt, 2010), and Investors Intelligence (Verma et al., 2008).

3. Data and Methodology

The data focuses on the U.S market, particularly on the constituents of the NASDAQ 100 index, from March 2011 to July 2020 in a daily and weekly frequency. The NASDAQ 100 index was the preferred index because it is the most volatile, due to its high-growth technological stocks, which could mean that it is highly influenced by behavioural factors. The companies were preventively selected as of 2011, through Bloomberg, to avoid survivorship bias. Only 66 companies were selected since the others had to be excluded due to inactivity or lack of data. On the other hand, the time-horizon was chosen in accordance with the data's availability and convergence.

The data for closing and opening prices, trading volume, highs and lows, and market capitalization and book value of equity were retrieved from Refinitiv Reuters. Returns were calculated in logarithmic form instead of arithmetic. Log-returns are usually considered to be normally distributed, which is one of the main assumptions of the Ordinary Least Square (OLS) regression. Consequently, returns were calculated as $log(\frac{Closing_{\ y}}{Closing_{\ y-1}})$, where $Closing_y$ is the closing price of day/week y and $Closing_{y-1}$ is the closing of the previous day/week. The OLS regression creates a linear regression between one dependent variable and one or more independent variables. The estimated relationship is done by minimizing the sum of squares, which results from the difference between the observed and the predicted values of the linear regression.

In this chapter, the variables will be organized according to the procedure that was followed in their construction and their corresponding heuristics will be identified after their introduction, Table 1 (Appendix). The first variables were constructed using only prices, volume and market capitalisation. Followed by analyst's recommendation, Twitter, and Google, and lastly, the departing factors, such as AAII, retail investor proxy and FFTFM data.

Basing the rationale on Byun et al. (2016) cross-sectional study of the signed volume ($SV_{i,t}$) method, two unorthodox variables using trading volume were developed: the retailed signalled volume, $RSV_{i,t}$, and the signalled volume with a one-day lag, $SV_{i,t}$. Because the previous study was done in a cross-sectional regression, the replication of its variable would result in forward looking bias[5]. Therefore, to signal volume, the retail investor sentiment was used instead in $RSV_{i,t} = \begin{cases} Vol_{i,t} \text{ } if \text{ } BR_{i,t} > BR_{i,t-1} \wedge BR_{i,t} > ABR_{i,t} \\ -Vol_{i,t} \text{ } if \text{ } BR_{i,t} \leq BR_{i,t-1} \vee BR_{i,t} \leq ABR_{i,t} \end{cases}$, where $BR_{i,t}$ is the percentage of bullish retail traders in the AAII and $ABR_{i,t}$ is the correspondent 8-week average (Overconfidence Bias). While on the $SV_{i,t}$, a one-day lag was applied. $SV_{i,t} = \begin{cases} Vol_{i,t-1}, \text{ } if \text{ } r_{i,t-1} > 0 \\ 0, if \text{ } r_{i,t-1} = 0 \\ -Vol_{i,t-1}, if \text{ } r_{i,t-1} < 0 \end{cases}$, where $Vol_{i,t-1} \text{ } and \text{ } r_{i,t-1}$ are stock i's trading volume and return with one day lag in respect to day/week t, respectively (Overconfidence Bias). It is expected from these variables to have a positive effect on returns as trade volume and returns are positively correlated (Byun et al., 2016). The closing prices and historical highs were used to construct the 52-week high ratio, $GH = \frac{Current \text{ } Price_{i,t}}{52-Week \text{ } High \text{ } Price_{i,t}}$ (George and Hwang (GH), 2004) (Anchoring Bias). It is expected to have a negative coefficient, since the anchoring causes reluctancy in investors to buy (sell) near historical highs (lows). As Campbell et al. (1993) suggested, a one-year backward moving average of turnover growth rate was subtracted to the log turnover series, $Turnover = log\left(\frac{Volume_{\text{ } i,t} * Current \text{ } Price_{i,t}}{Market \text{ } Capitalization_{i,t}}\right) - MA_{i,t-1y}$ (Overconfidence Bias). Its coefficient is presumably negative since, as previously explained, excess trading volume results in lower net returns. Incrementally, Ang et al. (2006) suggested that idiosyncratic volatility could be measured as, $IVOL_{i,t} = DIVol_{i,t-1}$, where $DIVoL_{\text{ } i,t-1}$ corresponds to the volatility of a stock's daily idiosyncratic returns over month_{t-1}.

[5] Forward looking bias occurs when the tests use data that would not have been available in the period being analysed. This usually leads the results to be closer to the desired outcomes of the test.

Idiosyncratic volatility should not persist after introducing other heuristic measurement variables since they should be able to price the variations of returns that are not captured by the traditional asset-pricing models, such as the FFTFM.

Quarterly data on analysts' number of estimates and their target prices' standard deviation, mean, median, high and low was extracted from I/B/E/S to calculate analysts' forecasts revisions (proxy for belief revision) and recommendations dispersion (a proxy for disagreement) (Pouget et al., 2014). These variables try to capture investors' reluctant attitude in revising their initial signals or beliefs. If investors have an initial positive (negative) belief or signal, they will be less willing to revise their beliefs downward (upward) or to pay attention to subsequent negative (positive) signals. Revisions were constructed as the log-change of the mean target price, $Revisions_{i,t} = \log\left(\frac{Target\ Price\ Mean_{i,t}}{Target\ Price\ Mean_{i,t-1}}\right)$, and analysts' dispersion is simply the standard deviation of their target prices' (Confirmation Bias). To prevent the frequency mismatch between the model, which is in a daily or weekly basis, and the analysts' data, which is in a quarterly basis, from originating missing data, a linear interpolation was constructed for $Revisions_{i,t}$. Based on Shefrin (2010), Kelly et al. (2011) and Bordalo et al. (2019) suggestion of using analysts' recommendations as a representativeness heuristic measurement, $RP_{i,t}$ was innovatively constructed as $RP_{i,t} = \frac{Closing\ Price_{i,t}}{Analysts'Recommendation_{i,t}}$. It corresponds to the closing price of stock i over its' mean analysts' recommendation price in day/week t. The rationale is that when investors receive a buy (sell) recommendation, ceteris paribus, they will disregard all other relevant information, such as the price of the security (Kelly et al., 2011).

Twitter data sentiment analysis has been widely accepted in the industry as an important indicator of market sentiment (Bollen at al., 2012; Ranco et al., 2015; Nisar and Yeung, 2018). Namely, it has been uncovered that Twitter can be used to assess the mood of the market and to correlate it with stock returns, yielding favourable results. Bollen et al. (2012) achieved an

86.7% accuracy in predicting the daily up and down movements in the closing prices of DJIA. Ranco et al. (2015) managed 1-2% abnormal returns by focusing on a Twitter sentiment strategy. Due to its importance, several platforms offer this type of data. Unfortunately, due to the inability to access most of them, data[6] on bull and bear messages, their spread and ratio, and total messages were extracted from Quantopian (Availability Bias). Unfortunately, Quantopian only yielded data until the 1st of May of 2020, thus, in order to maintain the initial time horizons' length, Twitter data was also extracted from Bloomberg from that point onward. Initially, Quantopian was preferred over Bloomberg because, within the same timeframe, Quantopian offered more information. In this case, $Bull_{i,t} = \frac{Bullish\ Messages\ Scanned_{i,t}}{Total\ Messages\ Scanned_{i,t}}$, corresponds to the fraction of stock i's bullish messages scanned over the number of total messages scanned in day/week t; $BBR_{i,t} = \frac{Bullish\ Messages\ Scanned\ _{i,t}}{Bearish\ Messages\ Scanned\ _{i,t}}$, where BBR stands for Bull and Bear Ratio, corresponds to the fraction of stock i's bullish scanned messages over the bearish ones in day/week t; finally, $Mess_{i,t} = \log\left(\frac{Total\ Messages\ Scanned_{i,t}}{Total\ Messages\ Scanned_{i,t-1}}\right)$, corresponds to the stock i's log-change of the total number of messages in day/week t.

Kristoufek (2013) proposed a novel approach to risk management by relying on Google queries to search for popular stocks, suggesting and proving that a stocks' popularity is positively correlated with its risk. Therefore, Google Trends data[7] was extracted. Google Trends provides scaled data ranging from 0, if no Google search queries were made that day/week, to 1, if, within the researched timeframe, that day/week was the one with the highest number of search queries. Risteski and Davcev (2014) revealed an interpolation technique that would enable daily data extraction from Google Trends (Availability Bias) for more than 90 days, which was the data period limit at the time. The daily period limit is now 9 months, which is

[6] Provided by Professor Qiwei Han. Assistant Professor of Data Science at NOVA School of Business & Economics.
[7] Available at https://github.com/qztseng/google-trends-daily

still minimal for the purpose of this study. Although there have been uncovered two more techniques of data extraction: the daily data method and the overlapping method, the interpolation method[8] was chosen due to its feasibility and easiness to use, besides being the second-best approach (Tseng, 2019). Thus, $Google_{i,t} = log\left(\frac{Google\ Trends\ Data_{i,t}}{Google\ Trends\ Data_{i,t-1}}\right)$ represented the log-change of Google Trends data between day/week t and t-1.

