Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER THESIS

Extending volatility models with market sentiment indicators

Author: **MSc Lenka Röhryová**

Supervisor: **PhDr. Ladislav Krištoufek PhD**

Academic Year: **2017/2018**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, January 5, 2018

Signature

Acknowledgments

This master thesis would have not been possible, without the people to whom I owe a great amount of gratitude. First and foremost, I would like to thank my thesis supervisor PhDr. Ladislav Krištoufek PhD for his valuable advice and guidance. Secondly, I would like to express my gratitude to my family and friends for their continuous support throughout my studies.

Abstract

In this thesis, we aim to improve forecast accuracy of a heterogenous autoregressive model (HAR) by including market sentiment indicators based on Google search volume and Twitter sentiment. We have analysed 30 companies of the Dow Jones index for a period of 15 months. We have performed out-of-sample forecast and compiled a ranking of the extended models based on their relative performance. We have identified three relevant variables: daily negative tweets, daily Google search volume and weekly Google search volume. These variables improve forecast accuracy of the HAR model separately or in a Twitter-Google combination. Some specifications improve forecast accuracy by up to 22% for particular stocks, others impair forecast accuracy by up to 24%. The combination of daily negative tweets and weekly search volume is a superior model to the basic HAR for 17 stocks according to RMSE and for 16 stocks according to MAE and MASE. The daily negative tweets specification outperforms the basic HAR for 17 and 19 stocks, respectively. And, the combination of daily negative tweets and daily search volume outpaces the basic HAR for 15 and 18 stocks, respectively. Based on the average MASE improvement, the combination of daily negative tweets and weekly search volume is a clear winner as it lowers the average MASE by 0.71%.

JEL Classification C32, C33, C52, G14 G17
Keywords volatility, market sentiment, HAR model

Author's e-mail lenkarohryova@gmail.com
Supervisor's e-mail ladislav.kristoufek@fsv.cuni.cz

Abstrakt

V této práci se snažíme zlepšit heterogenní autoregresivní model (HAR) rozšířením o ukazatele tržního sentimentu. Jako proxy tržního sentimentu používáme objem vyhledávání na Googlu a Twitter sentiment. Analyzovali jsme 30 společností Dow Jones indexu po dobu 15 měsíců. Pomocí out-of-sample předpovědi jsme sestavili žebříček modelů podle jejich přesnosti. Identifikovali jsme tři relevantní proměnné: denní negativní tweety, denní objem vyhledávání a týdenní objem vyhledávání. Tyto proměnné zlepšují přesnost předpovědi HAR modelu jednotlivě i v Twitter-Google kombinacích. Některé modely zlepšují přesnost předpovědi až o 22% pro určité akcie, jiné zhoršují přesnost předpovědi až o 24%. Kombinace denních negativních tweetů a týdenního objemu vyhledávání překoná základní model u 17 akcií podle RMSE a u 16 akcií podle MAE a MASE. Samotné denní tweety zlepšují přesnost základního modelu pro 17 a 19 společností podle užitého měřítka. Kombinace denních negativních tweetů a denního objemu vyhledávání zlepšuje přesnost základního modelu pro 15 respektive 18 společností. Na základě průměrného zlepšení MASE vítězí jednoznačně kombinace denních negativních tweetů a týdenního objemu vyhledávání, jelikož snižuje průměrnou MASE o 0.71%.

Klasifikace JEL	C32, C33, C52, G14 G17
Klíčová slova	volatilita, tržní sentiment, HAR model

E-mail autora	`lenkarohryova@gmail.com`
E-mail vedoucího práce	`ladislav.kristoufek@fsv.cuni.cz`

Contents

List of Tables

List of Figures

Acronyms

ARCH Autoregressive conditional heteroskedastic

DJIA Dow Jones Industrial Average

DM Diebold-Mariano test

EMH Efficient market hypothesis

FMH Fractal market hypothesis

GARCH Generalized autoregressive conditional heteroskedastic

HAR Heterogeneous Auto Regressive

HMH Heterogeneous market hypothesis

MAE Mean-absolute error

MASE Mean-absolute-scaled error

RMSE Root-mean-square error

Master Thesis Proposal

Author	MSc Lenka Röhryová
Supervisor	PhDr. Ladislav Krištoufek PhD
Proposed topic	Extending volatility models with market sentiment indicators

Topic characteristics

Understanding the volatility of stock returns is crucial for forecasting stock price movements. Although volatility is natural part of liquid market, excessive volatility leads to instability and potential crashes. Many economists have studied the nature and sources of volatility. Beside financial time series modelling, a behavioural approach to volatility has become popular during past decades. Scholars have been concerned with phenomena such as emotionality of investors, cognitive biases, or market sentiment.

There is no doubt that market sentiment affects stock price volatility. Baker and Wurgler (2007) outlined several approaches to behavioural finance and stock market. They conclude that investors' sentiment is an important factor of a stock market and showed that it can be measured. Economists have begun to incorporate market sentiment into volatility models and revealed novel findings. For instance, Uygur and Tas (2012) examined the effect of noise traders during high-sentiment and low-sentiment periods using EGARCH and TGARCH models. They provided an evidence that a mean-variance relationship is undermined during high-sentiment periods when noise traders are more active. Whereas Uygur and Tas used trading volume as a proxy for market sentiment, Ranco et al. (2015) took advantage of Twitter and used tweets about the companies as the proxy. They investigated 30 companies of Dow Jones index and found a significant dependence between the Twitter sentiment and market returns.

This thesis aims to refine Heterogeneous Auto-Regressive (HAR) volatility model (Corsi, 2009) by market sentiment extension. I will use information about the set of stocks from Dow Jones Industrial Average (DJIA) index from the dataset, which use Ranco et al in their paper. For each of 30 companies, there are financial data containing open, high low and close price and the date and twitter data consisting of number of positive, neutral and negative tweets, number of total tweets and date. Furthermore, I will extract data of searching volume from Google Trends as another proxy of market sentiment. I will provide out-of-sample forecast performance comparison and compare performance of the model with other volatility models (e.g. GARCH).

Methodology

This thesis aims to refine the Heterogeneous Auto-Regressive (HAR) volatility model (Corsi, 2009) by market sentiment indicators. We are going to use the dataset compiled byRanco et al. (2015). The panel captures DJIA companies for a period of 15 months. We have a financial data containing open, high low and close price, and a twitter data consisting of a number of positive, neutral and negative tweets, and a number of total tweets. Furthermore, we are going to extract searching volume from Google Trends as another proxy of market sentiment. We are going to derive an out-of-sample forecast and compare forecast accuracy of various specifications.

Expected contribution This thesis intends to contribute to deeper understanding of the effect of market sentiment on volatility. Particularly, it strives to introduce an extension of the HAR model incorporating market sentiment indicators. Since this HAR specification has not been deeply studied, we hope that we could provide valuable novel insights.

Outline

1. **Introduction:** We will introduce volatility, market sentiment and our motivation.
2. **Literature review:** We will summarize existing literature on market sentiment, measuring volatility, the HAR model and its extensions.
3. **Data description:** We will describe data sources and transformations.
4. **Methodology:** We will outline used methodology and HAR specifications.

5. **Empirical findings:** We will present outcome of the models and compare their performance using out-of-sample forecasts and benchmarking.
6. **Conclusion:** We will summarize the most remarkable findings, if any, and suggest further research opportunities.

Core bibliography

1. Baker, M. and Wurgler, J. (2007), 'Investor sentiment in the stock market', *The Journal of Economic Perspectives* 21(32), 129-151.
2. Corsi, F. (2009), 'A simple approximate long memory model of realized volatility', *Journal of Financial Econometrics 7,*174-196.,
3. Ranco, G., Aleksovski, D., Caldarelli, G., M., G. and Mozetic, I. (2015), 'The effects of twitter sentiment on stock price returns',*PloS one* 10(9), e0138441.
4. Uygur, U. and Tas, O. (2012), 'Modeling the effects of investor sentiment and conditional volatility in international stock markets', *Journal of Applied Finance and Banking* 2(5), 239

Author

Supervisor

Chapter 1

Introduction

Volatility measures variation of a trading price over time. In other words, it refers to an amount of uncertainty about a size of changes in returns. Higher volatility means that a price of an underlying security may disperse over a larger range of values. Thus, it is an important indicator of risk and subject to many analyses.

Volatility modelling has evolved over last fifty years from the random walk model to more sophisticated models accommodating unconditional and conditional variances, a leverage effect, or heterogeneity of investment horizons (Weron and Weron, 2000). Apart from more refined treatment of the information embodied in a stock price, a behavioural approach to volatility has become popular during past decades. Scholars have been concerned with phenomena such as emotionality of investors, cognitive biases, or market sentiment. There is no doubt that market sentiment affects stock price volatility. Baker and Wurgler (2007) concluded that investor sentiment is an important factor of stock market behaviour and showed that it can be measured. More scholars have incorporated market sentiment into their models (Hong and Stein, 2003, Shefrin, 2005). For instance, Uygur and Tas (2012) examined effect of noise traders during high-sentiment and low-sentiment periods using EGARCH and TGARCH models. They provide an evidence that a mean-variance relationship is undermined during high-sentiment periods when noise traders are more active.

With the advent of the big data and new technologies, novel data streams have become available. Whilst, most data sources used in economy are typically available with a significant delay, at a high level of aggregation, and for predetermined variables only, internet data are publicly available in real-time (Wu and Brynjolfsson, 2015). Those novel data streams can be analysed to extract market sentiment. Thus, whereas aforementioned authors used market data, e.g. trading volume, as a proxy for market sentiment, social networks, Google search engine, or online news channels offer alternative market sentiment indicators. Many authors have taken advantage of those novel sources of publicly available data. For instance, Bollen et al. (2011), Sprenger et al. (2014), Zhang et al. (2011) employed a microblogging platform Twitter to investigate link between market sentiment and stock returns.

The purpose of this thesis is to improve understanding of how can the internet data enhance performance of volatility models. Or stated more generally, whether are the internet data a good proxy of market sentiment. We have analysed 30 companies of Dow Jones index for a period of 15 months between May 31^{st} 2013 and September 18^{th} 2014. As a basic model, we have chosen the Heterogenous Auto-Regressive model (HAR) (Corsi, 2009) and we have extended it by Twitter data and Google search volume. Since HAR-type models work with high-frequency data and we work with daily data, we need to choose an alternative procedure. Hence, we use Yang-Zhang volatility estimator (Yang and Zhang, 2000) to capture volatility of underlying stocks. Twitter data are adapted from the paper "The Effects of Twitter Sentiment on Stock Price Returns" by Ranco et al. (2015). They consist of a daily number of negative, positive and neutral tweets. Google search volume is represented by a search index retrieved from https://trends.google.com/. We have performed out-of-sample forecast and compared resulting measurement errors.

We attempt to define superior specification of the HAR model extended by market sentiment indicators. Also, we would like to determine which are the best performing variables. We assume that phenomenon commonly known from recent literature, e.g. an asymmetric reaction to arrival of bad and good news (Chen et al., 2003), will be apparent also in our research.

The thesis is structured in the following way. Chapter 2 summarizes literature on volatility modelling and describes used theoretical concepts. Particularly, it focuses on evolution of market hypotheses, stylized facts about volatility and volatility models. Chapter 3 describes behavioural aspects of financial markets, market sentiment and latest application of internet data in financial modelling. Chapter 4 outlines used methodology and the dataset. The results are presented and discussed in the chapter 5. Finally, chapter 6 provides conclusion, theoretical contribution, limitation and suggestions for the further research.

Chapter 2

Theory of volatility

This chapter discusses literature on volatility modelling and describes used theoretical concepts. Particularly, it focuses on evolution of market hypotheses, stylized facts about volatility and volatility models.

2.1 Market theories

Economists have long time been interested in changes in stock returns. The first consensus on the source of price variation had emerged in early 60's suggesting that price volatility could be well captured by a random walk and thus was unforecastable. The random walk theory of asset pricing was advanced by Samuelson in 1965 who showed that in an informationally efficient market price variation must be unpredictable. Although the random walk model had proved to be empirically valid, it is rather a statistical statement than a coherent theory of asset pricing. (Pesaran, 2005)

2.1.1 Efficient market hypothesis

The efficient market hypothesis (EMH), developed based on the random walk model, is a theory describing behaviour of stock markets. Fama (1970) originally introduced three forms of EMH. The weak form states that share prices reflect all relevant information and thus their future movements cannot be predicted from past prices. The semi-strong form requires asset prices to change to fully reflect all publicly available information. And finally, the strong form postulates that stock prices reflect all information even if some investors have monopolistic access to some information.

Since, under EMH, stock prices reflect all relevant information and are immediately adjusted to new information, stocks are traded at their fair value

(Jarrow, 1988). EMH works quite well when markets are stable but deteriorates in presence of turbulence and shocks. This is not a surprising result as EMH is an equilibrium model and thus cannot deal with transition periods (Weron and Weron, 2000). With increasing frequency and volume of trade data, several other issues regarding clustering, slow-decaying autocorrelation and non-linear response occur (Sewell, 2011). EMH failed to fit the observed data and thus, there was a need to seek for a market hypothesis that would better describe heterogeneous components present in the financial markets. The fractal market hypothesis (FMH) introduced by Peters (1991, 1994) and the heterogeneous market hypothesis (HMH) proposed by Muller et al. (1993) provide an alternative explanation of market behaviour based on a chaos theory and fractal objects. Those hypotheses were proved to be suitable explanation of investors' heterogeneity with respect to their investment horizons. They provide a new theoretical background for more accurate modelling of non-linear market reaction, discontinuity and heterogeneity present in today's financial markets.

2.1.2 Fractal market hypothesis

The EMH's main weakness is a generic approach to market participants and information. It considers investors to be homogenous with respect to their behaviour, expectation, valuation technique and access to information. Moreover, it assumes investors to be rational price-takers who maximize return using all available information. Also, information is treated as a generic item drawn from a pool. Homogeneity of investors and information implies that a particular information influences all investors equally. (Weron and Weron, 2000)

Such conditions certainly do not reflect today's real markets. Thus, the fractal approach enables to embrace heterogeneity in data by analysing the objects on different scales, with different degrees of resolution, and comparing the results. (Muller et al., 1993). Peters (1994) proposed several assumptions considering investors' various investment horizons.

(i) *The market is made up of many individuals with a large number of different investment horizons.*

(ii) *Information has a different impact on different investment horizons.*

(iii) *The stability of the market is largely a matter of liquidity (balancing of supply and demand). Liquidity is available when the market is composed* *of many investors with many different investment horizons.*

(iv) *Prices reflect a combination of short-term technical trading and long-term fundamental valuation.*

(v) *If a security has no tie to the economic cycle, then there will be no long-term trend. Trading, liquidity, and short-term information will dominate.*

These five assumptions enable more precise approximation of market prices. Unlike the EMH, the fractal market hypothesis (FMH) captures diverse treatment of information in accordance with heterogeneity of investment horizons. Thus, as long as the market has a fractal structure, FMH proposes a stable model of market behaviour even during the turbulence periods. (Peters, 1994)

2.1.3 Heterogeneous market hypothesis

The heterogeneous market hypothesis (HMH) proposed by Muller et al. (1997) builds on the fractal theory and broadens heterogeneity of market participants by aspects of risk tolerance, information, institutional constraints, transaction costs etc. Muller et al. (1997) have described the HMH by following assumptions summarizing empirical findings.

(i) *Different actors in the heterogeneous market have different time horizons and dealing frequencies. The different dealing frequencies clearly mean different reactions to the same news in the same market. The market is heterogeneous, with a fractal structure of the participants' time horizons as it consists of short-term, medium-term and long-term components.*

(ii) *In a homogeneous market, the more agents are present, the faster the price should converge on the real market value, on which all agents with a rational expectation agree. In a heterogeneous market, different actors are likely to settle for different prices and decide to execute their transactions in different market situations. In other words, they create volatility.*

(iii) *The market is also heterogeneous in the geographic location of the participants.*

These additional assumptions describe better heterogeneity in markets and lead to a framework that realized volatility is an aggregation of the heterogeneous components. Moreover, Muller et al. (1997) observed that long-term volatility strongly influences short-term volatility but not vice versa. This is quite reasonable inference of investors' behaviour. Long-term volatility matters for short-term traders because it forms expectations of future risks and trends.

The statistical pattern created by traders of various frequencies can be statistically described as a *cascade of heterogeneous volatility components.*

*The geographical heterogeneity explains issues as the heat wave effect.*This meteorological analogy refers to persistence of market behaviour. In other words, it is likely that a hot day in New York will be followed by another one but it is not likely that it will be followed by another hot day in Tokyo. The markets exhibit similar behaviour. Thus, a high-volatility day will be likely followed by another high-volatility day in the same market.(Engle et al., 1990)

Another meteorological metaphor for volatility spillovers is the *meteor shower hypothesis. The volatility in one market transmits to another market and causes increased volatility in a geographically distant market opening several hours after closing of the original market. A combination of the heat wave effect* and the *meteor shower hypothesis* provides simplified description of volatility dynamics. (Engle et al., 1990) The impact of a shock also on subsequent returns, long memory and many other features have been formulated into styliyed facts.

2.1.4 Stylized facts about volatility

Over the years, scholars have formulated several stylized facts about stock market volatility. Cont (2001) in one of the most widely-cited articles in the financial literature "Empirical properties of asset returns: stylized facts and statistical issues" highlights some features of stock market returns relating to volatility.

1. **Absence of autocorrelations:** (linear) autocorrelations of asset returns are often insignificant, except for very small intraday time scales ($\simeq$ 20 minutes) for which microstructure effects come into play.
2. **Heavy tails:** the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular this excludes stable laws with infinite variance and the normal distribution.
3. **Gain/loss asymmetry:** one observes large drawdowns in stock prices and stock index values but not equally large upward movements.
4. **Aggregational Gaussianity:** as one increases the time scale t over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales.