Besides Google Trends and Twitter, which can possibly analyse global market sentiment, as seen in prior literature, retail investors also represent a substantial portion of the market and, consequently, have an important role within it. Although retail traders provide liquidity to the market, they are also responsible for the mispricing of securities (Black, 1986). Hence, the AAII and Institutional Ownership, extracted from Compustat, in accordance with Verma et al. (2008) and Hwang et al. (2016), were used to obtain retail investors sentiment and presence, respectively. Institutional ownership data was removed from the analysis due to lack of data. Nonetheless, it was possible to extract both $BBS_t = \%\ \text{Bullish Retail Trades}_\text{t} - \%\ \text{Bearish Retail Trades}_\text{t}\ and\ ABR_t =\ 8 - \text{week Average of}\ \%\ \text{Bullish Retail Trades}_\text{t}$ from AAII.

Due to the inability to access TAQ database and because of the lack of data from institutional ownership, retail investor presence had to be measured in an unorthodox manner. As Vrouenraets (2011) proposed in his thesis, retail investor activity can be proxied by analysing the trading volume of the biggest online brokerage firms. In his thesis, Vrouenraets defined three conditions for the selection of the brokers: being a U.S based company, being publicly listed in the U.S and having a stable company history[9]. Since the objective is to measure the fraction of retail investors in the market, the third criterion is not of interest to this

[8] An adjustment factor: the division of a lower frequency data (weekly/monthly) by a higher frequency (daily/weekly), when they converge, is multiplied to the desired interpolated high frequency data (daily/weekly).
[9] This criterion resulted in the exclusion of TD Ameritrade. The company had made to many large mergers and acquisitions.

model, hence, the resulting brokers were Charles Schwab Corporation, E-Trade Financial Corporation, TD Ameritrade and Interactive Brokers Group Inc. After collecting the quarterly data for the number of trades from the 10-Q fillings of the aforementioned companies, the following proxy was developed, $Retail\ Proxy_t = \frac{SumofBrokers'TotalTrades_t}{SumofNasdaq100Constituients'TotalVolume_t}$. The main limitation of this indicator is the difference in frequency between the number of trades and the volume of the constituents, since, in this case, the numerator will be static for one quarter while the denominator changes weekly.

FFTFM data was extracted from Kenneth R. French – Data Library, with the exception of market premium. Since it incorporates all companies from the NYSE, AMEX and NASDAQ, when the object of study is only the companies of the NASDAQ 100 and it caused consistency issues. Additionally, instead of the value-weighted return of the NASDAQ 100, the return of the corresponding ETF, QQQ, was used to proxy market returns, so as to avoid autocorrelation with stock returns, which was then subtracted for the Fama-French risk-free rate, $Market\ Premium_t = \text{QQQ Returns}_\text{t} - \text{Fama French risk} - \text{free rate}_\text{t}$. The momentum factor also followed a similar approach to Fama-French, mainly because in Daniel et al. (2002), Hou et al. (2009) and Asem and Tia (2010) studies, momentum was used as a heuristic measurement variable. However, momentum is usually built on a single heuristic, meaning that the momentum factor is created based on a winner and loser portfolio, screened by this one heuristic. Contrarily, this model uses several heuristics, therefore that momentum construction is not replicable. Hence, $Momentum_t = DWA\ Momentum\ ETF_t - \ QQQ\ ETF_t$, where DWA_t is the momentum ETF for NASDAQ 100. Similar to $Market\ Premium_t$, for the $Momentum_t$ factor, the DWA Momentum ETF was used instead of the Fama French Momentum factor because the former refers only to the NASDAQ 100 index.

The Y of the regression is the equally-weighted average of the excess returns derived from the FFTFM and the X's correspond to the equally-weighted average of every company's data for each bias. Since the purpose of this study is to understand if behavioural biases affect companies individually and if they serve as a forecasting tool for returns, the equally-weighted average was preferred over a weighted-average. Mainly because if different weights were attributed between companies, the most predominant ones would skew the regressions' resulting coefficients. To test the significance of the regression after accounting for size and book-to-market, the model was further divided into the 25% and 75% percentiles of market capitalisation and book-to-market, $\mathrm{BM}_{\mathrm{i,t}} = \frac{\text{Book value of equity}_{\mathrm{i,t}}}{\text{Market Capitalisation}_{\mathrm{i,t}}}$ (Fama and French, 1992), respectively. Furthermore, since the weekly data was always referent to Friday, an additional model was successfully built with data referent to Monday, with the intent to rule out any forward looking bias. See Table 1 (Appendix) for an overview on each bias.

4. Results

This study endeavoured to (1) investigate the explanatory power of heuristics relative to excess returns (behavioural model) and a possible conjunction between heuristics and the FFTFM (hybrid model), and to (2) investigate the predictability power of both the behavioural and hybrid models. For this purpose, this section will be divided into two main segments, the linear regression results, and the forecast results, for both the daily and weekly models.

Regression Results – Behavioural Daily Model

The models begin with the construction of the excess returns of FFTFM, Small Minus Big (SMB), High Minus Low (HML) and Market Premium, as previously defined. In the daily model it was redundant to create the excess returns derived from the FFTFM since the regression suffered from autocorrelation. When the lags were applied to correct for the phenomenon, the model became unapt, Table 2 (Appendix). Additionally, the coefficients and t-statistics were corrected with the Huber/White/sandwich test since the models suffered from heavy heteroskedasticity, as depicted in Figure 1 (Appendix).

The models will be composed of the 15 variables (1) that were previously presented in the Literature Review and Methodology.

$$(1)\ Y = Momentum_t + SV_{i,t} + RSV_{i,t} + GH_{i,t-1} + Turnover_{i,t} + IVOL_{i,t} + Dispersion_{i,t} + Revisions_{i,t} + RP_{i,t} + Bull_{i,t} + BBR_{i,t} + Mess_{i,t} + ABR_t + BBS_t + Google_{i,t} + Retail\ Proxy_t + e_{i,t}$$

After observing the null adjusted R-squared, excess returns were deemed inappropriate for the daily model and were then substituted with standard stock returns. Because excess returns are inappropriate and since $IVOL_{i,t}$ measures the volatility of idiosyncratic returns, the variable was excluded à priori from the model. $Retail\ Proxy_t$ was also removed due to the frequency gap between quarterly data from the brokers and the daily data of the volume. Due to the unfeasibility of calculating the signal with the weekly data of the AAII, $RSV_{i,t}$ was removed. $Turnover_{i,t}$, $Revisions_{i,t}$, $BBR_{i,t}\ and\ BBS_t$ were also excluded after constructing the correlation matrix, Table 3 (Appendix).

Variables	Results
$Momentum_t$	-0.156* (0.09)
$SV_{i,t}$	2.15E-10* (0.01)
$GH_{i,t-1}$	-0.023*** (2.79E-03)
$Dispersion_{i,t}$	-0.004 (4.37E-03)
$RP_{i,t}$	0.033*** (3.21E-03)
$Bull_{i,t}$	-0.004 (0.01)
ABR_t	-0.002 (2.31E-03)
$Mess_{i,t}$	0.001** (4.27E-04)
$Google_{i,t}$	-0.085* (0.05)
Constant	-0.009* (0.01)
R-Squared	7.63%
Durbin-Watson Test	2.10
Mean VIF	1.09

*Table 4 – Daily Heuristic Robust Linear Regression. *, **, and *** represent significance at 90%, 95% and 99% confidence, respectively.*

The daily regression yielded unfavourable results, with an R-Squared of 7.63%, Table 4. Indicating that these behavioural biases may not be relevant to explain returns on a daily frequency, even though the coefficients of $Momentum_{i,t}$, $SV_{i,t}$, $GH_{i,t-1}$, $RP_{i,t}$,

$Mess_{i,t}$ and $Google_{i,t}$ remained significant. One important note that will become clearer later on, is that, in this time frequency, $Google_{i,t}$ has a significant, at 90% confidence, and negative coefficient (-0.085), which proves Kristoufek (2013) proposition that famous companies usually have lower returns and higher risk.

Normally, daily frequency data is overlooked because lower frequencies for returns are usually, approximately normally distributed. The value of the normality assumption is not in its ability to replicate reality, but rather in the support that it grants the model to provide useful insights (Eugene Fama, 1976). Since the daily model was incapable of explaining the variations of returns, the construction of the hybrid model and the forecasts for daily basis were disregarded.

Regression Results – Behavioural Weekly Model

In the weekly model, Table 5 (Appendix), the FFTFM does not suffer from multicollinearity, demonstrated by the Variance Inflation Factor of 1.68, or autocorrelation, depicted by the Durbin Watson test equal to 1.02, therefore the excess returns were created.

Table 6 (Appendix) depicts the correlation matrix between the existing variables. $GH_{i,t}$ seems to be the main source of the problem, therefore, one additional lag was added. Despite the existence of other high correlation levels between variables (>0,5), the model was regressed, Table 7. Even then, most of the coefficients were still non-significant. Nonetheless, before pre-emptively removing these variables, the model was tested for market size and book-to-market ratio, specifically in the 25% and 75% percentiles of market capitalisation and book-to-market ratio, respectively. As Table 7 reveals, the predominant explanatory coefficients were of $RSV_{i,t}$, $SV_{i,t}$, $GH_{i,t-1}$, $Dispersion_{i,t}$, $RP_{i,t}$, $Bull_{i,t}$, $BBR_{i,t}$, ABR_t, BBS_t $and\ Retail\ Proxy_t$.

	Full Sample	25% Market Capitalisation	75% Market Capitalisation	25% Book-to-Market	75% Book-to-Market
$Momentum_{t}$	-0.041	-0.208*	0.014	-0.166	0.034
	0.10	(0.11)	(0.15)	(0.15)	(0.11)
$RSV_{i,t}$	3.00E-04***	3.90E-13**	7.52E-12***	2.17E-12***	8.58E-13***
	(1.03E-04)	(1.54E-13)	(2.06E-12)	(7.57E-13)	(2.38E-13)
$SV_{i,t}$	1.22E-10***	2.07E-11**	3.84E-10***	2.03E-10***	2.39E-11
	(3.23E-11)	(9.34E-12)	(1.15E-10)	(3.90E-11)	(2.89E-11)
$GH_{i,t-1}$	-0.152***	-0.162***	-0.078***	-0.095***	-0.162***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
$Turnover_{i,t}$	-0.024	0.022	-0.136***	-0.049	-0.026
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
$IVOL_{i,t}$	-0.614	-1.195	-1.351	-1.658**	0.335
	(0.73)	(0.83)	(0.88)	(0.78)	(0.79)
$Dispersion_{i,t}$	0.065***	-0.020***	0.103*	0.154***	0.055***
	(0.01)	(0.01)	(0.06)	(0.05)	(0.01)
$Revisions_{i,t}$	-0.002	0.046	0.022	0.105	0.110**
	(0.18)	(0.07)	(0.08)	(0.07)	(0.05)
$RP_{i,t}$	0.182***	0.165***	0.096***	0.099***	0.189***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
$Bull_{i,t}$	0.026***	0.081***	0.006	0.036**	0.031***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
$BBR_{i,t}$	-1.33E-04***	-2.55E-04***	-1.90E-04	-3.13E-04***	-1.06E-04***
	(4.01E-05)	(3.13E-05)	(1.58E-04)	(7.50E-05)	(4.00E-05)
ABR_{t}	0.027***	0.022***	0.030***	0.030***	0.024***
	(6.69E-03)	(0.01)	(0.01)	(9.81E-03)	(8.88E-03)
BBS_{t}	-0.008***	0.002	-0.012**	-0.011***	-0.005
	(3.29E-03)	(3.24E-03)	(4.89E-03)	(4.70E-03)	(3.63E-03)
$Mess_{i,t}$	0.006	0.006	0.010**	0.013***	0.004
	(3.84E-03)	(3.78E-03)	(4.25E-03)	(4.64E-03)	(3.69E-03)
$Google_{i,t}$	-0.010	0.087**	0.018	-0.002	0.013
	(0.04)	(0.05)	(0.03)	(0.03)	(0.04)
Retail Proxy $_{t}$	-0.045***	-0.007	0.003	-0.015**	-0.040***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	-0.042***	-0.024	0.084	0.019	-0.025
	(0.01)	(0.03)	(0.04)	(0.04)	(0.04)

Table 7 – Results from weekly excess return regressions of Full Sample, 25% and 75% percentile market capitalisation and book-to-market.

$Google_{i,t}$'s coefficient was not statistically significant, contrary to Kristoufek (2013), who found Google Trends to be a promising tool for risk diversification. Nonetheless, the results may be skewed due to the data interpolations that were forcibly constructed. Google Trends might still be a promising tool if used in a different time frequency or a shorter time-horizon, as previously demonstrated, in which it is possible to extract data without manipulating it. Similarly, $Revisions_{i,t}$, suffered linear interpolations to have complete data, which might have distorted the true significance of the variable.

$Mess_{i,t}$'s coefficient was not significant, which was not surprising as the variable had weak support within the literature. On the contrary, both coefficients of $Turnover_{i,t}$ and $Momentum_{t}$ were surprisingly non-significant. Alternatively, $Momentum_{t}$'s coefficient would have been significant if the Fama-French Momentum factor was used, however, it would have been unsuitable considering that the NASDAQ 100 is the index of study, rather than all NYSE, AMEX and NASDAQ firms.

Table 8 reveals the levels of correlation between the remaining variables and it points out some extremely high levels of correlation, which resulted in the exclusion $Retail\ Proxy_t$, $GH_{i,t-1}$, $BBR_{i,t}$ and BBS_t.

	RSV i,t	SV i,t	GH i,t-1	Dispersion i,t	RP i,t	Bull i,t	BBR i,t	ABR t	BBS t	Retail Proxy t
RSV i,t	1.0000									
SV i,t	0.1155	1.0000								
GH i,t-1	-0.0280	0.0175	1.0000							
Dispersion i,t	0.0717	-0.0657	-0.2325	1.0000						
RP i,t	0.1644	0.1638	0.6685	0.2390	1.0000					
Bull i,t	0.0273	0.1322	-0.0085	-0.1660	-0.0213	1.0000				
BBR i,t	0.0679	0.0126	0.1279	0.2084	0.2687	0.5867	1.0000			
ABR t	-0.0952	0.0545	0.4110	-0.1900	0.1785	0.0065	-0.2156	1.0000		
BBS t	0.4029	0.1083	0.4909	-0.1724	0.4288	0.0322	-0.0530	0.5739	1.0000	
Retail Proxy t	0.0166	-0.0393	-0.3074	0.7490	0.1451	-0.1814	0.0576	-0.2579	-0.2757	1.0000

Table 8 – Correlation Matrix of the predominant weekly heuristic indicators

$Retail\ Proxy_t$, was a valuable variable as it had a negative coefficient in all of the models, Table 7, and it had high positive correlation (0.7490) with $Dispersion_{i,t}$. The negative sign could corroborate the idea that when retail investors entail a higher presence in the market, since they are more prone to fall under heuristics (Barber and Odean, 2008), returns decrease. The correlation with $Dispersion_{i,t}$ could indicate a higher presence of retail investors in overvalued stocks, which happens when analysts' forecast dispersion is high (Hwang et al., 2016). $IVOL_{i,t}$ also yielded interesting results, despite its non-significance. Namely because, as Peterson and Smedema (2011) suggested, the variable had a negative correlation with forecast revisions. The $GH_{i,t-1}$ had high negative correlation (-0,7004) with $IVOL_{i,t}$ because, as Driessen et al. (2010) discovered, both betas and volatility decrease when approaching a historical high or low and increase when that high or low is broken. Additionally, as suggested by Hur and Singh (2019), high analysts coverage provides greater availability of information. Since stocks whose prices are further (closer) from their historical high experience more overpricing (under-pricing), due to informational reasons, there will be a positive correlation between analysts forecast revisions and the $GH_{i,t-1}$, as it was observed. Lastly, since the coefficient was negative and significant in all models at 99% confidence, Table 7, albeit the autocorrelation, the theory that investors become reluctant to incorporate good news and new information into their stock valuation when near 52-week high (George and Hwang, 2004) was confirmed.

After the removal of these variables, the model was regressed once again to test the significance of the remaining coefficients, Table 9. ABR_t had the only non-significant coefficient, most likely due to the removal of BBS_t since they were highly correlated, whereas the rest of the coefficients are jointly significant at 95% confidence.

Variables	Results
$RSV_{i,t}$	4.24E-04*** (1.04E-04)
$SV_{i,t}$	2.15E-10*** (3.85E-11)
$Dispersion_{i,t}$	0.074*** (0.01)
$RP_{i,t}$	0.039** (0.02)
$Bull_{i,t}$	0.019** (0.01)
ABR_t	-4.10E-04 (0.01)
Constant	-0.046*** (0.01)
R-Squared	34.17%
Durbin-Watson Test	2.01
Mean VIF	1.11

Table 9 – Robust Weekly Linear Regression of Heuristics

ABR_t and BBS_t coefficients' statistical non-significance was not expected, but not surprising either, as the AAII is considered a contrarian indicator (Burghardt, 2010).