5. **Intermittency:** returns display, at any time scale, a high degree of variability. This is quantified by the presence of irregular bursts in time series of a wide variety of volatility estimators.
6. **Volatility clustering:** different measures of volatility display a positive autocorrelation over several days, which quantifies the fact that high-volatility events tend to cluster in time.
7. **Conditional heavy tails:** even after correcting returns for volatility clustering (e.g. via GARCH-type models), the residual time series still exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution of returns.
8. **Slow decay of autocorrelation in absolute returns:** the autocorrelation function of absolute returns decays slowly as a function of the time lag, roughly as a power law with an exponent $\beta \in [0.2, 0.4]$. This is sometimes interpreted as a sign of long-range dependence.
9. **Leverage effect:** most measures of volatility of an asset are negatively correlated with the returns of that asset.
10. **Volume/volatility correlation:** trading volume is correlated with all measures of volatility.
11. **Asymmetry in time scales:** coarse-grained measures of volatility predict fine-scale volatility better than the other way round.

Economists have been trying to explain those stylized facts in different ways. For instance, long memory might be associated with many cognitive biases such as conservatism. Or, the leverage effect might relate to asymmetric reaction to positive and negative news. More about cognitive biases can be found in the chapter 3.

2.2 Volatility modelling

As the market theories have evolved, the underlying mathematical models follow the same path. Besides a random walk and linear models such as ARIMA, a new family of volatility models have arisen. During the eighties, the issue of autoregressive heteroscedasticity became widely discussed and led to discovery of non-linear models such as ARCH, GARCH and their modifications (Nelson, 1991). Recently, we have seen a surge of interest in models capturing volatility cascading, such as HAR models (Corsi, 2009). In this section, we summarize basic equations and properties of these models.

2.2.1 Generalized Autoregressive Conditional Heteroskedastic (GARCH) models

Conventional linear econometric models work with unrealistic assumption of a constant one-period forecast variance. A class of stochastic processes caused a breakthrough in volatility modelling. Engle (1982) introduced the ARCH (Autoregressive Conditional Heteroskedastic) model which is able to recognize unconditional and conditional variances. The ARCH processes are serially uncorrelated, zero mean processes. They allow different treatment of each type of variance. Whereas unconditional variance is constant over time, conditional variance is non-constant as it is a function of past errors. Using conditional densities, Engle specified ARCH as

$$\epsilon_t \mid \psi_{t-1} \sim N(0, h_t)$$
$$y_t = \epsilon_t \sqrt{h_t}$$
$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-1}^2$$

where p is the order of the ARCH process and α is a vector of unknown parameters and $V(\epsilon_t) = 1$.

ARCH processes have become important for modelling behaviour of many economic variables. However, linear declining lag structure in the conditional variance equation stemming from long memory property typical for many phenomena occurred in many applications of ARCH processes (Goudarzi, 2011). In response, Bollerslev (1986) introduced a new more general class of stochastic processes called Generalized Autoregressive Conditional Heteroskedastic (GARCH). GARCH-type models allow for a more flexible solution of a lag structure, long memory and more parsimonious explanation of the selected phenomena. The main advantage of GARCH processes is that they enable lagged conditional variances to specify forecasted conditional variance as well, whereas ARCH allows only past sample variances to enter the estimation process. Bollerslev defined the GARCH process as follows.

"Let ϵ_t denote a real-valued discrete-time stochastic process, and ψ_t the information set (σ-field) of all information through time t. The GARCH (p,q) (Generalized Autoregressive Conditional Heteroskedastic) process is then given by (Bollerslev, 1986):"

$$\epsilon_t = y_t - x_t^T \beta$$
$$\epsilon_t \mid \psi_{t-1} \sim N(0, h_t)$$
$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2 + \sum_{i=1}^{p} \beta_i h_{t-1}$$

$$= \alpha_0 + A(L)\epsilon_t^2 + B(L)h_t$$

$$\begin{array}{lll} p \geq 0, & q > 0 & \\ \alpha_0 > 0, & \alpha_i \geq 0 & i = 1, ..., q, \\ \beta_i \geq 0 & i = 1, ..., p. & \end{array}$$

For $p = q = 0$, ϵ_t is white noise, for $p = 0$ the process is described as ARCH(q).

In the early 1990s, economists revealed features associated with time dependence of conditional volatility that cannot be properly captured by GARCH models. Discovery of new volatility properties have called for modifications of the GARCH model. For instance, Baillie and Bollerslev (1990) found seasonal patterns in intra-day exchange spot rates, which can be well captured by seasonal GARCH with hours dummy. These patterns are also in line with previously discussed hypotheses, such as *meteor shower hypothesis* (Engle et al., 1990). GARCH models have been extended by jumps (Chen and Shen, 2004), leverage (Rodriguez and Ruiz, 2012) and many other effects.

2.2.2 Heterogeneous Auto Regressive (HAR) model

The idea that realized volatility is aggregation of a cascade of heterogeneous components leads to a Heterogeneous Auto-Regressive model proposed by Corsi (2009). Despite of its simplicity, it is able to accurately fit the observed market data and describe persistence of volatility.

A simple HAR model assumes that a variable X (e.g. a stock log-price) is driven by the stochastic process:

$$dX_t = \mu_t dt + \sigma_t dW_t + c_t dN_t \tag{2.1}$$

where μ_t is predictable, σ_t is cádlág and N_t is a stochastic Poisson process whose intensity is an adapted stochastic process λ_t, the timing of corresponding jumps is $(\tau_j)_{j=1,...,N_t}$ and c_j are i.i.d adapted random variables measuring the size of the jump at each time τ_j. Resulting quadratic variation is defined as:

$$\tilde{\sigma_t} = \int_t^{t+1} \sigma_s^2 ds + \sum_{t \leq \tau_j \leq t+1} c_{\tau_j}^2 \tag{2.2}$$

where the time unit is one day. For estimation of quadratic volatility, we use n observations in the interval [0, T]. The estimator of realised volatility is:

$$RV_t = \sum_{j=0}^{n-1} (\Delta_{t,j} X)^2 \tag{2.3}$$

given

$$\Delta_{t,j} X = X_{t+j/n} - X_{t+(j+1)/n} \tag{2.4}$$

which is consistent estimator, as $n \to \infty$, of $\tilde{\sigma}_t$.

Let's denote $\hat{V}_t$ a generic unbiased estimator of quadratic variation such that:

$$log\tilde{\sigma}_t = log\hat{V}_t + \omega_t \tag{2.5}$$

where ω_t is a zero mean and finite variance error term. We use a log form in order to avoid negativity issues and get approximately normal distribution.

To incorporate heterogeneity of components, we consider the aggregated values of $log\hat{V}_t$ as:

$$log\hat{V}_t^{(n)} = \frac{1}{n}(log\hat{V}_t + ... + log\hat{V}_{t-n+1}) \tag{2.6}$$

Let's assume two different time scales, of length n_1 and n_2, with $n_1 > n_2$. For the largest time scale, assume that $\tilde{\sigma}_t$ is determined by:

$$log\tilde{\sigma}_{t+n_1}^{(n_1)} = c^{n_1} + \beta^{n_1} log\hat{V}_t^{(n_1)} + \epsilon_{t+n_1}^{n_1} \tag{2.7}$$

where $\epsilon_t^{n_1}$ is IID zero mean and finite variance noise independent on ω_t. To capture influence of long-term volatility on short-term volatility, a volatility cascade from low to high frequencies is constructed so that shorter time scale volatility n_2 is influenced by the expected value of the largest time scale volatility n_1 (but not vice versa).

$$log\tilde{\sigma}_{t+n_2}^{(n_2)} = c^{n_2} + \beta^{n_2} log\hat{V}_{t+n_2}^{(n_2)} + \delta^{n_2} E_t[log\tilde{\sigma}_{t+n_1}^{(n_1)}] + \epsilon_{t+n_2}^{n_2} \tag{2.8}$$

where $\epsilon_t^{n_2}$ is IID zero mean and finite variance noise independent on ω_t and $\epsilon_t^{n_1}$.

By substitution, and using the equation 2.5 we obtain:

$$log\hat{V}_{t+n_2}^{(n_2)} = c + \beta^{n_2} log\hat{V}_t^{(n_2)} + \beta^{n_1} log\hat{V}_t^{(n_1)} + \epsilon_t \tag{2.9}$$

where ϵ_t is IID noise dependent on ω_t and $\epsilon_t^{n_1}$ and $\epsilon_t^{n_2}$. This model can be extended to d horizons such that: $n_1 > n_2 > .... > n_d$. Usually, there are three time horizons (monthly, weekly, daily) with length $n_1 = 22$ days, $n_2 = 5$ days and $n_3 = 1$ day. Since the shorter time-scale volatility is affected by the longer time-scale components, the auto-correlation function of the model (its memory) persistence increases. Thus, the HAR model is able to capture long memory as well as models belonging to the family of long memory processes. Similarly to GARCH-type models, HAR model has been extended by heterogenous jumps or leverage Corsi and Reno (2009).

Chapter 3

Market sentiment and financial modelling

This chapter briefly describes a behavioural approach to financial markets and market sentiment. Then, it highlights the current literature on financial modelling employing internet data.

3.1 Behavioural sources of mispricing

Whereas the efficient markets hypothesis has defined investors as rational, utility-maximizing individuals, cognitive psychology suggests that human decision processes are prone to several illusions. Behavioural finance proponents argue that biases caused by heuristic or arising from the adoption of "mental frames" cause market prices to deviate from their fundamental values (Singh, 2012).

The efficient market hypothesis fully neglects irrationality of investors and other behavioural aspects of pricing. On the other hand, the EMH does not require every single investor to act in a rational manner as long as the economic dominance of rational investors ensures the fair market prices (Singh, 2012). The FMH and the HMH incorporate investor-specific response to information and thus allow for behavioural elements. This is one of the most compelling reasons why incorporating market sentiment into the HAR model is interesting. It is possible that behavioural aspects mirrored in investors sentiment are already contained in the original HAR model. Since this thesis strives for deeper understanding of predicative value of market sentiment, we need to dive deeper into the behavioural features of volatility.

3.1.1 Cognitive and emotional biases

Behavioural finance employs principles that are less narrow than those based on Von Neumann-Morgenstern expected utility theory (Neumann and Morgenstern, 1953). The leading paradigm of behavioural approach is the Prospect theory. Kahneman and Tversky (1979) found that people assign value to gain and losses rather than to final assets and moreover, that people evaluate the outcomes with respect to a current level of status-quo. Resulting value function is concave for gains (due to risk aversion), convex for losses (risk seeking) and steeper for losses than for gains (as a result of risk aversion).

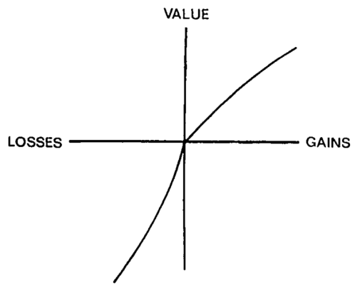


Figure 3.1: Hypothetical value function, Source: Kahneman and Tversky (1979)

The decision process itself comprises of two phases: the editing and the evaluation phase. During the editing, an individual organizes and reformulates outcomes according to certain rules in order to obtain simple preferential ordering. In many cases, an individual use heuristics and various cognitive easing to facilitate the process. In the second phase, the prospects are evaluated by the value function and the one with the highest utility is chosen. Some of the cognitive shortcuts are discussed below.

Heuristics Using the rule of thumb is the simplest approximation, which people use in order ease their choice. For example, when it comes to division of funds among n different investment possibilities, many people follow the *1/n* rule. Benartzi and Thaler (2001) have described many cases when people use the naive diversification heuristics when investing their retirement savings. They have conducted an experiment with two groups of employees, one was presented with four equity funds and one fixed-income securities fund and the second group was presented with one equity fund and four fixed-income securities funds. The first group allocated 68 percent of its funds to equity, whereas the second group ended up with only 43 percent in equities. Considerable influence of the 1/n rule (20% per option in this case) frequently leads to a suboptimal solution.

Overconfidence An overconfident investor overestimates her knowledge or abilities. As a result, she holds less diversified portfolio and sticks to what she is familiar with. Investors overestimate their ability to predict evolution of domestic markets and thus incline to invest more into local companies despite the potential gains from international diversification (Tesar and Werner, 1995). Another example of pernicious influence of overconfidence is that an overconfident CEO overestimates his ability to generate return and therefore undertakes value-destroying mergers and overpays for a target company (Malmendier and Tate, 2008).

Mental accounting Investors tend to categorize their current and future wealth into different, non-transferable groups. According to Thaler (1999), people divide financial assets into the separate mental accounts (current income versus future income) or budgets (housing, food, etc.). Each account operates within different propensity to consume and is re-evaluated at different frequency (daily, monthly, etc.). Mental accounting violates the economic principle of fungibility as it separates decisions that should be done together and thus distorts efficient decision-making process.

Representativeness Investors frequently label companies as good or bad based on their recent performance and fall in a trap of extrapolation bias. This leads to overpriced stocks of companies showing recent growth regardless of sustainability of the growth and vice versa. Ignoring a long-term average is also known as a law of small numbers as sufficient number of observations creates an image of what is "normal". For instance, high equity returns between 1982 and 2000 have led to a believe that high equtiy returns are normal. (Ritter, 2003)

Conservatism Conservatism bias is a counterpart to the representativeness bias. Whilst investors tend to overreact in a case of a sufficiently long pattern, they underreact to a sudden change. This phenomenon arises from anchoring on the expectations. Steenbergen (2001) showed that an information inconsistent with expectations is likely to be assigned a smaller weight than the expectations themselves and thus, even moderate amounts of new information contradicting the decision may not be enough to change it.

Anchoring Anchoring is the tendency to hold on to a one piece of information that is not adjusted in future. Individuals often base their decisions on the first information to which they are exposed (as initial purchase price of the stock). Many investors still anchor on the financial crisis 2008 what results in overall higher level of risk aversion. (Baker and Ricciardi, 2014)

Availability bias Information that is easily accessible, i.e. that is easily recalled from memory, overweighs others. Investors are likely to remember events that received a lot of attention by the media and base their evaluation on them. (De Bondt et al., 2008)

Herd behaviour Under certain circumstances, investors mimic the investment decisions of other investors and ignore substantive private information. Herd behaviour could partially explain excessive market volatility. By mimicking the behaviour of others rather than reacting to their private information, herd members amplify exogenous stock price shocks. Scharfstein and Stein (1990)

These are just few examples of cognitive and emotional biases leading to mispricing and inadequate volatility. The behavioural biases result, among others, in asymmetric reaction to price movements: returns and conditional volatility are negatively correlated (Bekaert and Wu, 2000). Morover, Veronesi (1999) found that in equilibrium, investors endeavour to hedge against changes in their own *uncertainty* drives stock prices overreact to bad news in good times and underreact to good news in bad times. Asymmetric reaction has given rise to a new class of models accommodating asymmetric volatility such as TGARCH (Zakoian, 1994, Glosten et al., 1993), EGARCH (Nelson, 1991), or extended version of the HAR model (Corsi and Reno, 2009). Those models are widely used. For instance, Chen and Shen (2004) found strong evidence supporting the asymmetrical hypothesis of stock returns by employing a double-threshold GARCH model. In other words, negative news causes a larger decline in a national stock return than an equal magnitude of good news.

3.1.2 Market sentiment

Behavioural finance advocates the effect of emotions on individual decision-making. However, does this also apply to societies at an aggregate level? The overall mood or tone of investors is captured by market sentiment. Market sentiment is the general prevailing attitude of investors towards anticipated development in a market (Baker and Wurgler, 2007). The question is no longer whether market sentiment influences stock prices but rather how to measure the effect.

One approach to quantify the effect of market sentiment is "bottom-up". Market sentiment is formed by investors' cognitive and emotional biases (see the subsection 3.1.1) and a variety of fundamental and technical factors. Considering irrationality and emotionality of investors, the standard finance models must have been augmented with alternative models. Related models explain mispricing by differences of opinion across investors combined with short sales constraints. They aspire to predict patterns in investor sentiment. (Hong and Stein, 2003, Shefrin, 2005)

However, real investors are too complicated to be neatly summarized by a few selected biases and trading frictions and thus, many of the bottom-up models result in a similar reduced form of variation over time in collective psychology (Baker and Wurgler, 2007). Thus, Baker and Wurgler (2007) define a "top-down" market sentiment approach operating on two assumptions. Firstly, investors are subject to sentiment that can be defined as a belief about future cash flows and investment risks (DeLong et al., 1990). Secondly, there are limits to arbitrage as betting against sentimental investors is costly and risky (Shleifer and Vishny, 1997). The top-down approach measures reduced-form, i.e. aggregate sentiment, and tracks its effects to returns of individual stocks. Therefore, this approach attempts to explain which stocks are likely to be most affected by sentiment, rather than simply arguing that the level of stock prices in the aggregate depends on market sentiment. Baker and Wurgler (2007) suggest that stocks of low capitalization, younger, unprofitable, high volatility, non-dividend paying, growth companies, and stocks of firms in financial distress are more likely to be disproportionately sensitive to broad waves of market sentiment.

We use the "top-down" market sentiment approach in our analysis. Aggregate market sentiment is a one-dimensional variable affecting all stocks to some extend but it affects some more than others. We analysed Twitter sentiment separately for each stock and thus, we can anticipate that some of those stocks will be more correlated with tweets based sentiment than the others.

3.2 Internet data and financial modelling

Social networks and new media have attracted a great deal of attention in the past. One of the greatest strength of the internet data is their availability for real-time predictions. Whilst, most data sources used in economy are typically available with a significant delay, at a high level of aggregation, and for predetermined variables only, internet data are publicly available in real-time (Wu and Brynjolfsson, 2015). Some researchers suggest that those novel data streams can be analysed to extract public sentiments to improve prediction of the market indicators (Lavrenko et al., 2000, Schumaker and Chen, 2009).

Some scholars tried to increase forecast accuracy of financial models by incorporating market sentiment. For instance, Bollen et al. (2011) investigated whether Twitter based the public mood is predictive of economic indicators. They approximated collective mood (e.g. Calm, Alert, Sure, Vital, Kind, and Happy) by large-scale Twitter feeds and analysed its correlation with the value of the Dow Jones Industrial Average (DJIA) over time. Their results indicate that the accuracy of DJIA predictions can be improved by the inclusion of specific public mood dimensions (particularly "Calm"). They obtained an accuracy of 86.7% in predicting the daily up and down changes in the closing values of the DJIA and a reduction of the Mean Average Percentage Error (MAPE) by more than 6%.