On the other hand, $Bull_{i,t}$ had a positive (0.019) and significant, at 95% confidence, coefficient. The rationale is that when Twitter sentiment is increasingly bullish then market returns are prone to increase (Bollen et al., 2012; Nisar and Yeung, 2018). $Dispersion_{i,t}$ had a positive (0.074) and significant, at 99% confidence, coefficient, which confirms Diether et al. (2002) and Johnson (2005) study, that demonstrates that investors tend to overvalue stocks with high levels of analysts' forecast dispersion. $SV_{i,t}$ and $RSV_{i,t}$ coefficients were both very small but positive (<0.001) and significant, at 99% confidence. When volume increases, whether signalled by market returns or retail investors' sentiment, returns will consequently raise (Byun et al., 2016). $RP_{i,t}$ had a positive (0.039) and significant, at 95% confidence, coefficient, which could imply that when prices move closer to the analysts' recommendations mean, representativeness is enhanced and returns increase.

Overall, despite the non-significance of some coefficients, the model provides vast conclusions regarding heuristics' measurement. The replication of all the variables for each bias supported in the literature would be impractical. However, the conjunction of the literature for each bias provided a potential variable of convergence. An in-depth study of these convergence variables in each bias could provide better indicators for the model. Nonetheless, despite the data that had to be manipulated or which had to be removed due to multicollinearity, the coefficients were mostly significant and in line with the literature.

Regression Results – Hybrid Weekly Model

The model seems to suggest that the behavioural model can explain the excess returns on a weekly basis. As previously stated, this paper also aims to incorporate the behavioural model in the FFTFM. In this case, the Y is the equally-weighted average of regular stocks returns. The correlation matrix between the two models led to the removal of $Dispersion_{i,t}$, Table 10 (Appendix), due to its -0.6485 correlation with $Market\ Premium_t$. According to Park (2005), analysts' recommendation dispersion can be a measure of the difference between investors' expectations rather than a measure of risk-aversion. This could imply that when $Dispersion_{i,t}$ is high, there is less confirmatory bias in the market, expectations become less favourable and the $Market\ Premium_t$ will be lower. The model was successfully combined as only $Bull_{i,t}$'s coefficient lost its significance, Table 11. Nisar and Yeung (2018) found a weak correlation between Mood, a Twitter sentiment variable, and FTSE 100 returns. The merger between the FFTFM and the behavioural biases can boost the R-Squared obtained by the FFTFM alone from 35.30% to 50.10%, without autocorrelation, heteroskedasticity or multicollinearity issues.

Variables	Results
SMB $_t$	0,002*** (3,85E-04)
HML $_t$	0,001* (3,70E-04)
Market Premium $_t$	0.246*** (0,03)
RSV $_{i,t}$	4,78E-04*** (1,12E-04)
SV $_{i,t}$	2,03E-10*** (3,89E-11)
RP $_{i,t}$	-0,064*** (0,01)
Bull $_{i,t}$	0,006 (0,01)
ABR $_t$	-0,015** (0,01)
Constant	-0.049*** (0.01)
R-Squared	50.10%
Durbin-Watson Test	1.73
Mean VIF	1.07

Table 11 – Weekly Hybrid Robust Linear Regression

Forecast Results

As proposed, $RSV_{i,t}$, $SV_{i,t}$, $Dispersion_{i,t}$, $RP_{i,t}$, $Bull_{i,t}$ and ABR_t will also be tested for their predictability power. The forecast will be split between an in-sample forecast, using the fitted values of the robust linear regression, and an ex-ante out-of-sample forecast, using an ARCH/GARCH model. Additionally, the results will always be divided between the behavioural biases model, which will only use the heuristics, and the hybrid model, which will merge the FFTFM with the heuristics.

In-Sample Forecast Results

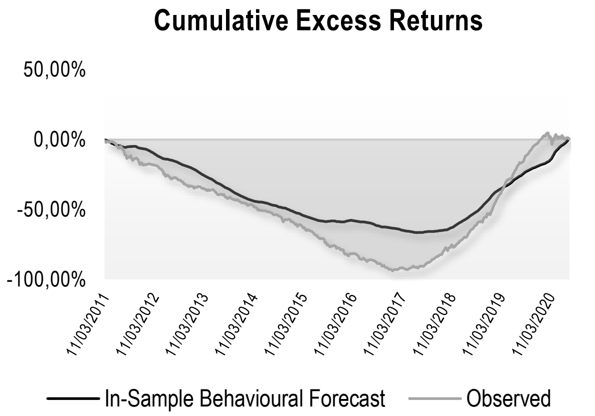
The in-sample forecasts, Table 12, presents considerably high accuracy in predicting both the upside (bull) and downside (bear) returns of the market, ≥ 70% in both models. Simultaneously, the Root Mean Squared Errors (RMSE) is approximately 0,00% for both models and the Mean Absolute Errors (MAE) is stable at around 0,50%. The predicted returns of both models in the 25%, 50% and 75% percentiles correspond to less than half of a standard deviation of the observed returns. The predicted return percentiles 0% and 100%, on the other hand, correspond to two-to-three times a standard deviation of the observed returns. This phenomenon can be observed

Indicators	Excess Return Model	Equally Weighted Return Model
Accuracy Bull	75,31%	77,16%
Accuracy Bear	74,19%	69,15%
MAE	0,52%	0,58%
RMSE	0,00%	0,00%
Standard Dev. Obs. Returns	0,89%	1,11%
Percentile 0%	-1,90%	-3,77%
Percentile 25%	-0,28%	-0,33%
Percentile 50%	0,01%	0,20%
Percentile 75%	0,38%	0,63%
Percentile 100%	1,39%	2,24%

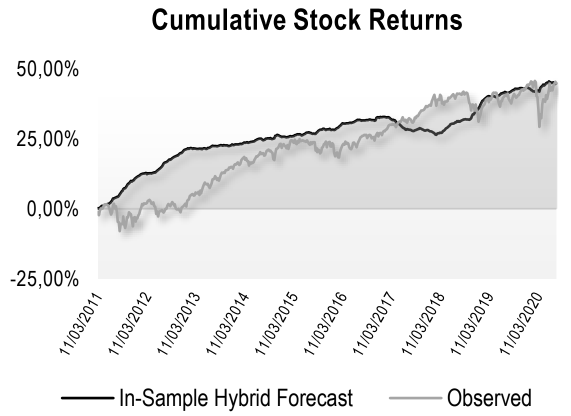
Table 12 – In-Sample forecast results for the behavioural model (left) and the hybrid model(right)

in Graphs 1 and 2 (Appendix), where there are huge dispersion peaks between the predicted and the observed returns.

Before moving forward to the out-of-sample forecasts, an in-sample with one additional lag on all variables was constructed to extend the robustness of the initial in-sample forecast. In this case, due to the added lags, only two coefficients were significant for both the behavioural and hybrid models, $Dispersion_{i,t-1}$ and $RP_{i,t-1}$, and $SMB_{i,t-1}$ and $RP_{i,t-1}$, respectively. Possibly because the lag does not change their data since it does not update on a weekly basis. The behavioural model, as in previous forecasts, tends to understate returns, while the hybrid model, possibly due to the addition of the FFTFM, tends to overstate returns. Consequently, the behavioural model has a higher accuracy in the downside of returns and the hybrid model in the upside of returns. Additionally, the dispersion measures are smaller than the previous forecasts, Table 13 (Appendix). Graphics 3 and 4 provide evidence that the forecasts may be more appropriate with one lag, as the overstating and understating of returns soothes down, possibly because the models are using less forecasting variables.



Graph 3 - Cumulative Returns of the in-sample lagged forecast and observed returns for the behavioural model



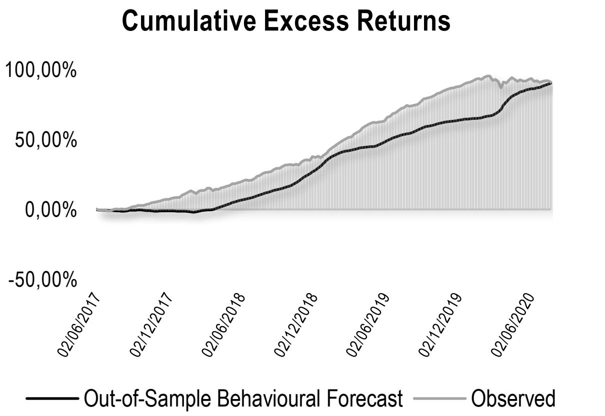
Graph 4 - Cumulative Returns of the in-sample lagged forecast and observed returns for the hybrid model

Out-of-Sample Forecasts Results

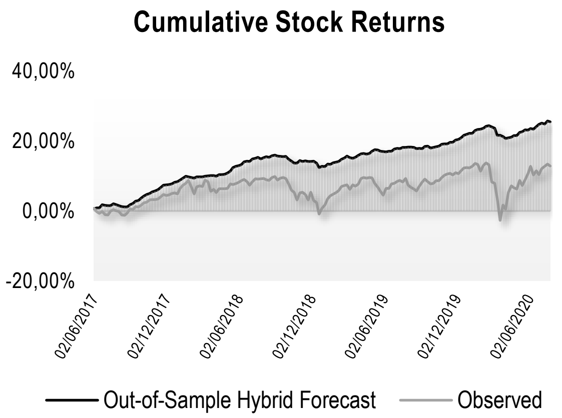
The ARCH/GARCH forecasting method is a benchmark for out-of-sample forecasts, thus it was the preferred forecasting tool. After ruling out seasonality, confirming that returns were stationary, and applying the proper lags to correct for the effects of autocorrelation, moving averages and GARCH effects, the ex-ante forecast was constructed. The in-sample data consisted of $2/3$ of the full sample and the remaining $1/3$ would correspond to the out-of-sample forecast, additionally, all the independent variables were lagged by one week. In essence, the forecast is only using available data at the time of the event, thus could be considered ex-ante forecasting. Both forecasts had good dispersion measures, Graph 5 and 6. However, there was a reversal in the behavioural model, since now it was not able to predict the downside returns, Table 14, possibly due to the absence of $Dispersion_{i,t}$ in this forecast. The significant coefficients for the behavioural and hybrid model were, $SV_{i,t}$ and $RP_{i,t}$, and $RSV_{i,t}$, $SV_{i,t}$ and $SMB_{i,t}$, respectively.

Indicators	Excess Return Model	Equally Weighted Return Model
Accuracy Bull	85,71%	82,47%
Accuracy Bear	18,75%	47,06%
MAE	0,24%	0,27%
RMSE	0,01%	0,08%
Standard Dev. Obs. Returns	1,01%	1,35%
Percentile 0%	-0,36%	-2,08%
Percentile 25%	0,18%	-0,06%
Percentile 50%	0,49%	0,15%
Percentile 75%	0,79%	0,44%
Percentile 100%	3,05%	0,95%

Table 14 - Out-of-Sample forecast results with one-week lag for the behavioural model (left) and the hybrid model(right)



Graph 5 - Cumulative Returns of the out-of-sample lagged forecast and observed returns for the behavioural model



Graph 6 - Cumulative Returns of the out-of-sample lagged forecast and observed returns for the hybrid model

5. Conclusion

Kahneman and Tversky's Prospect Theory unravelled a new perspective to investors' decision making. Since then, several authors have tried to prove that investors do not follow the traditional Bayesian rules for investment decisions. Rather, these decisions are skewed due to cognitive biases and heuristics. Based on a sample of the 2011 NASDAQ 100 stocks, this paper uses literature-wide indicators to proxy the arguably main heuristics in the financial markets, anchoring bias, availability bias, confirmation bias, overconfidence bias and representativeness bias. The first hypothesis is that heuristics are able to explain and predict excess returns and the second one is that heuristics can be combined with FFTFM to explain and predict stocks returns.

The daily model was incapable of explaining excess returns, presumably because of the statistical inability to provide useful insights when the normality principle is densely disrupted. However, although various coefficients were not statistically significant, the weekly model was able to capture vast variations of excess returns, and the variables intercorrelation provided evidence that validated their hypothesis in accordance with the literature. Heuristics were also successfully combined with FFTFM and were able to augment the traditional factors explanatory capabilities. Simultaneously, the in- and out-of-sample forecasts provide insights on the goodness-of-fit of the model's return predictions. The lagged in-sample forecast yielded better results than the in-sample forecast without the lags. The behavioural out-of-sample forecast has small accuracy in predicting the down movements of the market. Nonetheless, the hybrid model had small dispersion measures and high accuracy in both up and down movements, which indicates that it is highly capable of capturing the returns variations. Consequently, it might be a good forecasting model for future research.

In future research it would be interesting to assess if other models with slower time frequencies, such as monthly or quarterly data, arrive at similar indicators and significance levels for their corresponding coefficients in explaining excess returns. Additionally, it would be interesting to verify if the coefficients of Google and Revisions gain significance with this time frequency change, as previously discussed. Lastly, it would be interesting to study if other models can achieve valuable forecast results and apply them to different markets.

References

Abdin, S., Islam, M., Ahmad, M., and Hanif, H. 2020. "The Role of Heuristics Toward Stock Market Anomalies (Finding at Individual Investors)." *Abasyn University Journal of Social Sciences* 13 (1): 452-465.

Agarwal, S., Liu, C., and Rhee, C. H. 2008. "Investor demand for IPOs and aftermarket performance: Evidence from the Hong Kong stock market." *Journal of International Financial Markets, Institutions and Money* 18 (2): 176-190.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. 2006. "The Cross-Section of Volatility and Expected Returns." *The Journal of Finance* 61 (1): 259-299.

Asem, E. and Tian, G. 2010. "Market Dynamics and Momentum Profits." *The Journal of Financial and Quantitative Analysis* 45 (6): 1549-1562.

Banz, R. W. 1981. "The relationship between return and market value of common stocks." *Journal of Financial Economics* 9 (1): 3-18.

Barber, B. M. and Odean, T. 2000. "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *The Journal of Finance* 55 (2): 773-806.

Barber, B. M. and Odean, T. 2008. "All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies* 21 (2): 785-818.

Barber, B. M., Odean, T., and Zhu, N. 2009. "Do Retail Traders Move Markets?" *Review of Financial Studies* 22 (1): 151-186.

Basu, S. 1983. "The relationship between earnings yield, market value and return of NYSE common stock: Further evidence." *Journal of Financial Economics* 12 (1): 129-156.

Bates, D. S. 1991. "The Crash of '87: Was It Expected? The Evidence from Options Markets." *Journal of Finance* 46 (3): 1009-1044.

Belsky, G. and Gilovich, T. 1999. *Why Smart People Make Big Money Mistakes and How to Correct Them – Lessons From the New Science of Behavioral Economics.* Simon & Schuster, New York.

Bhootra, A. and Hur, J. 2013. "The timing of 52-week high price and momentum" *Journal of Banking & Finance* 37 (10): 3773-3782.

Black, F. 1972. "Capital market equilibrium with restricted borrowing." *The Journal of Business* 45 (3): 44-455.

Black, F. 1986. "Noise." *The Journal of Finance* 41 (3): 529-543.

Black, F. and Scholes, M. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81 (3): 637-654.

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

Bordalo, P., Gennaioli, N., La Porta, R., and Shleifer, A. 2017. "Diagnostic Expectations and Stock Returns." *The Journal of Finance* 74 (6): 2839-2874.

Boussaidi, R. 2013. "Overconfidence bias and overreaction to private information signals: the case of Tunisia" *Procedia Social and Behavioural Sciences* 81: 9-21.

Brase, G. L., Cosmides, L., and Tooby, J. 1998. "Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgement under uncertainty." *Journal of Experimental Psychology: General* 127 (1): 3-21.

Burghardt, W. 2010. "Retail Investor Sentiment and Behavior – an Empirical Analysis" Karlsruhe Institute of Technology.

Byun, S. J., Lim, S. S., and Yun, S. H. 2016. “Continuing Overreaction and Stock Return Predictability” *Journal of Financial and Quantitative Analysis* 51 (6): 2015-2046.

Cafferata, A. and Tramontana, F. 2019. “A financial market model with confirmation bias.” *Structural Change and Economic Dynamics* 51: 252-259.

Campbell, J. Y, Ramadorai, T., and Vuolteenaho, T. O. 2005. “Caught On Tape: Institutional Order Flow and Stock Returns.” NBER Working Paper No. 11439.

Campbell, J. Y., Sanford, J. G., and Jiang, W. 1993. “Trading volume and serial correlation in stock returns.” *Quarterly Journal of Economics* 108 (4): 905-939.

Chang, E., Lin, T., Luo, Y., and Ren, J. 2017. “Ex-Day Returns of Stock Distributions: An Anchoring Explanation.” *Management Science* 65 (3): 955-1453.

Charness, G. and Dave, C. 2017. “Confirmation bias with motivated beliefs” *Games and Economic Behaviour 104: 1-23.*

Colin, F. C. and George, L. 2004. “Behavioral Economics: Past, Present Future” In *Advances in Behavioural Economics*, edited by Colin F. Camerer, George Loewenstein, and Matthew Rabin, 3-51. Princeton: Princeton University Press.

Daniel, K., and Hirshleifer, D. 2015. “Overconfident Investors, Predictable Returns, and Excessive Trading.” *Journal of Economic Perspectives* 29 (4): 61-88.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. 2002. “Investor Psychology and Security Market Under- and Overreactions.” *The Journal of Finance* 53 (6): 1839-1885.

De Bondt, W. F. M. and Thaler, R. 1985. “Does the Stock Market Overreact?” *The Journal of Finance* 40 (3): 793-805.

De Bondt, W. F. M. and Thaler, R. 1987. “Further Evidence on Investor Overreaction and Stock Market Seasonality” *The Journal of Finance* 42 (3): 557-581.

Diether, K. B., Malloy C. J., and Scherbina, A. 2002. “Differences of opinion and the cross section of stock returns” *The Journal of Finance* 57 (5): 2113-2141.

Dimson, E. and Mussavian, M. 1999. “Three Centuries of Asset Pricing.” *Journal of Banking and Finance* 23 (12): 1745-1769.