Si et al. (2013) improved performance of a VAR model for short term (one day ahead) predictions of the S&P100 Index by including topic-based sentiment form Twitter. Their topic-based model shows better performance than existing state-of-the-art non-topic based methods.

Zhang et al. (2011) also aimed to predict stock market indicators by analysing Twitter posts. They got a randomized subsample of about one hundredth from a pool of the twitter feeds for a period of six months. They measured daily collective hope and fear and analysed the correlation with the stock market indicators. They discovered negative correlation between emotional indices and Dow Jones, NASDAQ and S&P 500, but significant positive correlation to VIX.

Recently, Ranco et al. (2015) examine the effect of twitter sentiment on stock price returns. In their paper, they present evidence of dependence between stock price returns and Twitter sentiment in tweets about the companies. The drawback is that dependence is significant only at the moments of increased activity of Twitter users. Their results show that aggregated Twitter sentiment predicts the direction of market response during the pre-selected events. This can be expected for both "known" events, like earnings announcements, and unexpected news.

Sprenger et al. (2014) arrived at similar results. There is an association between tweet sentiment and stock returns. Buy signals are accompanied and followed by abnormal returns but sell signals have no predictive power. Moreover, message volume corresponds to trading volume but not returns and volatility. In addition, they provide an analysis of information diffusion and revealed that users providing above-average investment advice are given credit via higher retweets and followers.

Twitter based sentiment is not the only one used for financial modelling. Antweiler and Frank (2004) used text classification to study Yahoo!Finance and Raging Bull message boards for the 45 companies of the Dow Jones Industrial Average and Dow Jones Internet Index. They point out that message volume implies trading volume and volatility. However, due to the presence of the internet bubble in the sample, the study has severe limitations.

Wu and Brynjolfsson (2015) used data from Google search engine to predict future housing market sales and prices. They showed strong correlation between the housing search index and house sales in the next quarter. They provided out-of-sample predictions, too. The extended model shows a mean absolute error of just 0.102 that is substantially lower than a mean absolute error of the baseline model (0.441).

Smith (2012) studied a link between Google Internet searches for particular keywords and volatility prediction in the market for foreign currency. He found that the keywords economic crisis, financial crisis and recession have incremental predictive power beyond the GARCH(1,1).

These papers have already shown that there is a great potential in investigating internet data as a storehouse of information for financial markets. Social network analysis might be used as a proxy for market sentiment. Google Internet search activity might be also a convenient method of detecting market turbulence and forecasting.

Chapter 4

Methodology and Data

This chapter outlines used methodology and the dataset. Firstly, it presents the dataset, including adjustment of the Google search volume. Then, it describes the volatility estimators used for simplified HAR model and our specification of the HAR model. Finally, it describes forecast accuracy measures used for evaluation of out-of-sample forecasts.

4.1 Data

In this section, we present the data and their transformation. Firstly, we describe market data, the Twitter data and the Google search volume. Secondly, we define winsorization: the transformation applied to the data to limit potentially spurious outliers.

4.1.1 Market data

The analysis is conducted on 30 companies of the Dow Jones index for a period of 15 months between May 31^{st} 2013 and September 18^{th} 2014. The volatility estimators (see section 4.2) are computed from opening, closing, high and low prices. This data are publicly available and can be downloaded from various sources, such as https://finance.yahoo.com/. The ticker list of the investigated stocks and the amounts of corresponding tweets are presented in the Table 4.1.

4.1.2 Search volume

We employ Google internet search volume as a proxy for interest in the specific companies. We use daily Google searches for each company that are publicly available at https://trends.google.com/. Google enables to adjust a searched

Ticker	Company	Total Tweets	Negative Tweets	Positive Tweets
AXP	American Express Co	21,941	1,340	3,369
BA	Boeing Co	51,799	3,780	9,461
CAT	Caterpillar Inc	38,739	4,840	6,698
CSCO	Cisco Systems Inc	57,427	5,380	10,235
CVX	Chevron Corp	29,477	2,547	4,958
DD	E I du Pont de Nemours and Co	17,340	1,076	2,594
DIS	Walt Disney Co	46,439	1,652	8,854
GE	General Electric Co	61,836	2,465	8,680
GS	Goldman Sachs Group Inc	91,057	10,050	13,850
HD	Home Depot Inc	30,923	2,374	6,316
IBM	International Business Machines Co	101,077	9,070	15,986
INTC	Intel Corp	68,079	4,802	13,630
JNJ	Johnson & Johnson	40,503	2,775	7,314
JPM	JPMorgan Chase and Co	108,810	19,762	12,412
KO	Coca-Cola Co	45,339	3,031	6,865
MCD	McDonald's Corp	45,971	6,311	6,312
MMM	3M Co	17,001	794	2,846
MRK	Merck & Co Inc	54,986	1,875	8,125
MSFT	Microsoft Corp	183,184	12,278	30,532
NKE	Nike Inc	29,220	1,927	7,523
PFE	Pfizer Inc	71,415	3,243	10,705
PG	Procter & Gamble Co	25,751	1,566	3,530
T	AT&T Inc	75,886	2,804	9,024
TRV	Travelers Companies Corp	12,184	912	1,587
UNH	UnitedHealth Group Inc	15,020	2,051	2,555
UTX	United Technologies Corp	16,123	995	3,065
V	Visa Inc	43,375	2,786	6,785
VZ	Verizon Communications Inc	45,177	2,284	8,508
WMT	Wal-Mart Stores Inc	63,405	8,562	7,318
XOM	Exxon Mobil Corp	46,286	3,381	7,275
Total		**1,555,770**	**126,713**	**246,912**

Table 4.1: Data overview

term related to the company for unrelated searches. For instance, it maps the number of searches for DuPont (a chemical company) adjusted for searches DuPont referring to DuPont analysis of ROE decomposition.

Google searches are normalized. In other words, for each selected period, the maximum equals to 100 and the rest of the time series is adjusted accordingly to preserve the trend. We need data for the time interval from May 2013 to September 2014. Since the longest period, for which the daily data are available, is 3 months we need to control for the period-specific normalization. We downloaded the overlapped data such as: May13-July13, July13-Sep13, Sep13-Nov13 etc. and normalized them over the entire period. We used the overlapping month to narrow imbalance of the subsamples. The weight W_i equals to a ratio of the i^{th} observations G_i from the two consecutive periods t and $t+1$ (Equation 4.1), where i denotes a date. We calculated an average misalignment of the consecutive subsamples and used it as a weighing factor for the latter period observations (Equation 4.2). Repeating this process, we obtained a consistent chain of searches S_i for the entire period.

$$W_i = \frac{G_i^{t+1}}{G_i^t} \tag{4.1}$$

$$S_i = \frac{G_i^{t+1} \cdot \sum_{i=1}^{n} W_i}{n} \tag{4.2}$$

The search volume follows similar pattern for all companies except Disney, Home Depot, McDonnalds', Nike and Walmart. The searches are highest on Tuesday, Wednesday and Thursday and moderate on Monday and Friday (Table 4.2). Searching activity decreases by average 66% on weekends. With respect to internet activity, this behaviour resembles an email open rate pattern revealed by many marketing agencies. According to Pietras (2013), the lowest email open rate is on Monday (18.2%) whereas the highest open rate is on Tuesday (19.9%). Tuesday is the day with the highest amount of email sent, too. Wednesday and Thursday also display high open rate and sent rate. The reasonable explanation for this is that people begin their work week on Monday and are overwhelmed by emails, planning and backlogs from the last week. On the other hand, on Tuesday, employees catch up with their work and are ready to solve new tasks. Similar reasoning could be used for high search volume on Tuesday, Wednesday and Thursday.

In terms of emails, Friday has the second higher open rate (19.6%). This could be caused by necessity to close up many weekly issues before a weekend. However, there is an opposing force of shifting people's focus to a weekend and postponing important decisions on the next week. As it is visible at Google searches, the second force overrules the first one and Friday's average search volume is only slightly higher than Monday's volume.

Moreover, revealed pattern of Google search volume is in line with research on trading activity. The day-of-the-week effects is a common phenomenon of trading activity and volatility. Chordia et al. (2001) argue that Fridays accompany significant decrease in both trading activity and liquidity, while Tuesdays show the opposite pattern. Berument and Kiymaz (2001) have examined several patterns regarding stock market volatility and returns. They studied the S&P 500 market index during the period of January 1973 and October 1997 and found presence of the week effect in volatility and returns. Wednesdays are associated with the highest returns and the lowest volatility while the highest volatility is observed on Friday and the lowest returns are observed on Monday.

Day of the week	**Search volume**	**Email open rate (%)**	**Volume of email sent (%)**
Monday	67	18.2%	16.59%
Tuesday	71	19.9%	17.93%
Wednesday	70	19.0%	16.08%
Thursday	69	18.9%	17.25%
Friday	67	19.6%	14.90%
Saturday	48	16.9%	8.58%
Sunday	45	17.1%	8.68%
Weekly average	62	18.61%	14.28%
Mon - Fri average	69	19.26%	16.55%

Table 4.2: Weekly pattern of Google search volume and emails

Regarding the five above mentioned companies with a different search pattern, the weekend searches are on average 17% higher than those during the workdays. Interest in these companies might be induced by general public searching for their services on the internet. Those searches overweight searching volume induced by financial analysts.

4.1.3 Twitter data

The second proxy of market sentiment stems from the micro-blogging platform Twitter and consists of relevant tweets during the period between June 1, 2013 and September 18, 2014. The data were used by Ranco et al. (2015) in their paper examining the effect of Twitter sentiment on stock price returns. The data was collected by the Twitter Search API using the stock cash-tag (e.g. BA for Boeing). All available tweets with cash-tags should be acquired. In total, the dataset consists of more than 1.5 million tweets sorted by their sentiment (Table 4.1).

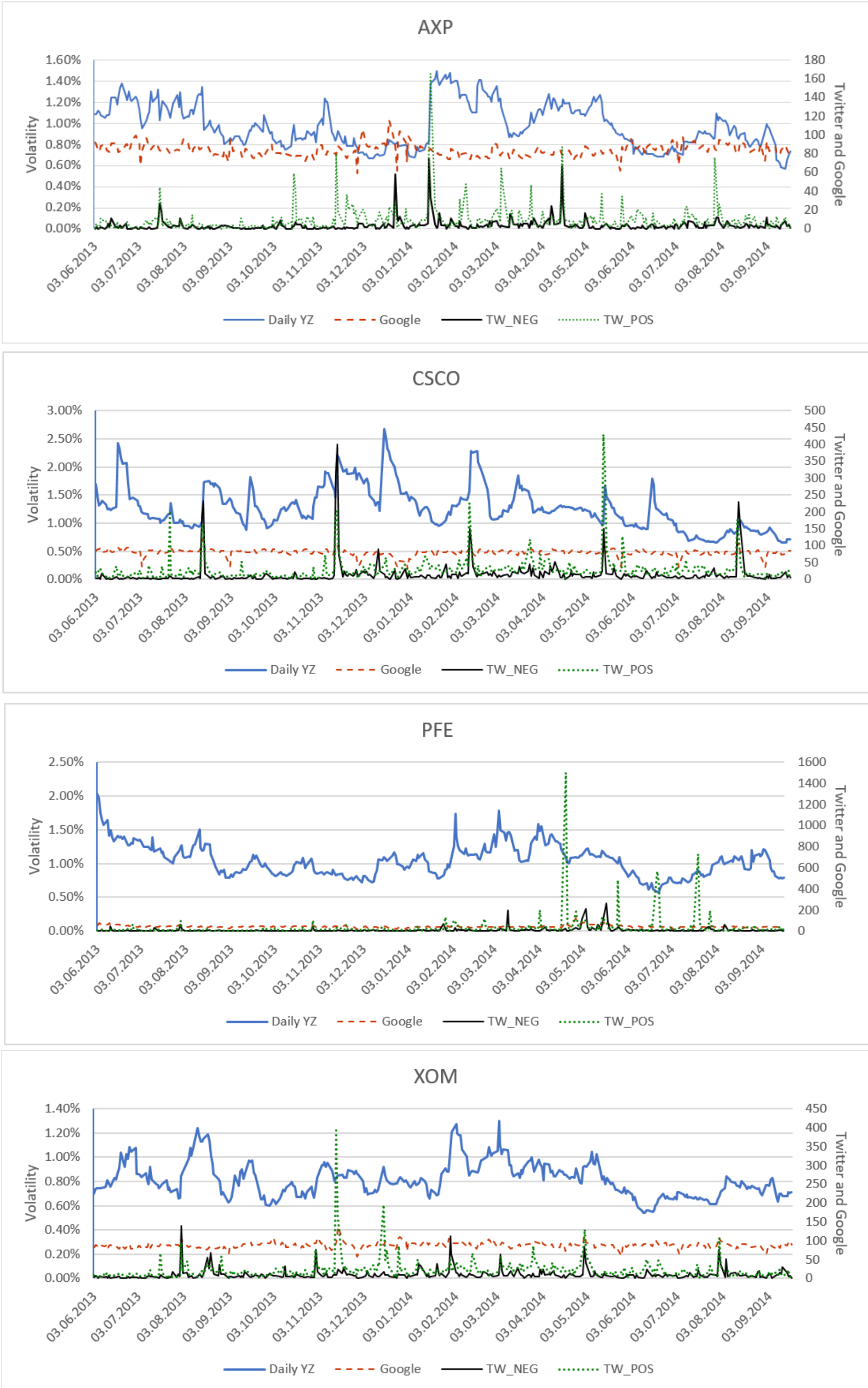


Figure 4.1: Volatility, Twitter data and Google data, Source: based on Ranco et al. (2015) and own analysis

The method of sentiment evaluation consists of two phases. Firstly, over 100,000 of tweets were evaluated by financial experts who labelled them with three sentiment labels: negative, neutral, positive. Afterwards, this labelling was used to set Support Vector Machine classification model, which was applied to the entire set of 1.5 million tweets (Ranco et al., 2015). The final dataset comprises the time series of total, negative, positive and neutral tweets for each day and company.

Figure 4.1 depicts volatility measured by Yang-zhang estimator, a number of positive and negative tweets, and Google search index of four randomly chosen stocks over time. Volatility is assign to the left axis while the rest is measured on the right axis. Visually, CSCO's tweet spikes corresponds to upturns in volatility better than those of AXP. On the other hand, PFE lacks visual link between tweets and volatility. Generally, negative tweets show higher fidelity than positive tweets. Google search volume is quite steady in all charts.

4.1.4 Winsorizing

The examined period has stable volatility and does not include any market crashes or bubbles. Also, the internet data are distributed steadily. There are six companies in the sample that show spikes in positive, neutral or negative tweets that are not reflected in neither market activity nor Google searches. These spikes are presented in the table below.

The spikes might be a product of wrong sentiment evaluation methodology or data collection. If we would suppose that those values are correct, the explanations may be twofold. Either, missing cardinality of sentiment evaluation equals market news that have different magnitude. Or, the spikes represent market anomaly or irrelevant event.

To reduce the effect of the outliers, we winsorized the time series (Dixon, 1960). This transformation replaces extreme values by upper and lower limits. Since the data are constrained by 0, we capped only the right tail outliers. We set mean (μ) plus three standard deviations (σ) as a limiting value. For normal distribution, this value corresponds to 99.73% probability distribution. Using this simple heuristic, we can assume that nearly all values lie within this interval.

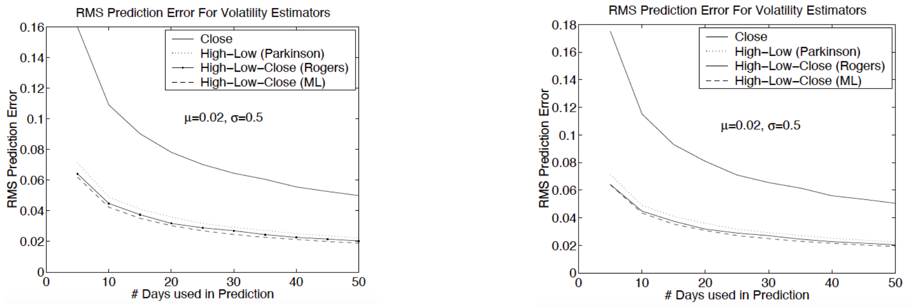
To support reliability of our conclusions, we performed robustness analysis. Additionally to $\mu + 3\sigma$ limit, we transformed data using $\mu + 2.5\sigma$, $\mu + 2\sigma$, and $\mu + 1.5\sigma$ values. More about robustness analysis can be found in the Section 4.4.4.

Company	Variable	Mean	St dev	Spike value	Spike date
Home Depot Inc	$T^{n_3,-}$	7.00	20.31	311	20.05.2014
Home Depot Inc	$T^{n_3,+}$	18.63	50.83	266	20.05.2014
Home Depot Inc	$T^{n_3,0}$	55.44	79.84	632	20.05.2014
IBM	$T^{n_3,+}$	47.16	76.70	1165	03.07.2014
IBM	$T^{n_3,+}$	47.16	76.70	339	15.07.2014
IBM	$T^{n_3,+}$	47.16	76.70	313	16.07.2014
IBM	$T^{n_3,+}$	47.16	76.70	368	17.07.2014
IBM	$T^{n_3,0}$	187.20	234.53	764	15.07.2014
IBM	$T^{n_3,0}$	187.20	234.53	2104	16.07.2014
IBM	$T^{n_3,0}$	187.20	234.53	1755	17.07.2014
Merck & Co Inc	$T^{n_3,+}$	23.97	88.50	1354	09.04.2014
Merck & Co Inc	$T^{n_3,+}$	23.97	88.50	833	10.04.2014
Pfizer Inc	$T^{n_3,+}$	31.58	102.96	1497	21.04.2014
Pfizer Inc	$T^{n_3,+}$	31.58	102.96	483	27.05.2014
Pfizer Inc	$T^{n_3,+}$	31.58	102.96	563	23.06.2014
Pfizer Inc	$T^{n_3,+}$	31.58	102.96	466	24.06.2014
Pfizer Inc	$T^{n_3,+}$	31.58	102.96	729	21.07.2014
Pfizer Inc	$T^{n_3,+}$	141.11	299.00	1219	21.04.2014
UnitedHealth Group Inc	$T^{n_3,-}$	6.05	54.97	1009	17.04.2014
UnitedHealth Group Inc	$T^{n_3,0}$	25.47	23.25	213	17.04.2014
UnitedHealth Group Inc	$T^{n_3,+}$	7.54	10.32	40	17.04.2014
Visa Inc	$T^{n_3,-}$	8.22	19.94	139	25.04.2014
Visa Inc	$T^{n_3,0}$	78.01	82.49	295	25.04.2014
Visa Inc	$T^{n_3,+}$	20.01	56.80	1005	25.04.2014

Table 4.3: Spikes in Twitter variables

4.2 Measuring realized volatility

The original HAR model assumes high-frequency data. Since we do not use high frequency data, we need to seek for an alternative solution, i.e. a volatility estimator which is able to handle low frequencies. We focus on traditional methods of volatility estimation using high, low and close price data, which are available. Magdon-Ismail and Atiya (2003) introduced maximum likelihood approach to the volatility modelling for an instrument following Brownian motion. They compare the results with three classical methods. The first estimator uses the close prices only (i.e. Close), the second one is Parkinson's estimator using the high and low values (Parkinson, 1980), and the third method is Rogers-Satchell estimator using the high, low and close prices (Rogers and Satchell, 1991). Comparison of RMS (root-mean-square) prediction error is depicted in the figure 4.2.



(a) The mean is known

(b) The mean has to be estimated

Figure 4.2: Comparison of the volatility estimators, Source: Magdon-Ismail and Atiya (2003)

The maximum likelihood estimator obtains improvement relative to the other estimators especially with less observations. Brandt and Kinlay (2005) also investigate methods of measuring historical volatility. They challenge the result of previous researches using geometric Brownian motion, that alternative volatility estimators offer efficiency improvements over the standard deviation estimators. Such conclusion depends on unrealistic assumptions of nature of the underlying process. Typically, the process is assumed to be continuous geometric Brownian motion with constant volatility and zero drift. Every departure from the ideal Brownian motion such as process drift, opening gaps or time-varying volatility may have effect on the performance of both the standard deviation estimators and the alternative estimators. They compare efficiency of traditional estimators, namely Parkinson (1980), Garman and Klass (1980), Rogers and Satchell (1991), Alizadeh et al. (2002) and Yang and Zhang (2000), and integrated volatility estimator with respect to sample size and frequency on simulated data. In comparison with classical close to close estimator, these estimators have 5 times greater efficiency, or even 7 to 8 times greater efficiency in the case of the German-Klass and Yang-Zang estimators. They conclude that each of selected estimators, except Alizadeh-Brandt-Diebold estimator, produces biased estimates unless very high frequencies are used. The anomalies further deteriorate performance of selected estimators. Since departures from geometric Brownian motion are almost certain, these findings call into question additional information obtained by the alternative (for instance maximum-likelihood) estimators.

Rogers and Satchell (1991) do not neglect the issue of non-zero process drift and thus their estimator significantly outperforms others with presence of time-varying drift in asset process. Yang and Zhang (2000) build their volatility estimator on a solid base of the Rogers-Satchel estimator. Their estimator is able to cope with a time-varying level of drift and is equipped to handle open price jumps. Since large amounts of information arrive during the period when the market is closed, there is a price gap between the opening and closing price. With respect to Yang-Zhang estimator ability to deal with these two common anomalies and its robust performance in Brandt and Kinlay measurement, we decided to use it for our study.

The Yang and Zhang estimator is composed by overnight volatility σ_o, open to close volatility σ_c and Rogers-Satchell estimator σ_{RS}. The Yang-Zhang estimator (Equation 4.3) is given by:

$$\sigma_{YZ} = \sqrt{F}\sqrt{\sigma_o^2 + k\sigma_c^2 + (1-k)\sigma_{RS}^2} \tag{4.3}$$

$$\sigma_o^2 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}[ln(\frac{o_i}{c_{i-1}}) - ln(\overline{\frac{o_i}{c_{i-1}}})]^2}$$

$$\sigma_c^2 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}[ln(\frac{o_i}{c_i}) - ln(\overline{\frac{o_i}{c_i}})]^2}$$

$$\sigma_{RS} = \sqrt{\frac{F}{N}\sum_{i=1}^{N} ln(\frac{h_i}{o_i})ln(\frac{h_i}{c_i}) + ln(\frac{l_i}{o_i})ln(\frac{h_i}{c_i})}$$

where F denotes frequency, N denotes the number of observations, h denotes the high price, l denotes the low price, c denotes the close price and o denotes the opening price.

For the given data point t, we examine volatility using prices from previous 21 days and the day t. In order to obtain weekly and monthly volatilities, we use moving average of daily volatilities of previous 5 and 22 days, respectively, as it corresponds to the number of trading days.

4.3 HAR specification

We constructed a three-levels volatility cascade with time horizons $n_1 = 22$ days, $n_2 = 5$ days and $n_3 = 1$ day. Thus, we arrive to the following equation:

$$V_t^{n_3} = \beta_0 + \beta_1\hat{V}_{t-1}^{n_3} + \beta_2\hat{V}_{t-1}^{n_2} + \beta_3\hat{V}_{t-1}^{n_1} + \epsilon_t \tag{4.4}$$

where,

$$\hat{V}_{t-1}^{n_3} = \sigma_{YZ,t-1}$$

$$\hat{V}^{n_2}_{t-1} = \frac{1}{5}\sum_{i=1}^{5}\sigma_{YZ,t-i}$$

$$\hat{V}^{n_1}_{t-1} = \frac{1}{22}\sum_{i=1}^{22}\sigma_{YZ,t-i}$$

$\hat{V}^{n_3}_{t-1}$ is the daily volatility obtained directly from Yang-Zhang estimator, $\hat{V}^{n_2}_{t-1}$ is the weekly volatility, and $\hat{V}^{n_1}_{t-1}$ is the monthly volatility, and ϵ is an error term. Weekly and monthly volatilities equal moving average of previous 5 or 22 days, respectively.

We extended the basic model by Twitter sentiment and Google search volume. We defined fifty specifications including the basic HAR. All models were estimated by Cochrane-Orcutt method. We did not control for stationarity because the HAR model implicitly assumes cointegration among the volatility cascade. To confirm this notion, we performed Engleâ€“Granger procedure.

4.3.1 Searching for the right specification

Due to a short period used for in-sample prediction (200 observations), the model can be easily over-specified by a large amount of market sentiment variables. Hence, we gradually added variables to balance trade-off of additional information and measurement error. Firstly, we enriched the basic model (4.4) by individual variables, as it is presented in the table 4.4.

Name	Tag	Description
Daily positive tweets	$T^{n_3,+}$	daily values
Daily negative tweets	$T^{n_3,-}$	daily values
Daily neutral tweets	$T^{n_3,0}$	daily values
Daily search volume	S^{n_3}	daily values
Weekly positive tweet	$T^{n_2,+}$	5-days moving average
Weekly negative tweets	$T^{n_2,-}$	5-days moving average
Weekly search volume	S^{n_2}	5-days moving average
Monthly positive tweet	$T^{n_1,+}$	22-days moving average
Monthly negative tweets	$T^{n_1,-}$	22-days moving average
Monthly search volume	S^{n_1}	22-days moving average

Table 4.4: Market sentiment variables

According to the performance of individual variables, we constructed various models including couples, triplets and more examined indicators as presented in the list 4.3.1. A list of all specifications can be found in the Appendix A.

- Daily positive tweets + Daily search volume
- Daily negative tweets + Daily search volume
- Daily positive tweets + Weekly search volume
- Daily negative tweets + Monthly search volume
- Weekly positive tweets + Daily search volume
- Monthly positive tweets + Daily search volume
- Daily positive tweets + Daily negative tweets + Daily search volume
- Daily positive tweets + Daily negative tweets + Monthly search volume
- Daily positive tweets + Daily negative tweets + Monthly positive tweets + Monthly negative tweets
- Daily positive tweets + Daily negative tweets + Daily neutral tweets + Daily search volume
- Daily positive tweets + Daily negative tweets + Monthly positive tweets + Monthly negative tweets + Monthly search volume
- etc.

List 4.3.1: An example of model extensions

Additionally, we constructed an alike cascade of market sentiment variables to preserve dynamic of the model. Thus, we create weekly, and monthly values of each variable using 5 and 22 days moving averages of the daily values respectively. The resulting equation is given by:

$$\begin{aligned} V_t^{n_3} = \beta_0 + \beta_1 \hat{V}_{t-1}^{n_3} + \beta_2 \hat{V}_{t-1}^{n_2} + \beta_3 \hat{V}_{t-1}^{n_1} + \gamma_1 T_t^{n_3,+} + \gamma_2 T_t^{n_3,-} + \gamma_3 T_t^{n_3,0} + \\ + \gamma_4 T_t^{n_2,+} + \gamma_5 T_t^{n_2,-} + \gamma_6 T_t^{n_2,0} + \gamma_7 T_t^{n_1,+} + \gamma_8 T_t^{n_1,-} + \gamma_9 T_t^{n_1,0} + \gamma_{10} S_t^{n_3} + \\ + \gamma_{11} S_t^{n_2} + \gamma_{12} S_t^{n_1} + \epsilon_t \end{aligned} \tag{4.5}$$

where T^+ denotes a number of positive tweets, T^- denotes a number of negative tweets, T^0 denotes a number of neutral tweets, and S represents the searching volume. Each variable follows the month-week-day cascade so that $T^{n_3,-}$ denotes daily negative tweets, $T^{n_2,-}$ stands for weekly average of negative tweets etc. We assume that as the volatility cascade captures different treatment of information with respect to different investment horizons, risk aversion and other investor-specific characteristics, the sentiment cascade could capture different effect of market sentiment on various types of investors.

4.4 Out-of-sample forecast approach

Firstly, we divided the time series into a training set (200 observations) and a test set (138 observations). Then, we estimated the parameters using the training set and performed an out-of-sample forecast using the test set. To evaluate accuracy of our models, we used following measures: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Scaled Error (MASE). Furthermore, we ran Diebold-Mariano test to test statistical significance of resulting forecasts.

4.4.1 Scale-dependent errors

The scale dependent errors are based on forecast errors $e_i = y_i - \hat{y}_i$. They are on the same scale as the data and thus cannot be used to make comparisons between series that are on different scales. The two scale-dependent measures that are used in this study are Mean absolute error (MAE), and Root Mean Square Error (RMSE) (Hyndman and Athanasopoulos, 2014).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|e_i| \tag{4.6}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}e_i^2} \tag{4.7}$$

Both measures are indifferent to the direction of errors and can range from 0 to infinity, where lower values indicate the better performing model. The main difference is that RMSE gives relatively high weight to large errors as the errors are squared before they are averaged. The high sensitivity to outliers is particularly useful when large errors are undesirable. RMSE grows larger than MAE with increasing test samples and thus may be problematic when comparing different sized test samples (Chai and Draxler, 2014). This is not our concern as all test samples have equal number of observations. To increase reliability of our study, we use both measures.

4.4.2 Scaled errors

Scaled errors were proposed by Hyndman and Koehler (2006) as an alternative to percentage errors for comparing forecast accuracy across series on different scales. The percentage errors have extreme values or are undefined when y_i is close to zero. This might be easily the case of volatility estimators and thus we decided for Hyndman and Koehler (2006) specification.

The Mean absolute scaled error (MASE) is MAE scaled by scaling statistics Q calculated from in-sample naïve forecasts for non-seasonal time series(Equation 4.8), in-sample seasonal naive forecasts for seasonal time series (Equation 4.9) and in-sample mean forecasts for non-time series data (Equation 4.10) (Hyndman and Koehler, 2006).

$$Q = \frac{1}{N-1} \sum_{i=2}^{N} |y_i - y_{i-1}| \tag{4.8}$$

$$Q = \frac{1}{N-m} \sum_{i=m+1}^{N} |y_i - y_{i-m}| \tag{4.9}$$

$$Q = \frac{1}{N} \sum_{i=1}^{N} |y_i - \bar{y}| \tag{4.10}$$

$$MASE = \frac{MAE}{Q} \tag{4.11}$$

MASE is asymptotically normal and symmetric, i.e. positive and negative errors are penalized equally as well as errors in large and small forecasts. It is easily interpretable. MASE is lower than one if it arises from a better forecast than the average naïve forecast computed on the training set and greater than one if the forecast is worse.(Hyndman and Koehler, 2006)

4.4.3 Diebold-Mariano test

Error-based measures compare forecast accuracy across the examined sample. Thus, the result does not say anything about statistical significance of the forecasts in the population. Diebold-Mariano test (DM test) makes inference on the population rather than on the sample (Diebold and Mariano, 1995) and therefore shows statistical strength of predictions. Diebold (2015) does not encourage using it for comparing models so we use it as a complementary measure to look at the phenomenon from a different angle.

The DM test compare forecast accuracy of two forecasts (Equation 4.12). Under the null hypothesis, competing forecasts have the same accuracy. We use the one-sided version of the test, where the alternative hypothesis is that second forecast is more accurate than the first one.

$$\begin{aligned} d_t &= g(e_{1t}) - g(e_{2t}) \\ H_0 &: E(d_t) = 0 \quad \forall t \\ H_A &: E(d_t) > 0 \end{aligned} \tag{4.12}$$

where $g(e_{1t})$ and $g(e_{2t})$ are loss functions of the first and the second forecast respectively.

The nature of DM test may be particularly problematic when comparing forecasts from nested models. If the null hypothesis is true, the forecast errors from the examined models are exactly the same and perfectly correlated at a population level. Hence, the numerator and denominator of a DM test are each limiting to zero as the estimation sample grows. But, when the size of the estimation sample remains finite, parameters are prevented from reaching their probability limits and the DM test is asymptotically valid. (Giacomini and White, 2003)

4.4.4 Ranking

To decide on predictive power of particular models, we created ranking system based on the obtained measurement errors. This procedure is described in steps below.

1. All 50 models were executed for all 30 companies.
2. The models were ordered in ascending orders according to obtained measurement errors, i.e. the first model shows the lowest error and the last one shows the highest error.
3. Percental difference (δ) relative to the winning model was assigned to each specification. Such that:

$$\delta_1 = 0\%$$

$$\delta_i = \frac{Error_i - Error_1}{Error_1} \qquad i \in \langle 2, 30 \rangle \subset \mathbb{N}$$

   where i stands for the ordered model, i.e. $i = 2$ for the second best performing model, $i = 2$ for the third best performing model etc., and *Error* refers to MAE, RMSE, and MASE respectively.
4. Total error rate Δ^j and average error rate $\bar{\delta^j}$ were calculated by averaging and summing deltas for each specification over the 30 companies. Such that:

$$\bar{\delta^j} = \frac{1}{n}\sum_{i=1}^{30} \delta_i^j$$

$$\Delta^j = \sum_{i=1}^{30} \delta_i^j$$

where i denotes a company and j indicates a model. This has been done for each error separately.

5. The models were ordered according to the sum of deltas Δ^j, which is scale independent. We obtained three different rankings as it has been done for each measurement error separately.
6. This procedure was executed for all dataset, i.e. original data, and $\mu+3\sigma$, $\mu + 2.5\sigma$, $\mu + 2\sigma$, and $\mu + 1.5\sigma$ winsorized data.

Top ten performing models according to each measurement error were further investigated.

Chapter 5

Results

This chapter presents results of out-of-sample forecast and robustness analysis. Firstly, out-of-sample approach is summarized. The measurement errors are presented and models' predictive power is evaluated. Moreover, results of Diebold-Mariano tests are outlined. Secondly, an output of the robustness analysis is discussed.

5.1 Out-of-sample forecast results

The tables 5.1 and 5.2 summarize averaged measurement errors $\bar{\delta}^j$ and summed errors Δ^j of the top ten performing models and the basic model. Complete results can be found in the Appendix A. The first table presents error rates based on the original dataset whereas the other shows errors stemming from the $\mu + 3\sigma$ winsorized time series.

The average error rate and the total error rate can be interpreted as indicators of the model fit. A benchmark is the best performing model, i.e. the model with lowest RMSE, MASE and MAE respectively, for each company. A unit of the average error rate is percent, as it averages the percentage error rates (δ_i) across examined companies. It can be interpreted as model's average error relatively to other examined models. On the other hand, the total error rate does not have any direct interpretation. It is a sum of the percentage error rates (δ_i) and thus it represents an absolute indicator of the model fit.

There is significant disparity between errors obtained from the original dataset and the winsorized dataset. Deltas have greater variance as well as a larger mean for the original data suggesting model sensitivity to extreme values. Speaking of MASE and MAE, the difference between the best and worst performing model is approximately 2.5 larger for the original data than for

	Average error rate $\tilde{\delta^j}$ (%)			Total error rate Δ^j		
Model	**RMSE**	**MASE**	**MAE**	**RMSE**	**MASE**	**MAE**
S^{n_2}	0.36	0.00	0.00	1.79	2.34	2.33
S^{n_3}	0.00	0.09	0.18	1.74	2.42	2.41
HAR basic	0.10	0.78	0.50	1.77	2.62	2.62
$T^{n_3,-} + S^{n_2}$	11.98	1.41	1.50	5.87	2.86	2.85
S^{n_1}	0.58	1.57	1.49	1.92	2.88	2.88
$T^{n_3,-}$	12.16	2.09	1.98	5.98	3.11	3.11
$T^{n_3,-} + S^{n_3}$	12.33	2.13	2.23	6.05	3.14	3.13
$T^{n_3,-} + S^{n_1}$	12.24	2.28	2.44	6.03	3.23	3.22
$T^{n_3,+} + S^{n_2}$	15.73	4.21	4.01	6.29	3.56	3.55
$T^{n_2,-} + S^{n_2}$	13.13	3.75	3.76	6.19	3.66	3.66
$T^{n_3,+}$	15.24	4.68	4.34	6.17	3.74	3.73
Average error	8.53	2.09	2.04	4.53	3.05	3.05

Table 5.1: Average and total error rates, Original data, Source: own analysis

the winsorized one. Also, the average delta is larger for the original data, i.e. 0.0209 for MASE and 0.0204 for MAE on average, than for the spike-adjusted data, i.e. 0.0090 for MASE and 0.0083 for MAE on average.