Driessen, J., Lin, TC., and Van Hermet, O. 2010. “How the 52-week high and low affect beta and volatility” Paper presented in the 8th NTU International Conference on Economics, Finance and Accounting, Taiwan, IEFA, June 21-23.

Fama, E. F. 1970. “Efficient Capital Markets: A Review of Theory and Empirical Work.” *The Journal of Finance* 25 (2): 383-417.

Fama, E. F. and French, K. R. 1992. “The Cross-Section of Expected Stock Returns.” *The Journal of Finance* 47 (2): 427-465.

Fama, Eugene F. 1976. Foundations of finance: portfolio decisions and securities prices. New York: Basic Books.

Fazzini, A., and Lamont, O. A. 2008. “Dumb Money: Mutual fund flows and the cross-section of stock returns.” *Journal of Financial Economics* 88 (2): 299-322.

Ganzach, Y. 2001. “Judging Risk and Return of Financial Assets.” *Organizational Behavior and Human Decision Processes* 83 (2): 353-370.

George, T. J. and Hwang, C. 2004. “The 52-Week High and Momentum Investing.” *The Journal of Finance* 59 (5): 2145-2176.

Gigerenzer, G. 1991. “From tools to theories: A heuristic of discovery in cognitive psychology.” *Psychological Review* 98 (2): 254-267.

Gigerenzer, G. 1996. “On narrow norms and vague heuristics: A reply to Kahneman and Tversky.” *Psychological Review* 103 (3): 592-596.

Gigerenzer, G. and Murray, D. J. 1987. *Cognitive as Intuitive Statistics.* Hillsdale, NJ: Lawrence Erlbaum.

Goetzmann, W. N., and Massa, M. 2002. “Daily Momentum and Contrarian Behaviour of Index Fund Investors.” *The Journal of Financial and Quantitative Analysis* 37 (3): 375-389.

Graham, B. and Dodd, D. L. 2009. *Security Analysis,* 6th ed. New York: McGraw-Hill.

Grinblatt, M. and Keloharju, M. 2001. “What Makes Investors Trade?” *The Journal of Finance* 56 (2): 589-616.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., and Merill, L. 2009. “Feeling Validated Versus Being Correct: A Meta-Analysis of Selective Exposure to Information.” *Psychological Bulletin* 135 (4): 555-588.

Hirshleifer, D. 2001. “Investor Psychology and asset pricing” *The Journal of Finance* 56 (4): 1533-1597.

Hou, K., Peng, L., and Xiong, W. 2009. A Tale of Two Anomalies: The Implications of Investor Attention for Price and Earnings Momentum. (SSRN working paper).

Hsu, Y., and Shiu, C. Y. 2010. “The overconfidence of investors in the primary market” *Pacific-Basin Finance Journal* 18 (2): 217-239.

Hur, J. and Singh, V. 2019. “How do disposition effect and anchoring bias interact to impact momentum in stock returns?” *Journal of Empirical Finance* 53: 238-256.

Hur, J., Pritamani, M., and Sharma, V. 2010. “Momentum and Disposition Effect: The Role of Individual Investors.” *Financial Management* 39 (3): 1155-1176.

Hwang, C., Wong, K., and Yi, L. 2016. “Why do High Dispersion Stocks Earn Low Returns? Evidence from Institutional Ownership” Paper Presented in the 28th SFM Conference – Conference on the Theories and Practices of Securities and Financial Markets, Taiwan, Kaohsiung.

Johnson, T. C. 2005. “Forecast Dispersion and the Cross Section of Expected Returns” *The Journal of Finance* 59 (5): 1957-1978.

Kahn, R. N. 2018. *The Future of Investment Management.* CFA Institute Research Foundation Publications.

Kahneman, D. and Tversky, A. 1974. “Judgement under Uncertainty: Heuristics and Biases.” *Science* 185 (4157): 1124-1131.

Kahneman, D. and Tversky, A. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47 (2): 263-291.

Kaniel, R., Saar, G., and Titman, S. 2008. “Individual Investor Trading and Stock Returns.” *The Journal of Finance* 63 (1): 273-310.

Kelly, K., Low, B., Tan, H., and Tan, S. 2011. “Investors’ Reliance on Analysts’ Stock Recommendations and Mitigating Mechanisms for Potential Overreliance.” *Contemporary Accounting Research* 29 (3): 991-1012.

Khan, H. H., Naz, I., Qureshi, F., and Ghafoor, A. 2017. “Heuristics and stock buying decision: Evidence from Malaysian and Pakistani stock markets.” *Borsa Istanbul Review* 17 (2): 97-110.

Kristoufek, L. 2013. “Can Google Trends search queries contribute to risk diversification?” *Scientific Reports* 3 (1): 2713-2718.

Lee, C. MC, Shleifer, A. and Thaler, R. H. 1991. “Investor Sentiment and the Closed-End Fund Puzzle.” *Journal of Finance* 46 (1): 75-109.

Leroy, S. F. and Porter, R. D. 1981. “The present-value relation: Tests based on implied variance bounds.” *Econometrica* 49 (3): 555-574.

Li, J. and Yu, J. 2012. “Investor attention, psychological anchors, and stock return predictability.” *Journal of Financial Economics* 104 (2): 401-419.

Lintner, J. 1965. “The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets.” *Review of Economics and Statistics* 47 (1): 13-37.

Marewski, J. N., Gigerenzer, G. and Gaissmaier, W. 2010. “Good judgements do not require complex cognition.” *Cognitive Processing* 11 (2): 103-121.

Markowitz, H. 1952. “Portfolio Selection” *Journal of Finance* 7 (1): 77-91.

McDermott, R. 2001. “The Psychological Ideas of Amos Tversky and Their Relevance for Political Science.” *Journal of Theoretical Politics* 13 (1): 5-33.

Mehra, R. and Prescott, E. C. 1985. “The equity premium: A puzzle.” *Journal of Monetary Economics* 15 (2): 145-161.

Mossin, J. 1966. “Equilibrium in a Capital Asset Market.” *Econometrica* 34 (4): 768-783.

Nickerson, R. 1998. “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises.” *Review of General Psychology* 2 (2): 175-220.

Nisar, T. M. and Yeung, M. 2017. “Twitter as a tool for forecasting stock market movements: A short-window event study.” *The Journal of Finance and Data Science* 4 (2): 101-119.

Nisar, T. M. and Yeung, M. 2018. “Twitter as a tool for forecasting stock market movements: A short-window event study” *The Journal of Finance and Data Science* 4 (2): 101-119.

Odean, T. 1999. “Do Investors Trade Too Much?” *American Economic Review* 89 (5): 1279-1298.

Park, C. 2005. “Stock Return Predictability and the Dispersion in Earnings Forecasts” *The Journal of Business* 78 (6): 2351-2376.

Park, J., Konana, P., Gu, B., Kumar, A., and Raghunathan, R. 2013. “Information Valuation and Confirmation Bias in Virtual Communities: Evidence from Stock Message Boards.” *Information Systems Research* 24 (4): 1050-1067.

Peterson, D. R. and Smedema, A. R. 2011. “The return impact of realized and expected idiosyncratic volatility” *Journal of Banking of Finance* 35 (10): 2547:2558.

Pouget, S., Sauvagnat, J., and Villeneuve, S. 2017. “A Mind is a Terrible Thing to Change: Confirmation Bias in Financial Markets” *The Review of Financial Studies* 30 (6): 2066-2109.

Raharja, B. S., Suhaeli, D., and Mranani, M. 2017. “Research of the Stock Price Overreaction and Investor Overconfidence Issues.” *Business Management and Education* 15 (1): 127-139.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. 2015. “The Effects of Twitter Sentiment Stock Price Returns.” PLoS ONE 10 (9): e0138441.

Risteski, D. and Davcev, D. 2014. “Can We Use Daily Internet Search Query Data to improve Predictive Power of EGARCH Models for Financial Time Series Volatility?” Paper presented at the International conference on Computer Science and Information Systems, Dubai, UAE.

Rode, C., Cosmides, L., Hell, W., and Tooby, J. 1999. “When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory.” *Cognition* 72 (3): 269-304.

Roll, R. 1983. “Vas Ist Das?” *The Journal of Portfolio Management* 9 (2): 18-28.

Ross, S. A. 1976. “The arbitrage theory of capital asset pricing.” *Journal of Economic Theory* 13 (3): 341-360.

Samson, A. (Ed.). 2020. *The Behavioural Economics Guide 2020 (with and Introduction by Colin Camerer).*Behavioral Science Solution Ltd. https://www.behavioraleconomics.com/be-guide/the-behavioral-economics-guide-2020/

Schemling, M. 2007. "Institutional and individual sentiment: Smart money and noise trader risk?" *International Journal of Forecasting* 23 (1): 127-145.

Sharpe, W. 1964. "Capital asset prices: A theory of market equilibrium under conditions of risk." *Journal of Finance* 19 (3): 425-442.