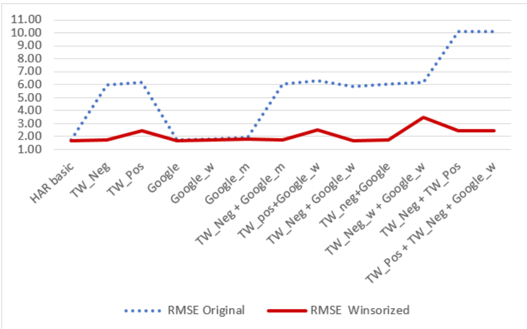
In terms of RMSE, the difference is even more apparent. No wonder, as RMSE penalizes large errors, Twitter-driven outliers pronouncedly affect the overall performance of our models. The difference between the best and worst performing model is approximately 27 times larger for the original data than for the winsorized one and the average delta is 10 times larger.

The common feature of all top-performing models is an low amount of extending variables. The most complex model contains three sentiment variables in addition to the basic volatility cascade. Even among these specifications, the more parsimonious models win. Since we estimate the models over a relatively short period, the regression equations might become easily overspecified. Also, all additional variables are a proxy of market sentiment. Using a proxy variable is naturally accompanied by risks of inappropriate selection. Assuming that google searches and Twitter data well-capture market sentiment, they are still burdened by measurement error and unrelated noise. Inappropriate methodology in data collection or data evaluation might be a source of such noise. Despite all limitations, selected variables contribute worthwhile information.

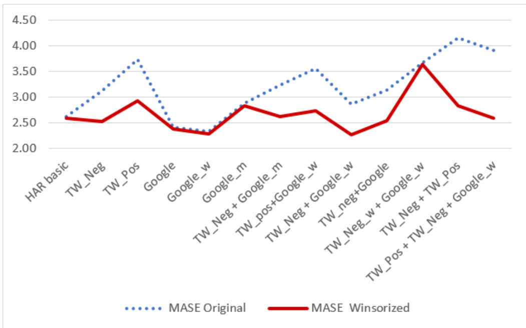
The Figure 5.1 shows development of an error rate of the top performing models. The total error rate is measured by total sum of RMSE deltas and

Model	Average error rate $\bar{\delta}^j$ (%)			Total error rate Δ^j		
	RMSE	**MASE**	**MAE**	**RMSE**	**MASE**	**MAE**
$T^{n_3,-} + S^{n_2}$	0.08	0.00	0.00	1.67	2.27	2.26
S^{n_2}	0.36	0.21	0.21	1.71	2.29	2.29
S^{n_3}	0.00	0.31	0.40	1.66	2.38	2.36
$T^{n_3,-}$	0.07	0.66	0.46	1.71	2.53	2.52
$T^{n_3,-} + S^{n_3}$	0.21	0.70	0.70	1.77	2.55	2.55
$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	2.78	1.09	1.05	2.44	2.59	2.58
HAR basic	0.10	0.99	0.71	1.69	2.59	2.58
$T^{n_3,-} + S^{n_1}$	0.16	0.85	0.92	1.76	2.63	2.62
$T^{n_3,+} + S^{n_2}$	3.33	1.83	1.64	2.54	2.74	2.73
S^{n_1}	0.58	1.79	1.71	1.83	2.83	2.83
$T^{n_3,-} + T^{n_3,+}$	2.68	1.77	1.58	2.45	2.84	2.83
Average error	0.94	0.93	0.85	1.93	2.57	2.56

Table 5.2: Average and total error rates, $\mu + 3\sigma$ winsorized data, Source: own analysis



(a) RMSE



(b) MASE

Figure 5.1: Original and winsorized errors comparison, Source: own analysis

MASE deltas respectively across all companies as described in the Chapter 4. The term is more stable for the winsorized data. The difference is naturally present only in specifications containing Twitter variables $T^{n_3,-}$, $T^{n_2,-}$, and $T^{n_2,-}$ because google searches and volatility estimators were not adjusted.

By creating percentage difference δ, we obtained scale-independent MAE and thus, it is not surprising that MAE and MASE results are similar. On the other hand, RMSE based deltas differ due to the sensitivity to large errors. Therefore, we present two different rankings according to the used error (see Tables 5.3 and 5.4).

As can be seen, limiting the extreme values have a significant effect on forecast accuracy. Spurious effect of the Twitter spike tumble performance of the

Rank	MASE (MAE) total error rate ranking	RMSE total error rate ranking
1	S^{n_2}	S^{n_3}
2	S^{n_3}	HAR basic
3	HAR basic	S^{n_2}
4	$T^{n_3,-} + S^{n_2}$	S^{n_1}
5	S^{n_1}	$T^{n_3,-} + S^{n_2}$
6	$T^{n_3,-}$	$T^{n_3,-}$
7	$T^{n_3,-} + S^{n_3}$	$T^{n_3,-} + S^{n_1}$
8	$T^{n_3,-} + S^{n_1}$	$T^{n_3,-} + S^{n_3}$
9	$T^{n_3,+} + S^{n_2}$	$T^{n_3,+}$
10	$T^{n_2,-} + S^{n_2}$	$T^{n_2,-} + S^{n_2}$
11	$T^{n_3,+}$	$T^{n_3,+} + S^{n_2}$

Table 5.3: Top ten performing models, Original data, Source: own analysis

Rank	MASE (MAE) total error rate ranking	RMSE total error rate ranking
1	$T^{n_3,-} + S^{n_2}$	$T^{n_3,-} + S^{n_2}$
2	S^{n_2}	S^{n_3}
3	S^{n_3}	HAR basic
4	$T^{n_3,-}$	$T^{n_3,-}$
5	$T^{n_3,-} + S^{n_3}$	S^{n_2}
6	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	$T^{n_3,-} + S^{n_1}$
7	HAR basic	$T^{n_3,-} + S^{n_3}$
8	$T^{n_3,-} + S^{n_1}$	S^{n_1}
9	$T^{n_3,+} + S^{n_2}$	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$
10	$T^{n_3,-} + T^{n_3,+}$	$T^{n_3,-} + T^{n_3,+}$
11	S^{n_1}	$T^{n_3,+} + S^{n_2}$

Table 5.4: Top ten performing models, μ+3σ winsorized data, Source: own analysis

model regardless of additional information. Whilst the basic model occupies the second, respectively the third, position for the original dataset, six, respectively two, extended models outperform the basic model for the winsorized data. RMSE ranking is more sensitive to noise induced by the internet variables and thus, the basic model performs relatively better. However, the overall results are satisfactory as both rankings contain eight common specifications. It indicates a certain level of robustness to the error methodology.

Hereafter, we decided to focus on results based the winsorized data only for reasons discussed above. Differences in predictive power of the individual specifications are rather subtle. The average error rate of the basic model is only 0.99% larger than that of the winning model in case of MASE. In terms of RMSE is this difference even smaller, approximately 0.1%. Even difference between the best and worst performing model does not exceed 2% in terms of MASE and MAE.

RMSE offer interesting view into impact of positive tweets. Whereas $\bar{\delta}^j$ takes value from 0% to 0.58% for most of the selected models, it goes up to 3% for specifications containing positive tweets. It suggests that there is more irrelevant noise in the positive tweets than in the negative tweets.

The table 5.5 shows average values of MASE of the top performing models. The complete overview of average MASE, MAE and RMSE can be found in the Appendix A. All values are very closed to 1 and varies in the third decimal place. It indicates high quality of the original model. Nevertheless, six models including the basic HAR attain values lower than 1 a thus outperform a naïve forecast.

Model	**MASE**
$T^{n_3,-} + S^{n_2}$	0.99223
S^{n_2}	0.99437
S^{n_3}	0.99618
$T^{n_3,-}$	0.99675
$T^{n_3,-} + S^{n_3}$	0.99918
HAR basic	0.99931
$T^{n_3,-} + S^{n_1}$	1.00135
$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	1.00266
$T^{n_3,-} + T^{n_3,+}$	1.00787
$T^{n_3,+} + S^{n_2}$	1.00854
S^{n_1}	1.00919

Table 5.5: Average MASE, $\mu + 3\sigma$ winsorized data, Source: own analysis

We performed Diebold-Mariano test to reveal statistical significance of predictive power of examined models. We set an alternative hypothesis that the extended model outperforms the basic HAR. Tables 5.6 and 5.7 present the test results. Apart from six companies, i.e. Boeing, Caterpillar, Goldman Sachs Group, 3M, Merck & Co, and Procter & Gamble, p-values always exceeded 10% level. We conclude that the results are not statistically significant. However, it is not a surprising result considering negligible differences in the model performance.

As we stated in the Chapter 4, we performed Engle-Granger test to check for cointegration. We ran a Dickey-Fuller test on residuals of the basic HAR model. A resulting p-value is smaller than 0.01. Thus, we can confirm that volatility variables are cointegrated.

Model/Company	**AXP**	**BA**	**CAT**	**CSCO**	**CVX**	**DD**
S^{n_3}	0.3762	0.2313	0.2723	0.3018	0.4312	0.1976
S^{n_2}	0.4681	0.1284	0.1483	0.2661	0.3893	0.3342
S^{n_1}	0.3676	0.2509	0.3182	0.4576	0.4416	0.3084
$T^{n_3,+}$	0.3793	0.1598	0.0504	0.4055	0.3143	0.2007
$T^{n_3,-}$	0.2156	0.2368	0.2203	0.2904	0.1853	0.1724
$T^{n_3,-} + S^{n_1}$	0.2098	0.2105	0.2193	0.2909	0.1802	0.1500
$T^{n_3,-} + S^{n_2}$	0.2166	0.1358	0.1895	0.2547	0.1836	0.1892
$T^{n_3,-} + S^{n_3}$	0.2093	0.2233	0.2193	0.2805	0.1680	0.1700
$T^{n_3,+} + S^{n_2}$	0.3791	0.1065	0.0435	0.2630	0.3015	0.2089
$T^{n_3,-} + T^{n_3,+}$	0.2863	0.1607	0.1570	0.2701	0.1847	0.1762
$T^{n_2,-} + S^{n_2}$	0.3618	0.0706	0.1567	0.1356	0.3895	0.3395
$T^{n_3,+} + T^{n_3,-} + S^{n_2}$	0.2853	0.1080	0.1377	0.2401	0.1829	0.1838
Model/Company	**DIS**	**GE**	**GS**	**HD**	**IBM**	**INTC**
S^{n_3}	0.3903	0.1778	0.3147	0.3934	0.2664	0.4808
S^{n_2}	0.4387	0.3121	0.4301	0.2005	0.1359	0.4926
S^{n_1}	0.4105	0.4855	0.4582	0.4653	0.1547	0.3977
$T^{n_3,+}$	0.2216	0.4888	0.0698	0.3174	0.3703	0.3106
$T^{n_3,-}$	0.2772	0.4801	0.1524	0.2126	0.3225	0.3659
$T^{n_3,-} + S^{n_1}$	0.2689	0.4785	0.1534	0.2126	0.3118	0.3582
$T^{n_3,-} + S^{n_2}$	0.2741	0.3192	0.1524	0.1904	0.2841	0.3682
$T^{n_3,-} + S^{n_3}$	0.2675	0.1858	0.1518	0.2125	0.3073	0.3644
$T^{n_3,+} + S^{n_2}$	0.2061	0.3113	0.0710	0.2898	0.1959	0.3126
$T^{n_3,-} + T^{n_3,+}$	0.1712	0.4762	0.1121	0.2114	0.3202	0.3118
$T^{n_2,-} + S^{n_2}$	0.3111	0.3224	0.3217	0.2060	0.2060	0.4215
$T^{n_3,+} + T^{n_3,-} + S^{n_2}$	0.1669	0.3142	0.1119	0.1894	0.2817	0.3135

Table 5.6: D-M test (p-values), $\mu + 3\sigma$ winsorized data, Source: own analysis

Model/Company	JNJ	JPM	KO	MCD	MMM	MRK
S^{n_3}	0.2595	0.2244	0.3692	0.1983	0.4462	0.1741
S^{n_2}	0.3354	0.3644	0.4671	0.3187	0.1893	0.4446
S^{n_1}	0.3667	0.4956	0.4752	0.4915	0.4794	0.4705
$T^{n_3,+}$	0.1391	0.2175	0.1500	0.4814	0.2986	0.0236
$T^{n_3,-}$	0.1905	0.4237	0.4140	0.2640	0.4138	0.0546
$T^{n_3,-} + S^{n_1}$	0.1749	0.4249	0.4059	0.2636	0.4128	0.0552
$T^{n_3,-} + S^{n_2}$	0.1366	0.3339	0.4016	0.2544	0.1786	0.0660
$T^{n_3,-} + S^{n_3}$	0.1655	0.2282	0.3317	0.2523	0.4081	0.0946
$T^{n_3,+} + S^{n_2}$	0.1283	0.2062	0.1971	0.3107	0.1781	0.0233
$T^{n_3,-} + T^{n_3,+}$	0.1203	0.1890	0.1470	0.2345	0.2987	0.0236
$T^{n_2,-} + S^{n_2}$	0.2079	0.3523	0.4146	0.2944	0.0474	0.1393
$T^{n_3,+} + T^{n_3,-} + S^{n_2}$	0.1065	0.1748	0.1963	0.2276	0.1784	0.0234
Model/Company	**MSFT**	**NKE**	**PFE**	**PG**	**T**	**TRV**
S^{n_3}	0.2080	0.2055	0.3763	0.2216	0.4741	0.3671
S^{n_2}	0.3671	0.2697	0.4558	0.3855	0.4311	0.4357
S^{n_1}	0.4915	0.3296	0.2742	0.4573	0.3942	0.4150
$T^{n_3,+}$	0.2184	0.2246	0.1877	0.4233	0.4443	0.4777
$T^{n_3,-}$	0.1408	0.2619	0.2436	0.2680	0.1650	0.2266
$T^{n_3,-} + S^{n_1}$	0.1408	0.2393	0.2000	0.2677	0.1660	0.2114
$T^{n_3,-} + S^{n_2}$	0.1433	0.2417	0.2409	0.2323	0.1578	0.2208
$T^{n_3,-} + S^{n_3}$	0.1387	0.2271	0.2379	0.0940	0.1647	0.2227
$T^{n_3,+} + S^{n_2}$	0.2184	0.1961	0.1805	0.3818	0.4187	0.4463
$T^{n_3,-} + T^{n_3,+}$	0.1910	0.2351	0.1557	0.2969	0.1677	0.1791
$T^{n_2,-} + S^{n_2}$	0.2220	0.2067	0.3569	0.0945	0.2540	0.3620
$T^{n_3,+} + T^{n_3,-} + S^{n_2}$	0.1910	0.2172	0.1493	0.2644	0.1618	0.1723
Model/Company	**UNH**	**UTX**	**V**	**VZ**	**WMT**	**XOM**
S^{n_3}	0.3747	0.1369	0.4851	0.4956	0.2935	0.3611
S^{n_2}	0.2415	0.3293	0.4251	0.4507	0.3923	0.1980
S^{n_1}	0.4132	0.2138	0.3897	0.4846	0.2800	0.3157
$T^{n_3,+}$	0.2463	0.2236	0.2530	0.3919	0.3987	0.2061
$T^{n_3,-}$	0.1703	0.4406	0.4113	0.3398	0.1212	0.3613
$T^{n_3,-} + S^{n_1}$	0.1706	0.2081	0.4007	0.3398	0.1162	0.3058
$T^{n_3,-} + S^{n_2}$	0.1637	0.3211	0.4087	0.3346	0.1202	0.2478
$T^{n_3,-} + S^{n_3}$	0.1702	0.1401	0.4113	0.3331	0.1198	0.3445
$T^{n_3,+} + S^{n_2}$	0.2189	0.2094	0.2525	0.3862	0.3863	0.1373
$T^{n_3,-} + T^{n_3,+}$	0.1672	0.2083	0.2245	0.3509	0.1118	0.2516
$T^{n_2,-} + S^{n_2}$	0.2536	0.2953	0.2329	0.3980	0.2757	0.1617
$T^{n_3,+} + T^{n_3,-} + S^{n_2}$	0.1611	0.1978	0.2241	0.3462	0.1104	0.1760

Table 5.7: D-M test (p-values), $\mu + 3\sigma$ winsorized data, Source: own analysis

5.1.1 The best performing specifications

Let's focus on the models outperforming the basic HAR. We arrive at seven specifications:

1. Daily negative tweets + Weekly search volume
2. Weekly search volume
3. Daily search volume
4. Daily negative tweets
5. Daily negative tweets + Daily search volume
6. Daily negative tweets + Daily positive tweets + Weekly search volume
7. Daily negative tweets + Monthly search volume

The reason why we include also the last specification *Daily negative tweets + Monthly search volume* is twofold. Firstly, it outperforms the basic HAR in terms of the average error term. Secondly, the total error rate differs only negligibly and thus we cannot infer any ultimate conclusions out of it.

MASE value	HAR basic	$T^{n_3,-} + S^{n_2}$	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-} + S^{n_3}$	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	$T^{n_3,-} + S^{n_1}$
Average	0.9993	0.9922	0.9944	0.9962	0.9968	0.9992	1.0027	1.0013
Median	0.9815	0.9941	0.9763	0.9659	0.9572	0.9773	0.9959	0.9718
Max	1.4397	1.4455	1.4399	1.4416	1.4451	1.4482	1.4603	1.4613
Min	0.5438	0.4941	0.4791	0.4850	0.5407	0.5119	0.5343	0.5403
Stdev	0.2124	0.2166	0.2265	0.2208	0.2085	0.2144	0.2268	0.2148
$Pr[X < 1]$	56.67%	50.00%	53.33%	56.67%	53.33%	53.33%	50.00%	56.67%

Table 5.8: MASE values statistics, $\mu + 3\sigma$ winsorized data, Source: own analysis

Table 5.8 summarizes MASE values of top performing models. The complete results, i.e. MASE, RMSE and MAE values of the top performing models, can be found in the Appendix A. The last two models (Daily negative tweets + Daily positive tweets + Weekly search volume, and Daily negative tweets + Monthly search volume) show comparatively worse results than rest of the models. In terms of absolute values, the most sophisticate model combining Daily negative tweets, Daily positive tweets, and Weekly search volume is among top 3 performing models for 15 stocks (see the Appendix A). However, it does not show robust results across all examined stocks. If it does not fit the stock well, the results deteriorate significantly. Thus, it is outperformed by more parsimonious models.

Differences in other MASE results are subtle and almost contradictory. For instance, Daily negative tweets + Weekly search volume specification has the lowest average MASE but almost the highest median MASE. Moreover, the probability that MASE will be equal or lower than 1 is only 50% compared 56,67% for the basic HAR. On average, the Weekly search volume specification and the Daily search volume specification bring the most robust results and outperform HAR almost in every aspect described in the table. Diving deeper into MASE statistical distributions, good performance of daily and weekly search volume is apparent in Figure 5.2.

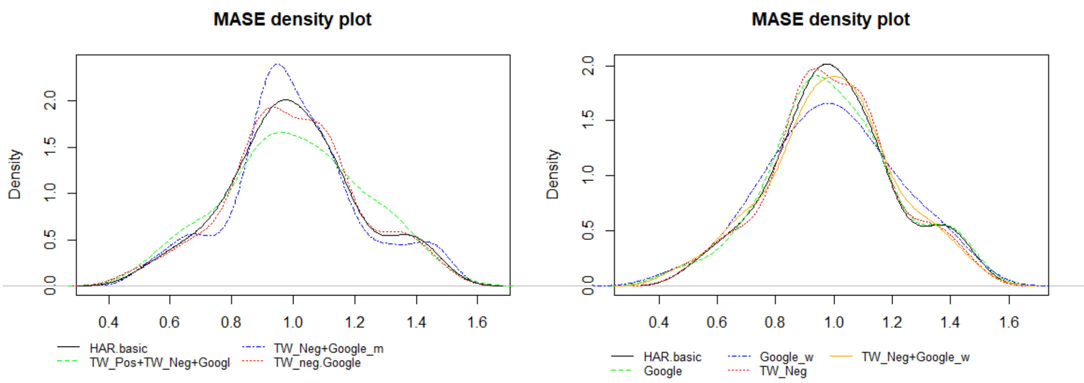


Figure 5.2: Distributions of MASE, Source: own analysis

Generally, selected models exploit overall activity expressed by Google searches and arrival of negative news (innovations) represented by negative tweets. Whereas Google search volume indicates a magnitude of the innovation rather than its sign, negative tweets represent a sign of the news as well as its magnitude. With respect to these relations, the best performing model combines the effect of Daily negative tweets and support them by a mid-term activity measure, i.e. Weekly search volume. It leads the table regardless of measure used. Also, both variables increase accuracy of the basic HAR individually. Daily search volume is another important variable suggesting that searching volume is a suitable proxy of market activity.

These findings are in line with recent literature. Many studies refer to an asymmetric reaction to positive and negative news. For instance, Laakkonen and Lanne (2008) studied the impact of positive and negative macroeconomic news in different phases of the business cycle on the high-frequency volatility of the EUR/USD exchange rate. They conclude that bad news increases volatility more than good news and the news effect depends on the state of the economy. Bad news increases volatility more in good times than in bad times, while there is no difference between the volatility effects of good news in bad and good times. Goudarzi (2011) found a similar pattern in the Indian stock markets, i.e. that the negative news has a greater impact on volatility than a positive news. More recent literature is presented in the 3.

Coherence of negative tweets effect and the state of art of volatility behaviour is certainly a positive sign. But how can we explain short-term nature of the significant variables? All influential factors are daily or weekly. Monthly search volume shows satisfactory results only together with a daily variable.

Figure 5.3 shows Google searches associated with four randomly chosen companies. A dotted green line denotes daily search volume, a solid red line shows weekly search volume and a blue dashed line shows monthly search volume. An obvious consequence of averaging daily values into weekly and monthly clusters is smoothening the volatility. However, a method of moving average is accompanied by another detrimental effect: shifting the ups and downs in the data. This trend is visible in all charts. In particular, the more is the spike "sharp", i.e. it lasts a short period, the more is the shift visible. This may be a source of a low predicative value of the monthly variables.

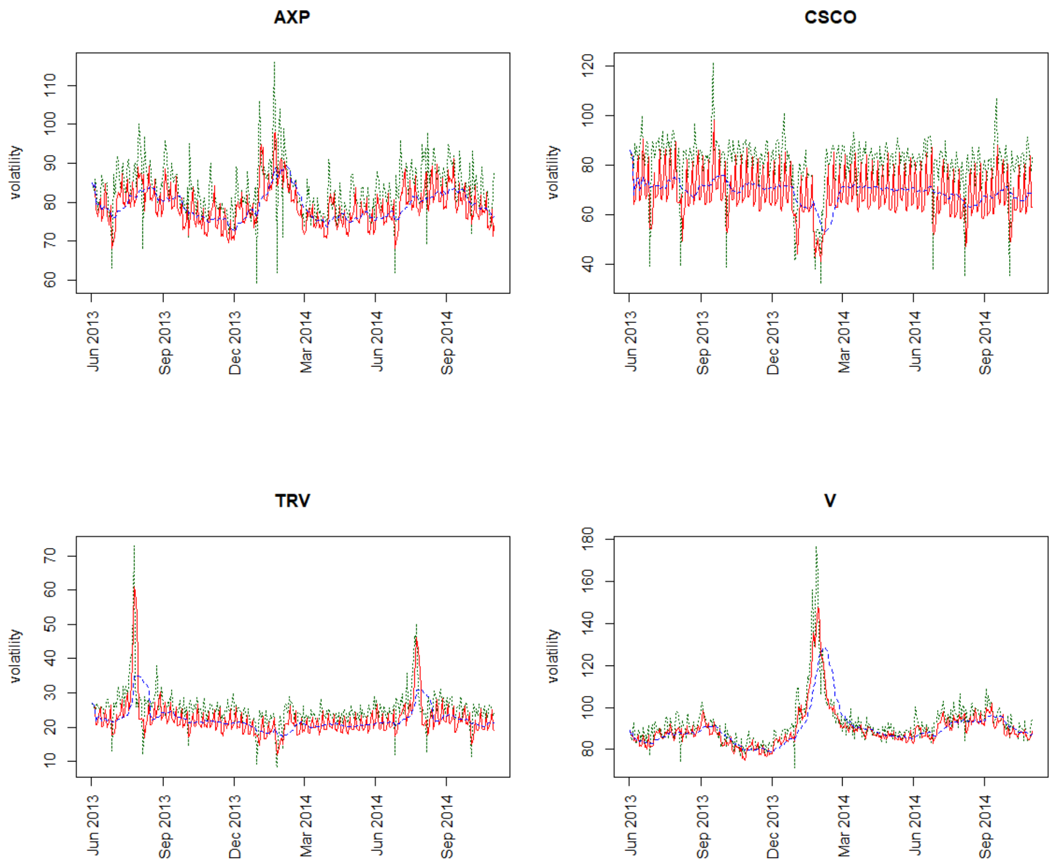


Figure 5.3: Google searches, Source: own analysis

5.2 Individual stocks analysis

The drawback of our approach is that it aggregates results across examined companies. As discussed in the chapter 3, market sentiment affects all stocks to some extend but it affects some more than others. Since we have analysed companies separately, we can observe link between market sentiment and volatility at a company level.

The best stocks	**RMSE**	**MAE**	**MASE**
JNJ	0.00051	0.00041	1.10808
MMM	0.00045	0.00034	0.65880
XOM	0.00044	0.00036	0.77927
MSFT	0.00070	0.00059	0.54380
DIS	0.00059	0.00046	0.67121
Average	**0.00078**	**0.00058**	**0.99931**

Table 5.9: Original HAR model: the best stocks, Source: own analysis

The worst stocks	**RMSE**	**MAE**	**MASE**
CSCO	0.00132	0.00098	0.96921
INTC	0.00113	0.00093	1.13783
KO	0.00096	0.00069	1.43971
T	0.00085	0.00057	1.36575
WMT	0.00087	0.00065	1.39228
Average	**0.00078**	**0.00058**	**0.99931**

Table 5.10: Original HAR model: the worst stocks, Source: own analysis

Firstly, the original HAR model performs best for JNJ, MMM and XOM in terms of MAE and RMSE[1]. In terms of MASE, DIS, MMM and MSFT show the best results. Interestingly, the basic HAR model is outperformed by naïve forecast for JNJ (see the table 5.9). On the other hand, the basic HAR model reveals the worst forecast accuracy for CSCO, INTC and KO according to MAE and RMSE and KO, T and WMT according to MASE (see the table 5.10). Discrepancy between MAE, RMSE and MASE performance indicates different nature of underlying volatility processes. MASE is scaled

[1]For company names, see the table Data overview in the chapter 4.

by scaling statistics Q calculated from in-sample naïve forecasts. The naïve forecast assumes the value t to be equal to the value $t-1$. In terms of the HAR model, the daily volatility t equals to the daily volatility $t-1$, i.e. the first part of the volatility cascade. The results suggest different contribution of individual elements of the volatility cascade for individual companies. For instance, Cisco (CSCO) shows poor values of RMSE and MAE but a superb value of MASE. It indicates importance of weekly and monthly elements of the volatility cascade.

No wonder, extended models perform similarly to the basic HAR model. But interestingly, sentiment variables contribution differs across companies and specifications. MSFT, XOM and CSCO are among models with highest positive impact of market sentiment variables. Table 5.11 presents improvement of forecast errors of extended models. Negative values indicate enhanced performance of the extended models over the basic HAR model. Besides those three stocks, few stocks show substantially lower forecast errors for particular specifications. For instance, weekly Google searches improve MAE and MASE of HD by 15% and negative tweets improve MAE and MASE of IBM by 11%. The complete results can be found in the Appendix A.

Stock	Error	$T^{n_3,-}$ + S^{n_2}	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-}$ + S^{n_3}	$T^{n_3,-}$ + $T^{n_3,+}$ + S^{n_2}	$T^{n_3,-}$ + S^{n_1}
MSFT	**RMSE**	-9%	-9%	-5%	-3%	-6%	1%	-3%
	MAE	-9%	-12%	-11%	-1%	-6%	-2%	-1%
	MASE	-9%	-12%	-11%	-1%	-6%	-2%	-1%
CSCO	**RMSE**	-9%	-9%	-4%	-2%	-4%	-7%	-3%
	MAE	-21%	-19%	-9%	-4%	-8%	-22%	-6%
	MASE	-21%	-19%	-9%	-4%	-8%	-22%	-6%
XOM	**RMSE**	-9%	-8%	-2%	-1%	-2%	-10%	-5%
	MAE	-10%	-9%	-2%	-1%	-3%	-12%	-5%
	MASE	-10%	-9%	-2%	-1%	-3%	-12%	-5%

Table 5.11: The difference between errors of the extended model and those of the original model: the best stocks, Source: own analysis

Table 5.12 outlines the worst performing stocks, namely PFE, CAT and DD. The higher values suggest worse performance compare to the basic HAR model. Extending the basic HAR model by any Google search variable significantly hinders performance of the model in the case of CAT. For DD and PFE, tweets impair forecast accuracy of the model more. HD and IBM show an interesting mirror effect. Whilst, weekly Google searches improve MAE and MASE of HD by 15% they deteriorate the same errors of IBM by 14%. Also, MAE and MASE of HD are negatively influenced by negative tweets, i.e. higher by 11% (see table 5.13). Apart from these stocks, the most sophisticated model combining negative and positive tweets, and weekly Google search volume deteriorates measurement error of MRK substantially. RMSE is higher by 54% and MAE and MASE are higher by 25%. Those extreme values impair average performance of the model.

Stock	**Error**	$T^{n_3,-}$ + S^{n_2}	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-}$ + S^{n_3}	$T^{n_3,-}$ + $T^{n_3,+}$ + S^{n_2}	$T^{n_3,-}$ + S^{n_1}
PFE	**RMSE**	10%	0%	1%	12%	14%	24%	9%
	MAE	9%	0%	1%	10%	12%	19%	6%
	MASE	9%	0%	1%	10%	12%	19%	6%
CAT	**RMSE**	7%	8%	2%	0%	0%	13%	2%
	MAE	12%	16%	2%	-2%	-2%	17%	6%
	MASE	12%	16%	2%	-2%	-2%	17%	6%
DD	**RMSE**	7%	-1%	-4%	6%	0%	3%	11%
	MAE	13%	3%	4%	7%	9%	9%	18%
	MASE	13%	3%	4%	7%	9%	9%	18%

Table 5.12: The difference between errors of the extended model and those of the original model: the worst stocks, Source: own analysis

Stock	**Error**	$T^{n_3,-}$ + S^{n_2}	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-}$ + S^{n_3}	$T^{n_3,-}$ + $T^{n_3,+}$ + S^{n_2}	$T^{n_3,-}$ + S^{n_1}
HD	**RMSE**	0%	-4%	-1%	7%	7%	0%	7%
	MAE	-7%	-15%	-4%	11%	10%	-6%	11%
	MASE	-7%	-15%	-4%	11%	10%	-6%	11%
IBM	**RMSE**	1%	8%	3%	-6%	-3%	-2%	1%
	MAE	2%	14%	4%	-11%	-7%	-2%	4%
	MASE	2%	14%	4%	-11%	-7%	-2%	4%

Table 5.13: The difference between errors of the extended model and those of the original model: HD and IBM, Source: own analysis

Table 5.14 presents a number of stocks where the extended model outperforms the basic HAR model. In terms of RMSE, performance of the extended model and the basic model is comparable. Four out of seven models show equivalent performance for the extended and the basic model. MAE and MASE results are identical. The best performing models contain $T^{n_3,-}$ and S^{n_2}, $T^{n_3,-}$, $T^{n_3,-}$ and S^{n_3}, $T^{n_3,-}$ and $T^{n_3,+}$ and S^{n_2}. The last specification performs very poorly for few stocks. Considering RMSE errors, we conclude that $T^{n_3,-}$ and S^{n_2}, $T^{n_3,-}$, and $T^{n_3,-}$ and S^{n_3} are the best performing models across examined stocks.

Better model	Error	$T^{n_3,-}$ + S^{n_2}	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-}$ + S^{n_3}	$T^{n_3,-}$ + $T^{n_3,+}$ + S^{n_2}	$T^{n_3,-}$ + S^{n_1}
Extended	RMSE	17	15	11	17	15	15	14
Basic HAR	RMSE	13	15	19	13	15	15	16
Extended	MAE	16	14	13	19	18	17	15
Basic HAR	MAE	14	16	17	11	12	13	15
Extended	MASE	16	14	13	19	18	17	15
Basic HAR	MASE	14	16	17	11	12	13	15

Table 5.14: Comparison of extended model performance and the basic HAR model performance: number od stocks , Source: own analysis

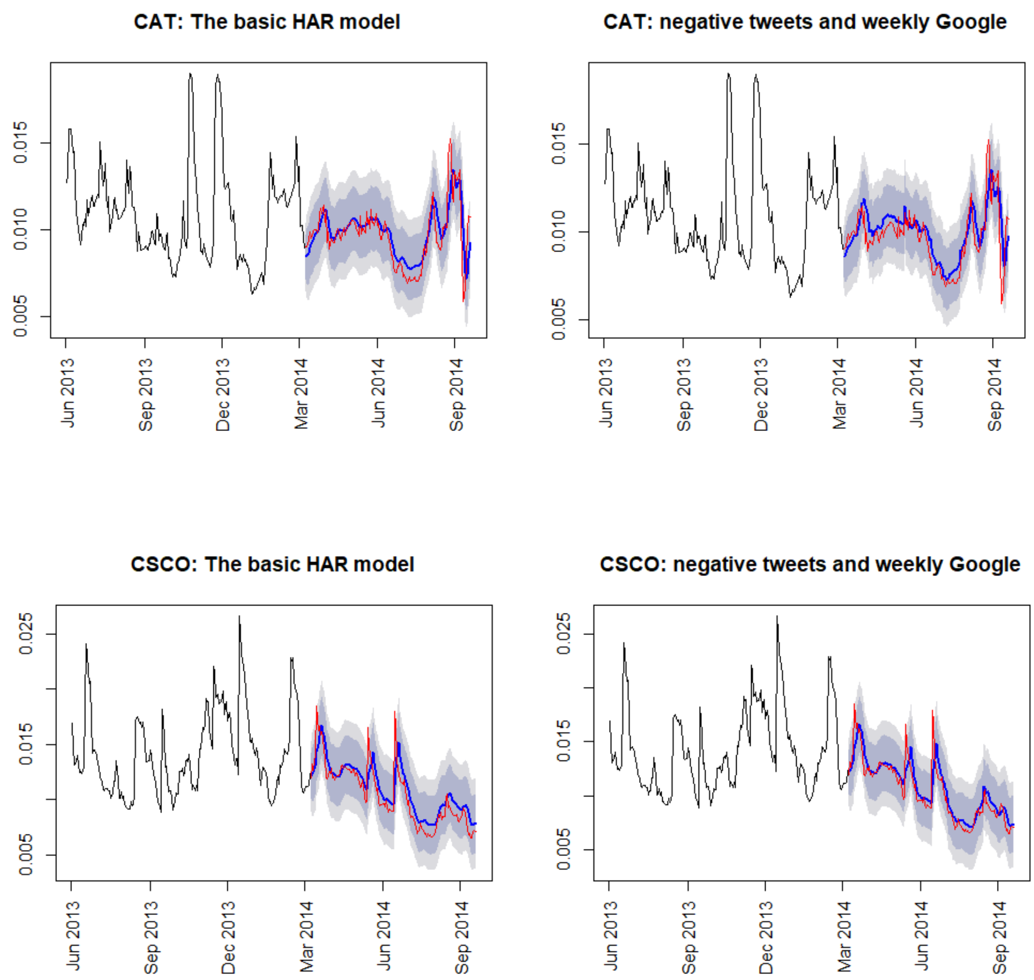


Figure 5.4: Out-of-sample forecast plots, CAT and CSCO, Source: own analysis

Figure 5.4 depicts out-of-sample forecasts of the basic HAR and the specification combining daily negative tweets and weekly search volume. As an example, we present CSCO and CAT. The difference between those two models is apparent in the chart. Whereas in the CSCO plot, a line of the extended model copies true volatility values more accurately than that of the basic HAR model, in the CAT plot, the HAR model plot is the more precise. Generally, all depicted models show very solid performance as a low level of measurement errors suggests.