Shefrin, H. 2010. "Behavioralizing Finance." *Foundations and Trends in Finance* 4 (1-2): 1-184.

Shiller, R. J. 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *The American Economic Review* 71 (3): 421-436.

Shiller, R. J. 2003. "From Efficient Markets Theory to Behavioral Finance." *Journal of Economic Perspectives* 17 (1): 83-104.

Sias, R. W., and Whidbee, D. A. 2010. "Insider Trades and Demand by Institutional and Individual Investors." *The Review of Financial Studies* 23 (4): 1544-1595.

Stein, J. C. 1989. "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behaviour." *The Quarterly Journal of Economics* 104 (4): 655-669.

Tseng, Q. 2019. "Reconstruct Google Trends Daily Data for Extended Period". *Towards Data Science,* October 2. https://towardsdatascience.com/reconstruct-google-trends-daily-data-for-extended-period-75b6ca1d3420

Tversky, A and Kahneman, D.1973. "Availability: A heuristic for judging frequency and probability." *Cognitive Psychology* 5 (2): 207-232.

Verma, R., Baklaci, H., and Soydemir, G. 2008. "The impact of rational and irrational sentiments of individual and institutional investors on DJIA and S&P500 index returns." *Applied Financial Economics* 18 (16): 1303-1317.

Vrouenraets, M. 2011. "How does retail trading activity relate to market performance?" Tilburg University.

Williams, J. B. 1938. *The Theory of Investment Value.* Cambridge, MA: Harvard University Press.

Yang, N., Chu, H., Ko, K., and Lee, S. 2018. "Continuing Overreaction and momentum in a market with price limits." *Pacific-Basin Finance Journal* 48: 56-71.

Appendix

BIASES	REPRESENTS	FORMULAS
Anchoring Bias:		
$GH_{i,t-1}$	52-Week High Ratio	$\frac{Current\ Price_{i,t}}{52 - Week\ High\ Price_{i,t}}$
Availability Bias:		
$Google_{i,t}$	Google Trends Data	$log\left(\frac{Google\ Trends\ Data_{i,t}}{Google\ Trends\ Data_{i,t-1}}\right)$
$Bull_{i,t}$	Twitter Bullish Sentiment Tweets	$\frac{Bullish\ Messages\ Scanned_{i,t}}{Total\ Messages\ Scanned_{i,t}}$
$BBR_{i,t}$	Bull and Bear Sentiment Tweet Ratio	$\frac{Bullish\ Messages\ Scanned\ _{i,t}}{Bearish\ Messages\ Scanned\ _{i,t}}$
$Mess_{i,t}$	Total Number of Tweets Analysed	$\log\left(\frac{Total\ Messages\ Scanned_{i,t}}{Total\ Messages\ Scanned_{i,t-1}}\right)$
Confirmation Bias:		
$Dispersion_{i,t}$	Analysts' Recommendations Dispersion	$Std.Dev.Analysts'\ Recommendations$
$Revisions_{i,t}$	Analysts' Recommendations Revisions	$\log\left(\frac{Target\ Price\ Mean_{i,t}}{Target\ Price\ Mean_{i,t-1}}\right)$
Overconfidence Bias:		
$Momentum_t$	Momentum Factor	$DWA\ Momentum\ ETF_t -\ QQQ\ ETF_t$
$SV_{i,t}$	1-day Lagged Signalled Volume	$\begin{cases} Vol_{i,t-1},\ \ if\ r_{i,t-1} > 0 \\ 0, if\ r_{i,t-1} =\ 0 \\ -\ Vol_{i,t-1}, if\ r_{i,t-1} < 0 \end{cases}$
$RSV_{i,t}$	Retail Investor Signalled Volume	$\begin{cases} Vol_{i,t}\ if\ BR_{i,t} > BR_{i,t-1} \land BR_{i,t} > ABR_{i,t} \\ -Vol_{i,t}\ if\ BR_{i,t} \leq BR_{i,t-1} \lor BR_{i,t} \leq ABR_{i,t} \end{cases}$
$Turnover_{i,t}$	Stocks Turnover	$log\left(\frac{Volume\ _{i,t} *\ Current\ Price_{i,t}}{Market\ Capitalization_{i,t}}\right) - MA_{i,t-1y}$
Representativeness Bias:		
$RP_{i,t}$	Representativeness Heuristic	$\frac{Closing\ Price_{i,t}}{Analysts'Recommendation_{i,t}}$
Idiosyncratic Volatility:		
$IVOL_{i,t}$	Idiosyncratic Volatility Measure	$IVOL_{i,t} =\ DIVol_{i,t-1}$

Retail Investors' Sentiment:		
$\boldsymbol{ABR_{i,t}}$	8-Week Average of Bullish Retailers from AAII	$8 - \text{week Average of \% Bullish Retail Trades}_t$
$\boldsymbol{BBS_{i,t}}$	Bullish – Bearish Retailers Sentiment from AAII	$\text{\% Bullish Retail Trades}_t$ $-\text{\% Bearish Retail Trades}_t$
Retail Investors' Presence:		
$\boldsymbol{Retail\ Proxy_{i,t}}$	Retail Investors' Presence Proxy	$\frac{Sum\ of\ Brokers'\ Total\ Trades_t}{Sum\ of\ Nasdaq\ 100\ Constituients'\ Total\ Volume_t}$

Table 1 – Overview of each Biases variables

Variables	Initial Regression	Autocorrelation Correction
$\boldsymbol{SMB_t}$	0.001*** (1.42E-04)	
$\boldsymbol{HML_t}$	0.001*** (1.26E-04)	
$\boldsymbol{Market\ Premium_t}$	0.676*** (0.02)	
$\boldsymbol{SMB_{t-1}}$		3.75E-04 (3.89E-04)
$\boldsymbol{HML_{t-1}}$		-4.87E-04 (3.22E-04)
$\boldsymbol{Market\ Premium_{t-1}}$		-0.096 (0.04)
R-Squared	69.81%	1.58%
Durbin-Watson Test	0.93	2.03
Mean VIF	1.02	1.02

Table 2 – Daily Robust FFTFM Linear Regression

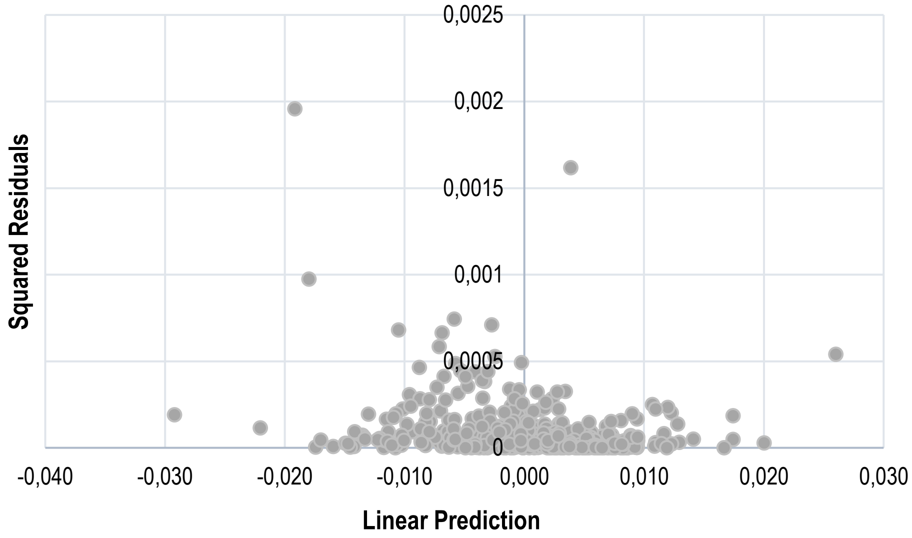


Figure 1 – Two-way Scatter Plot between Squared Residuals and Linear Prediction

	Momentum	SV	GH	Turnover	Dispersion	Revisions	RP	Bull	BBR	ABR	BBS
Momentum	1.0000										
SV	0.0285	1.0000									
GH	-0.0379	0.0397	1.0000								
Turnover	0.0102	-0.0197	-0.1827	1.0000							
Dispersion	0.0011	-0.0127	0.2068	-0.1778	1.0000						
Revisions	-0.0280	0.0522	0.2354	-0.1860	-0.0558	1.0000					
RP	-0.0268	0.1148	0.3623	-0.5283	0.2518	0.5687	1.0000				
Bull	0.0428	0.0212	-0.1375	0.1391	-0.0761	-0.0286	-0.1016	1.0000			
BBR	0.0311	0.0537	-0.0100	-0.1196	0.1800	0.0523	0.2626	0.3912	1.0000		
ABR	-0.0102	0.0101	0.1541	0.0521	-0.1884	0.2205	0.1533	0.0174	-0.2055	1.0000	
BBS	-0.0095	0.0316	0.2668	-0.1955	-0.1657	0.3115	0.4054	-0.0022	-0.0597	0.5633	1.0000
Mess	0.0557	0.0209	-0.0005	0.0897	-0.0006	-0.0054	0.0085	0.0758	0.1698	-0.0018	0.0001
Google	-0.0324	-0.0282	-0.0349	-0.0519	0.0205	-0.0004	0.0009	0.0171	0.0245	-0.0178	0.0030