5.3 Robustness analysis

Winsorizing significantly affects performance of our models. As a final spike-adjusted time series, we use data capped by $\mu + 3\sigma$ value. Figure 5.5 shows the density functions of adjusted variables of the six critical companies. A solid black line is assigned to positive tweets and a dashed red line denotes negative tweets. Vertical lines show the winsorizing limit equal to $\mu + 3\sigma$.

We do not include density functions of neutral tweets, although we constrained them, as they do not have a positive effect on model's performance. All densities are leptokurtic and positively skewed. The minimal values are 0, as an underlying variable is a number of daily tweets. Disbalance between positive and negative tweets is apparent in the chart 5.5. The table *Data overview* in the chapter 4 shows that the absolute number of positive tweets is almost twice larger than the number of negative tweets. However, the density plots show that positive tweets have mean similar to that of negative tweets but differs in variance.

It brings us to the interesting point regarding our dataset. How were the tweets collected and evaluated? Why are there more positive tweets than negative tweets? And, is a magnitude of a positive tweet comparable to that of a negative tweet? These questions are discussed in the last chapter.

The $\mu + 3\sigma$ limit is a very conservative level considering a field of social science where 2σ limits are commonly used. Thus, we performed a robustness analysis in order to support reliability of our conclusions. We additionally constrained the data by $\mu + 2.5\sigma$, $\mu + 2\sigma$, and $\mu + 1.5\sigma$ values. We arrived at five different rankings including the original dataset.

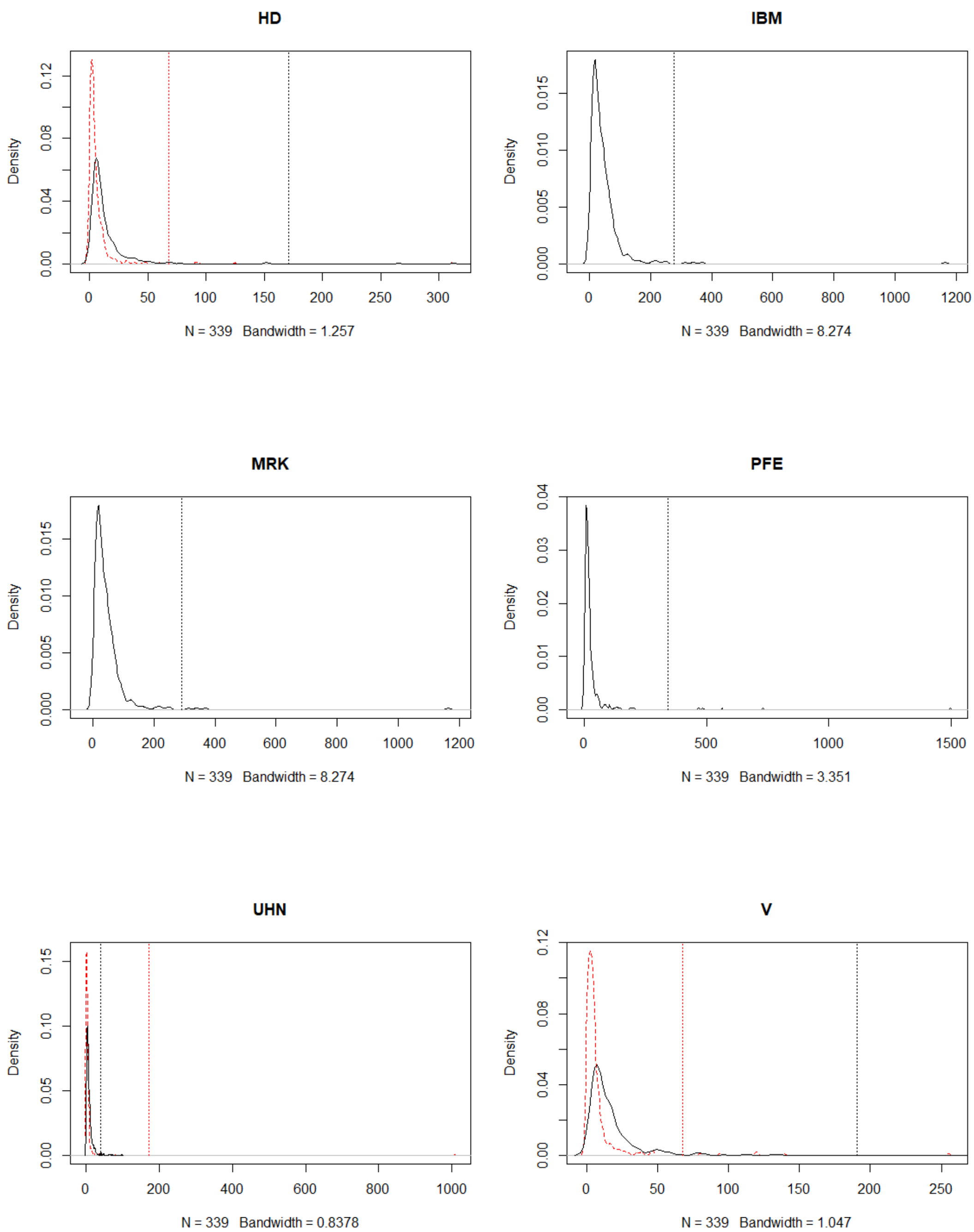


Figure 5.5: Density functions and limiting vaulues, Source: own analysis

The results of robustness analysis are satisfactory. MASE was used as a measure for the ranking presented in Table 5.15. Exact values of MASE coefficients can be found in the Appendix A. The original data and the $\mu + 1.5\sigma$ data are extreme cases where either outliers are not limited at all or are limited very insolently. Nevertheless, even those rankings contain almost the same specification. The exception are $T^{n_3,-} + T^{n_3,+} + S^{n_1}$ that occurs only in strictly winsorized datasets, i.e. $\mu + 2.5\sigma$ and less, $T^{n_3,+}$ that occurs only in the original data, and S^{n_1} that occurs in the original data and the $\mu + 3\sigma$ data. However, all of these specifications occupy the bottom places of an ranking ladder and none outperforms the basic HAR.

Rank	**Original data**	**3sigma**	**2.5 sigma**	**2 sigma**	**1.5 sigma**
1	S^{n_2}	$T^{n_3,-} + S^{n_2}$	$T^{n_3,-} + S^{n_2}$	$T^{n_3,-} + S^{n_2}$	$T^{n_3,-} + S^{n_2}$
2	S^{n_3}	S^{n_2}	S^{n_2}	S^{n_2}	S^{n_2}
3	HAR basic	S^{n_3}	S^{n_3}	S^{n_3}	S^{n_3}
4	$T^{n_3,-} + S^{n_2}$	$T^{n_3,-}$	$T^{n_3,-}$	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$
5	S^{n_1}	$T^{n_3,-} + S^{n_3}$	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	$T^{n_3,-}$	$T^{n_3,-}$
6	$T^{n_3,-}$	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	$T^{n_3,-} + S^{n_3}$	$T^{n_3,-} + S^{n_3}$	$T^{n_3,-} + S^{n_3}$
7	$T^{n_3,-} + S^{n_3}$	HAR basic	HAR basic	HAR basic	$T^{n_3,-} + S^{n_1}$
8	$T^{n_3,-} + S^{n_1}$	$T^{n_3,-} + S^{n_1}$	$T^{n_3,-} + S^{n_1}$	$T^{n_3,-} + S^{n_1}$	$T^{n_3,+} + S^{n_2}$
9	$T^{n_3,+} + S^{n_2}$	$T^{n_3,+} + S^{n_2}$	$T^{n_3,+} + S^{n_2}$	$T^{n_3,+} + S^{n_2}$	HAR basic
10	$T^{n_2,-} + S^{n_2}$	S^{n_1}	$T^{n_3,-} + T^{n_3,+}$	$T^{n_3,-} + T^{n_3,+}$	$T^{n_3,-} + T^{n_3,+}$
11	$T^{n_3,+}$	$T^{n_3,-} + T^{n_3,+}$	$T^{n_3,-} + T^{n_3,+} + S^{n_1}$	$T^{n_3,-} + T^{n_3,+} + S^{n_1}$	$T^{n_3,-} + T^{n_3,+} + S^{n_1}$

Table 5.15: Ranikng for various winsorization levels, Source: own analysis

By closer analysis of the middle datasets, i.e. from $\mu + 3\sigma$ to $\mu + 2\sigma$, we may confirm that the ranking is stable and thus, relatively robust to a level of winsorization. Moreover, we can observe a trend in improvement forecast accuracy of more complex models with increasing winsorization. As extreme values are more dramatically capped, the more complex models are shifting upward at the expense of the more parsimonious models. Interestingly, the upper 3 ranks remain unchanged for all levels of winsorization.

5.4 Results summary

Out-of-sample approach has shown some minor improvements in forecast accuracy. Notably, the MASE based error rate and the MAE based error rate have improved for some extended models. The RMSE based error rate has decreased only slightly due to lower robustness to irrelevant noise induced by market sentiment proxies. However, Diebold-Mariano test has shown that the difference in forecast accuracy is not statistically significant. A ranking of the top ten models according to predictive power contains almost the same specifications for the original data and the $\mu + 3\sigma$ winsorized data. Although, the basic HAR occupies higher ranks for the original data due to the detrimental effect of Twitter spikes. Thus, we decided to elaborate on the $\mu + 3\sigma$ winsorized data.

The models outperforming the basic HAR contain particularly Daily negative tweets, Weekly search volume, and Daily search volume. Daily positive tweets have satisfactory results only in combination with Daily negative tweets and Weekly search volume. Robustness analysis has shown that the ranking is robust to a level of winsorization.

In terms of individual stocks, all models show decent results. The basic HAR model fit some stocks better and thus subsequently, all extended models attain higher accuracy. Considerable discrepancy between MAE, RMSE and MASE performance suggests different nature of underlying volatility processes. For instance, strong MASE performance and poor RMSE and MAE performance indicate importance of weekly and monthly elements in the volatility cascade. Contribution of sentiment variables differs across companies and specifications. Whereas, some specifications for particular stocks improve forecast accuracy up to 22%, others impair forecast accuracy up to 24% (in one extreme case up to 54%). By a thorough analysis of a number of stocks where the extended model outperforms the basic HAR model, we arrived at following results. The best performing models contain $T^{n_3,-}$ and S^{n_2}, $T^{n_3,-}$, and $T^{n_3,-}$ and S^{n_3}. Moreover, by analysing relation between market sentiment and volatility at a company level, we can efficiently identify stocks which volatility is strongly correlated with market sentiment. Although, we cannot decide on which specification is the ultimately best performing models for all stocks, we can say which stocks are interesting for market sentiment analysis.

Chapter 6

Conclusion

The aim of this thesis is to improve understanding of information hidden in social networks and internet data. Particularly, can it serve as a market sentiment indicator and improve predictive power of volatility models? The answer is yes.

As the basic model, we had chosen the heterogenous autoregressive model (HAR) and extended it by a variety of market sentiment factors. We have approximated market sentiment by Google search volume and a quantity of positive, negative and neutral tweets. The Twitter variables contain extreme values and thus, we winsorized them using $\mu + 3\sigma$ limit. To support reliability of our results, we performed robustness analysis. We have used two different methodologies to evaluate the effect of market sentiment. Firstly, we have performed out-of-sample forecast, compared measurement errors, tested statistical significance of the difference in forecast accuracy, and compiled a ranking of the extended models based on their relative performance. Secondly, we have carried out an analysis on the top seven performing models at a stock level. The final ranking indicates models that outperform the basic HAR model.

We have arrived at seven specifications that slightly outperform the basic HAR. Those models include the basic HAR enriched by daily negative tweets and weekly search volume, weekly search volume, daily search volume, daily negative tweets, daily negative tweets and daily search volume, daily negative tweets and daily positive tweets and weekly search volume, and daily negative tweets and monthly search volume, respectively. These extended models have shown better performance, particularly in terms of MASE and MAE. The RMSE based error rate has decreased only slightly due to lower robustness to irrelevant noise induced by market sentiment proxies. Robustness analysis has shown that the ranking is robust to a level of winsorization. However, as the

Diebold-Mariano test has shown, the differences in forecast accuracy are not statistically significant.

The individual stocks analysis has brought affirmative results. By an analysis of a number of stocks where the extended model outperforms the basic HAR model, we have shown that the best performing models contain $T^{n_3,-}$ and S^{n_2}, $T^{n_3,-}$, and $T^{n_3,-}$ and S^{n_3}. The first model combining daily negative tweets and weekly search volume is a superior model for 17 stocks according to RMSE and 16 stocks according to MAE and MASE. The daily negative tweets outperforms the basic HAR for 17 and 19 stocks, respectively. And, the last specification containing daily negative tweets and daily search volume outpaces the basic HAR for 15 and 18 stocks, respectively. Based on the average MASE and MAE improvement, a combination of daily negative tweets and weekly search volume is a clear winner. It lowers the average MAE and the average MASE by 0.98% and 0.71% respectively. Based on the average RMSE, the daily search volume model leads the table with 0.05% improvement.

Moreover, three important findings have been revealed. Firstly, the basic HAR model has very good predictive power even without additional variables. This is not a surprising phenomenon. For instance, Hansen and Lunde (2005) compared 330 ARCH-type models in terms of their ability to describe the conditional variance. They did not find any evidence that a GARCH(1,1) is outperformed by more sophisticated models apart from those accommodating a leverage effect.

Secondly, assessment of the ultimate best model is difficult. A model that captures well volatility of one stock is outperformed by other models for the other stocks. Moreover, the difference between individual HAR-type models is so subtle that within a certain range, the used error methodology is unable to distinguish superior and inferior models. Therefore, we present a set of the best performing models.

Thirdly, market sentiment variables contain a substantial amount of information. We identified three relevant variables, daily negative tweets, daily Google search volume and weekly Google search volume. These variables improve forecast accuracy of the basic HAR model separately or in a Twitter-Google combination. The absence of positive tweets suggests that negative news influence volatility more than positive news. In terms of more sophisticated models, a set of daily negative tweets, daily positive tweets and weekly Google search volume outperforms the basic HAR. This model is among top 3 performing models for 15 stocks and outperforms the basic model for 17 stocks but, it does not show robust results across all examined stocks. The main drawback of this model is that if it does not fit the stock well, the results deteriorate significantly. Thus, it is outperformed by more parsimonious models.

All relevant market sentiment variables are in a weekly or daily form. The absence of monthly values is arguably caused by a shift in the data. The averaging of monthly values introduces long-memory and thus shift extreme values forward. Interestingly, the weekly search volume outperforms the daily search volume particularly, in a combination with negative tweets. There are multiple ways how to explain it. Either, it suggests that weekly search volume better reflects market activity. Or, the short-term market activity is embodied in a magnitude of tweets, i.e. the number of tweets, and thus, daily Google search volume provide partially duplicate information. Or, different types of investors, with respect to various investment horizons, use different information channels. For instance, increased daily market activity does not have such a significant impact on search volume since investors may use Bloomberg services instead of googling the company. Alternatively, investors betting on short-term market activity might have a better idea of what they are looking for and thus do not increase google activity as significantly as short-term investors.

A combination of market sentiment and a HAR model is particularly interesting due to the heterogeneous component contained in both concepts. The HAR model implicitly assumes heterogeneity in volatility generating process. Market sentiment encompasses cognitive and emotional biases of heterogenous investors and a variety of fundamental and technical factors. Those components might be partially overlapping. Thus, extending a heterogenous model by market sentiment indicators, that are by definition formed by heterogenous elements, may introduce additional noise that is not compensated by additional information.

6.1 Theoretical contribution

From a theoretical perspective, the research brings novel findings about information contained in publicly available data. It has shown that Google search volume that can be easily obtained and processed in real time contains a sufficient amount of information to improve predictive power of an econometrics model. We have also pointed out the challenges associated with Twitter data. Evaluation and classification of tweets is a technically and intellectually difficult task. Twitter posts contain a lot of unrelated noise that can, if untreated, impair their forecasting power.

Moreover, we have confirmed strong performance of HAR-type models even without use of high-frequency data. The basic HAR model as well as other selected models have shown excellent results. Mean absolute scaled errors take values significantly lower than 1 for most of the stocks. Even average MASE including all examined stocks is below 1. Those are results that are commonly arduous to achieve.

6.2 Limitation

One of the limitations is uncertain origin of twitter data. We have to believe in methodology of Ranco et al. (2015) as we do not know how were the tweets collected and evaluated. Disbalance between positive and negative tweets discussed in the chapter 5 bring us to an interesting point. Why are there more positive tweets than negative tweets? Is a magnitude of a positive tweet comparable to that of a negative tweet? Or, is more unrelated noise present in positive tweets than in negative tweets? Since we cannot answer these questions, we cannot make any conclusions on worse performance of positive tweets than negative tweets. It is very likely, and in line with literature, that arrival of negative news influence volatility more than arrival of positive news but it is also possible that this effect is caused by inappropriate selection of tweets.

The next limitation is also associated with a system of tweets evaluation. A positive-neutral-negative scale enables sort tweets in three disjointed sets but does not allow for cardinality. In other words, are the more and less negative tweets in the same set? We might assume that a number of tweets, i.e. a magnitude of news, might foster the relevance of the post. The logic behind that is that a relevant post should be re-tweeted by peers and therefore increased in a magnitude. However, reasons for re-tweeting might vary from economical relevance to personal affection to a writer or a subject. Such tweets may create clusters of economically unrelated noise and reduce performance of the models.

Another great limitation is the used methodology. We made many decisions about model selection, length of a sample in out-of-sample forecast, error rate ranking, or a stock level analysis. Each of these decisions might influence results and thus, researchers using different methodology might arrive at slightly different results. However, we tried to describe and follow used methodology as precisely as possible in order to increase reliability and replicability of our results.