	Mess	Google
Mess	1.0000	
Google	0.0255	1.0000

Table 3 – Daily Correlation Matrix

Variables	Results
SMB	0.002*** (4,37E-04)
HML	0.001 (4,02E-04)
Market Premium	0.277*** (0.03)
Constant	0.004*** (4,10E-04)
R-Squared	35.30%
Durbin-Watson Test	1.68
Mean VIF	1.02

Table 5 – Results from weekly excess return regression with FFTFM

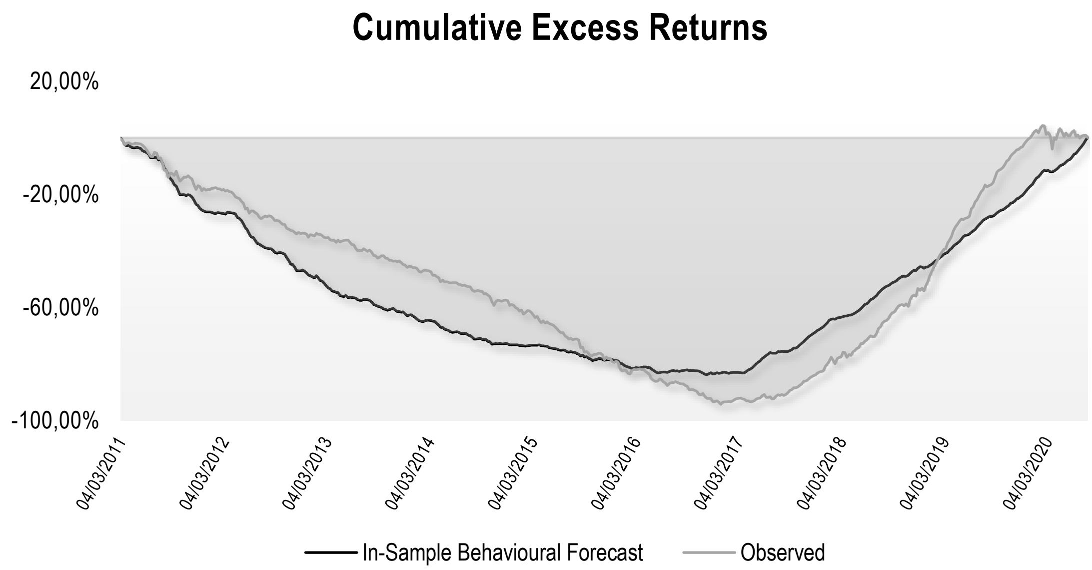
	Momentum	*RSV*	*SV*	*GH*	*Turnover*	*IVOL*	*Dispersion*	*Revisions*	*RP*	*Bull*	*BBR*
Momentum	1,0000										
RSV	-0,0413	1,0000									
SV	-0,0142	0,1155	1,0000								
GH	-0,0088	-0,0280	0,0175	1,0000							
Turnover	-0,0416	-0,1554	-0,0544	0,0429	1,0000						
IVOL	-0,0801	-0,0245	-0,0163	-0,7004	-0,0447	1,0000					
Dispersion	-0,0304	0,0717	-0,0657	-0,2325	-0,0004	0,1887	1,0000				
Revisions	-0,0302	-0,0621	0,0439	0,6870	0,1272	-0,5070	-0,1203	1,0000			
RP	-0,0358	0,1644	0,1638	0,6685	-0,0831	-0,5775	0,2390	0,5775	1,0000		
Bull	-0,0234	0,0273	0,1322	-0,0085	0,0256	-0,0789	-0,1660	-0,0106	-0,0213	1,0000	
BBR	-0,0310	0,0679	0,0126	0,1279	0,0566	-0,1915	0,2084	0,0382	0,2687	0,5867	1,0000
ABR	0,0035	-0,0952	0,0545	0,4110	0,0333	-0,1052	-0,1900	0,3007	0,1785	0,0065	-0,2156
BBS	-0,0351	0,4029	0,1083	0,4909	-0,1148	-0,3133	-0,1724	0,3178	0,4288	0,0322	-0,0530
Mess	-0,0713	-0,0213	0,0575	0,0003	0,3378	0,0201	-0,0094	0,0664	0,0412	0,0998	0,1177
Google	-0,0259	0,0567	-0,0195	0,0162	0,0428	-0,0431	0,0229	-0,0153	0,0153	0,0460	0,0404
Retail Proxy	-0,0139	0,0166	-0,0393	-0,3074	-0,0569	0,3000	0,7490	-0,0532	0,1451	-0,1814	0,0576

	ABR	*BBS*	*Mess*	*Google*	*Retail Proxy*
ABR	1,0000				
BBS	0,5739	1,0000			
Mess	-0,0265	0,0015	1,0000		
Google	0,0123	0,0294	0,0437	1,0000	
Retail Proxy	-0,2579	-0,2757	-0,0158	-0,0342	1,0000

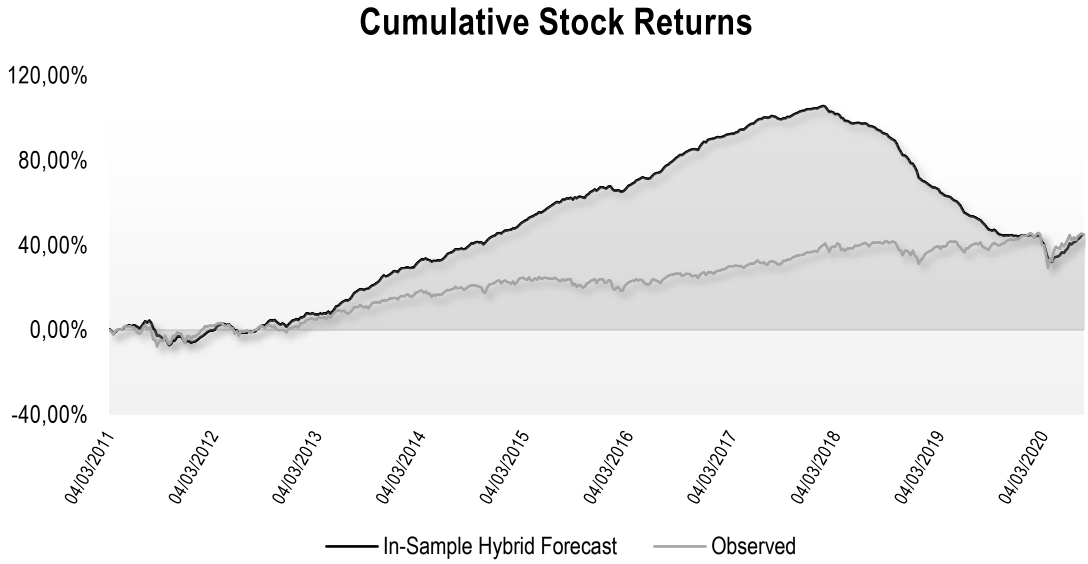
Table 6 – Correlation Matrix for all initial weekly heuristic indicators

	SMB	HML	Market Premium	RSV	SV	RP	Bull	Dispersion	ABR
SMB	1.0000								
HML	0.0515	1.0000							
Market Premium	0.1674	0.0318	1.0000						
RSV	0.1558	0.0520	0.0642	1.0000					
SV	0.0836	-0.1104	0.2141	0.1112	1.0000				
RP	0.1443	0.0594	0.0785	0.1640	0.1636	1.0000			
Bull	0.0157	-0.0386	0.0388	0.0266	0.1325	-0.0213	1.0000		
Dispersio n	-0.0206	-0.0984	-0.6485	0.0745	-0.0687	0.2386	-0.1663	1.0000	
ABR	-0.0372	-0.0560	0.0635	-0.0998	0.0596	0.1780	0.0073	-0.1932	1.0000

Table 10 – Weekly Hybrid Correlation Matrix



Graph 1 – Cumulative Returns of the in-sample forecast and observed returns for the behavioural model



Graph 2 - Cumulative Returns of the in-sample forecast and observed returns for the hybrid model

Indicators	Excess Return Model	Equally Weighted Return Model
Accuracy Bull	58,85%	64,93%
Accuracy Bear	82,59%	36,32%
MAE	0,57%	0,79%
RMSE	0,00%	0,00%
Standard Dev. Returns	0,89%	1,11%
Percentile 0%	-0,48%	-0,53%
Percentile 25%	-0,25%	-0,06%
Percentile 50%	-0,11%	0,09%
Percentile 75%	0,26%	0,22%
Percentile 100%	1,37%	1,27%

Table 13 - In-Sample forecast results with one-week lag for the behavioural model (left) and the hybrid model(right)