6.3 Future research

The topic of extending models by market sentiment is still relatively unexplored. Channels such as social networks hide untapped potential of novel data streams. Hand in hand, latest technologies in big data analysis or artificial intelligence offer more efficient way how to process those data. This combination represents unique opportunity in unravelling a new class of financial models.

In this research, we have shown that market sentiment variables contain a substantial amount of information. However, we did not arrive at the ultimate best model. It can be achieved by another methodology of model selection or, for instance, averaging models. Averaging models is widely used technique in forecasting and thus, extending this research by weighted average of the selected models is a logical step.

Furthermore, we described substantial discrepancy among MAE, RMSE and MASE performance, and distinct contribution of sentiment variables across companies and specifications. New questions have arisen. Does volatility of individual stocks follow slightly different processes? For instance, various elements of the volatility cascade are of various importance for diverse stocks. And, which stocks tend to be more sensitive to market sentiment? Those and many other questions should be properly answered.

Bibliography

Alizadeh, S., Brandt, W. M. and Diebold, X. F. (2002), 'Range-based estimationof stochastic volatility models', *Journal of Finance* **57**, 1047–1091.

Antweiler, W. and Frank, M. Z. (2004), 'Is all that talk just noise? the information content of internet stock message boards', *Journal of Finance* **59**, 1259–1294.

Baillie, R. T. and Bollerslev, T. (1990), 'Intra day and inter market volatility in foreign exchange rates', *Review of Economic Studies* **58**, 565–585.

Baker, H. K. and Ricciardi, V. (2014), 'How biases affect investor behaviour', *The European Financial Review* .
URL: *http://www.europeanfinancialreview.com/?p=512*

Baker, M. and Wurgler, J. (2007), 'Investor sentiment in the stock market', *The Journal of Economic Perspectives* **21**(32), 129–151.

Bekaert, G. and Wu, G. (2000), 'Asymmetric volatility and risk in equity markets', *The review of financial studies* **13**(1), 1–42.

Benartzi, S. and Thaler, R. H. (2001), 'Naive diversification strategies in defined contribution saving plans', *American economic review* pp. 79–98.

Berument, H. and Kiymaz, H. (2001), 'The day of the week effect on stock market volatility', *Journal of Economics and Finance* **25**(2), 181–193.

Bollen, J., Mao, H. and Zeng, X. (2011), 'Twitter mood predicts the stock market', *Journal of computational science* **2**(1), 1–8.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Economics* **31**(3), 307–327.

Brandt, M. W. and Kinlay, J. (2005), 'Estimating historical volatility', *Research Article, Investment Analytics* .

Chai, T. and Draxler, R. R. (2014), ‘Root mean square error (rmse) or mean absolute error (mae)? â€“ arguments against avoiding rmse in the literature’, *Geoscientific Model Development* **7**(3), 1247–1250.

Chen, C. W., Chiang, T. C. and So, M. K. (2003), ‘Asymmetrical reaction to us stock-return news: evidence from major stock markets based on a double-threshold model’, *Journal of Economics and Business* **55**(5-6), 487–502.

Chen, S. W. and Shen, C. H. (2004), ‘Garch, jumps and permanent and transitory components of volatility: the case of the taiwan exchange rate’, *Mathematics and Computers in Simulation* **67**(3), 201–216.

Chordia, T., Roll, R. and Subrahmanyam, A. (2001), ‘Market liquidity and trading activity’, *Journal of Finance* **56**(2), 501â€“530.

Cont, R. (2001), ‘Empirical properties of asset returns: stylized facts and statistical issues’, *Quantitative finance* **1**, 223–236.

Corsi, F. (2009), ‘A simple approximate long memory model of realized volatility’, *Journal of Financial Econometrics* **7**, 174–196.

Corsi, F. and Reno, R. (2009), ‘Har volatility modelling with heterogeneous leverage and jumps’.
URL: *Available at SSRN 1316953*

De Bondt, W. F., Muradoglu, Y., Shefrin, H. and Staikouras, S. K. (2008), ‘Behavioral finance: Quo vadis?’, *Journal of Applied Finance (Formerly Financial Practice and Education)* **18**(2).

DeLong, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990), ‘Noise trader risk in financial markets’, *Journal of Political Economy* **98**(4), 703â€“738.

Diebold, F. (2015), ‘Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of dieboldâ€“mariano tests’, *Journal of Business and Economic Statistics* **33**(1), 1–1.

Diebold, F. and Mariano, R. (1995), ‘Comparing predictive accuracy’, *Journal of Business and Economic Statistics* **13**, 253–263.

Dixon, W. J. (1960), ‘Simplified estimation from censored normal samples’, *The Annals of Mathematical Statistics* **31**, 385â€“391.

Engle, R. F. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation', *Econometrica* **50**, 987–1008.

Engle, R. F., Ito, T. and Lin, W. (1990), 'Meteor showers or heat wave? heteroskedastic intra-daily volatility in the foreign exchange market', *Econometrica* **58**(3), 525–542.

Fama, E. F. (1970), 'Efficient capital markets: A review of theory and empirical work', *The Journal of Finance* **25**, 383–417.

Garman, M. and Klass, M. (1980), 'On the estimation of security price volatilities from historical data', *The Journal of Business* **53**, 67–78.

Giacomini, R. and White, H. (2003), 'Tests of conditional predictive ability', *Working Paper No. 572* .

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The Journal of Finance* **48**, 1779–1801.
URL: *doi:10.1111/j.1540-6261.1993.tb05128.x*

Goudarzi, H. (2011), 'Modeling asymmetric volatility in the indian stock market', *International Journal of Business and Management* **6**(3).

Hansen, P. R. and Lunde, A. (2005), 'A forecast comparison of volatility models: does anything beat a garch(1,1)?', *Journal of Applied Econometrics* **20**(7), 873–889.
URL: *http://dx.doi.org/10.1002/jae.800*

Hausman, J. (1978), 'Specification tests in econometrics', *Econometrica* **46**(6), 1251–1271.

Honda, Y. (1985), 'Testing the error components model with non–normal disturbances', *Review of Economic Studies* **52**, 681–690.

Hong, H. and Stein, J. (2003), 'Differences of opinion, short-sales constraints and market crashes', *Review of Financial Studies* **16**(2), 487â€“525.

Hyndman, R. and Athanasopoulos, G. (2014), *Forecasting: principles and practice*, OTexts, chapter 2.5: Evaluating forecast accuracy.

Hyndman, R. and Koehler, A. B. (2006), 'Another look at measures of forecast accuracy', *International Journal of Forecasting* **22**(4), 679–688.

Jarrow, J. R. (1988), *Finance Theory*, Prentice-Hall, Englewood Cliffs.

Kahneman, D. and Tversky, A. (1979), 'Prospect theory: An analysis of decision under risk', *Quantitative Finance* **47**(2), 263–291.

Laakkonen, H. and Lanne, M. (2008), 'Asymmetric news effects on volatility: Good vs. bad news in good vs. bad times', *HECER Discussion Paper No. 207* .

Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D. and Allan, J. (2000), 'Mining of concurrent text and time series', *In Proceedings of the 6th KDD Workshop on Text Mining* pp. 37–44.

Magdon-Ismail, M. and Atiya, A. F. (2003), 'A maximum likelihood approach to volatility estimation for a brownian motion using high, low and close price data', *Quantitative Finance* **3**(5), 376–384.

Malmendier, U. and Tate, G. (2008), 'Who makes acquisitions? ceo overconfidence and the market's reaction', *Journal of financial Economics* **89**(1), 20–43.

Muller, U., Dacorogna, M., Dave, R., Olsen, R., Pictet, O. and vonWeizsacker, J. (1997), 'Volatilities of different time resolutions analyzing the dynamics of market components', *Journal of Empirical Finance* **4**, 213–239.

Muller, U., Dacorogna, M., Dave, R., Olsen, R., Pictet, O. and Ward, J. R. (1993), 'Fractals and intrinsic time - a challenge to econometricians', *Research Report UAM* **1993-08-16**.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica* **59**(2), 347–370.
URL: *http://doi.org/10.2307/2938260*

Neumann, J. v. and Morgenstern, O. (1953), *Theory of Games and Economic Behavior*, Princeton, NJ. Princeton University Press.

Parkinson, M. (1980), 'The extreme value method for estimating the variance of the rate of return', *Journal of Business* pp. 61–65.

Pesaran, M. H. (2005), 'Market efficiency today', *IEPR working paper 05.41* .

Peters, E. E. (1991), 'Chaos and order in the capital markets'.

Peters, E. E. (1994), 'Fractal market analysis', *Applying Chaos Theory to Investment and Economics* .

Pietras, M. (2013), 'Best time to send email'.
URL: *http://blog.getresponse.com/new-infographic-best-day-to-send-email-2013.html*

Ranco, G., Aleksovski, D., Caldarelli, G., M., G. and Mozetic, I. (2015), 'The effects of twitter sentiment on stock price returns', *PloS one* **10**(9), e0138441.

Ritter, J. R. (2003), 'Behavioral finance', *Pacific-Basin Finance Journal* **11**(4), 429–437.

Rodriguez, M. J. and Ruiz, E. (2012), 'Revisiting several popular garch models with leverage effect: Differences and similarities', *Journal of Financial Econometrics* **10**(4), 637–668.

Rogers, L. C. G. and Satchell, S. E. (1991), 'Estimating variance from high, low and closing prices', *The Annals of Applied Probability* **1**(4), 504–512.

Scharfstein, D. S. and Stein, J. C. (1990), 'Herd behavior and investment', *The American Economic Review* **80**(3), 465–479.

Schumaker, R. P. and Chen, H. (2009), 'Textual analysis of stock market prediction using breaking financial news', *ACM Transactions on Information Systems* **27**(2), 1–19.

Sewell, M. (2011), 'Characterization of financial time series', *UCL Department of computer science, Research Note* **11**(01).

Shefrin, H. (2005), 'A behavioral approach to asset pricing', *Boston: Elsevier Academic Press* .

Shleifer, A. and Vishny, R. (1997), 'The limits of arbitrage', *The American Economic Review* **52**(1), 35–55.

Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H. and Deng, X. (2013), 'Exploiting topic based twitter sentiment for stock prediction', *The Journal of Global Business Management* **ACL**(2), 24–29.

Singh, S. (2012), 'Investor irrationality and self-defeating behavior: Insights from behavioral finance', *The Journal of Global Business Management* **8**(1).

Smith, G. P. (2012), 'Google internet search activity and volatility prediction in the market for foreign currency', *Finance Research Letters* **9**(2), 103–110.

Sprenger, T., Tumasjan, A., Sandner, P. and Welpe, I. (2014), 'Tweets and trades: The information content of stock microblogs', *European Financial Management* **20**(5), 926â€“957.

Steenbergen, M. (2001), 'The reverend bayes meets john q. public: Patterns of political belief updating in citizens'.

Tesar, L. L. and Werner, I. M. (1995), 'Home bias and high turnover', *Journal of international money and finance* **14**(4), 467–492.

Thaler, R. H. (1999), 'Mental accounting matters', *Journal of Behavioral decision making* **12**(3), 183.

Uygur, U. and Tas, O. (2012), 'Modeling the effects of investor sentiment and conditional volatility in international stock markets', *Journal of Applied Finance and Banking* **2**(5), 239.

Veronesi, P. (1999), 'Stock market overreaction to bad news in good times: a rational expectations equilibrium model', *The review of financial studies. Winter 1999* **12**(5), 975–1007.

Weron, A. and Weron, R. (2000), 'Fractal market hypothesis and two power-laws', *Chaos, Solitons and Fractals* **11**(1), 289–296.

Wu, L. and Brynjolfsson, E. (2015), 'The future of prediction: How google searches foreshadow housing prices and sales', *Economic analysis of the digital economy* pp. 89–118.

Yang, D. and Zhang, Q. (2000), 'Drift independent volatility estimation based on high, low, open and close prices', *Journal of Business* **73**, 477–491.

Zakoian, J. M. (1994), 'Threshold heteroskedastic models', *Journal of Economic Dynamics and control* **18**(5), 931–955.

Zhang, X., Fuehres, H. and Gloor, P. A. (2011), 'Predicting stock market indicators through twitter "i hope it is not as bad as i fear"', *Procedia-Social and Behavioral Sciences* **26**, 55–62.

Appendix A

Appendix

A.1 All specifications

1. the basic HAR model
2. Daily positive tweets
3. Daily negative tweets
4. Daily neutral tweets
5. Daily search volume
6. Weekly positive tweet
7. Weekly negative tweets
8. Weekly search volume
9. Monthly positive tweet
10. Monthly negative tweets
11. Monthly search volume
12. Daily positive tweets + Daily search volume
13. Daily negative tweets + Daily search volume
14. Daily positive tweets + Weekly search volume
15. Daily negative tweets + Weekly search volume
16. Daily positive tweets + Monthly search volume
17. Daily negative tweets + Monthly search volume
18. Weekly positive tweets + Daily search volume
19. Weekly negative tweets + Daily search volume
20. Monthly positive tweets + Daily search volume
21. Monthly negative tweets + Daily search volume
22. Weekly positive tweets + Weekly search volume
23. Weekly negative tweets + Weekly search volume
24. Monthly positive tweets + Weekly search volume
25. Monthly negative tweets + Weekly search volume
26. Weekly positive tweets + Monthly search volume
27. Weekly negative tweets + Monthly search volume
28. Monthly positive tweets + Monthly search volume
29. Monthly negative tweets + Monthly search volume

30. Weekly negative tweets + Daily negative tweets
31. Monthly negative tweets + Daily negative tweets
32. Daily negative tweets + Daily positive tweets
33. Daily negative tweets + Weekly positive tweets
34. Daily negative tweets + Monthly positive tweets
35. Weekly negative tweets + Daily positive tweets
36. Monthly negative tweets + Daily positive tweets
37. Weekly positive tweets + Daily positive tweets
38. Monthly positive tweets + Daily positive tweets
39. Daily positive tweets + Daily negative tweets + Daily search volume
40. Daily positive tweets + Daily negative tweets + Weekly search volume
41. Daily positive tweets + Daily negative tweets + Monthly search volume
42. Daily positive tweets lag1 + Daily negative tweets lag1 + Monthly search volume lag1
43. Daily positive tweets + Daily negative tweets + Monthly positive tweets + Monthly negative tweets
44. Daily positive tweets + Daily negative tweets + Daily neutral tweets + Daily search volume
45. Daily positive tweets + Daily negative tweets + Daily neutral tweets + Weekly search volume
46. Daily positive tweets + Daily negative tweets + Daily neutral tweets + Monthly search volume
47. Daily positive tweets lag1 + Daily negative tweets lag1 + Daily neutral tweets lag1 + Monthly search volume lag1
48. Daily positive tweets + Daily negative tweets + Monthly positive tweets + Monthly negative tweets + Monthly search volume
49. Daily positive tweets + Daily negative tweets + Daily neutral tweets + Monthly positive tweets + Monthly negative tweets + Monthly neutral tweets + Daily search volume
50. Complete D-W-M cascade

Figure A.1: All specifications of HAR model

A.2 Measurement errors

Model	Average error rate			Total error rate		
	RMSE	MASE	MAE	RMSE	MASE	MAE
HAR basic	0.10%	0.78%	0.50%	1.7691	2.6206	2.6152
$T^{n_3,+}$	15.24%	4.68%	4.34%	6.1702	3.7389	3.7335
$T^{n_3,-}$	12.16%	2.09%	1.98%	5.9828	3.1160	3.1106
$T^{n_3,0}$	7.41%	7.63%	7.34%	4.1988	4.9758	4.9706
S^{n_3}	0.00%	0.09%	0.18%	1.7444	2.4177	2.4123
S^{n_2}	0.36%	0.00%	0.00%	1.7937	2.3350	2.3294
S^{n_1}	0.58%	1.57%	1.49%	1.9215	2.8815	2.8761
$T^{n_2,-}$	13.98%	5.10%	4.91%	6.5174	4.1518	4.1462
$T^{n_1,-}$	14.59%	17.67%	17.26%	6.7458	8.3899	8.3845
$T^{n_2,+}$	26.57%	13.85%	12.77%	9.2588	6.4053	6.4000
$T^{n_1,+}$	22.43%	22.50%	22.30%	8.4635	9.3580	9.3526
$T^{n_3,+} + S^{n_1}$	15.58%	5.27%	5.15%	6.2824	3.9544	3.9490
$T^{n_3,-} + S^{n_1}$	12.24%	2.28%	2.44%	6.0274	3.2261	3.2207
$T^{n_3,-} + T^{n_3,+}$	26.33%	5.65%	5.53%	10.0999	4.1533	4.1479
$T^{n_3,+} + S^{n_2}$	15.73%	4.21%	4.01%	6.2850	3.5599	3.5543
$T^{n_3,-} + S^{n_2}$	11.98%	1.41%	1.50%	5.8715	2.8605	2.8549
$T^{n_3,-} + S^{n_3}$	12.33%	2.13%	2.23%	6.0499	3.1403	3.1349
$T^{n_3,-} + S^{n_3}$	15.76%	4.67%	4.48%	6.3500	3.7538	3.7485
$T^{n_2,-} + S^{n_3}$	13.84%	4.56%	4.69%	6.4823	3.9964	3.9909
$T^{n_1,-} + S^{n_3}$	13.66%	16.28%	16.56%	6.4962	8.0146	8.0092
$T^{n_2,+} + S^{n_3}$	25.49%	13.53%	12.55%	8.9469	6.2884	6.2832
$T^{n_1,+} + S^{n_3}$	21.39%	20.21%	20.61%	8.1862	8.6879	8.6826
$T^{n_2,-} + S^{n_2}$	13.13%	3.75%	3.76%	6.1899	3.6627	3.6570
$T^{n_1,-} + S^{n_2}$	10.96%	13.50%	13.71%	5.4747	6.9561	6.9506
$T^{n_2,+} + S^{n_2}$	27.78%	13.94%	12.72%	9.5871	6.3494	6.3440
$T^{n_1,+} + S^{n_2}$	22.18%	20.51%	20.78%	8.3367	8.6679	8.6623
$T^{n_2,-} + S^{n_1}$	13.97%	4.83%	5.12%	6.5410	4.1561	4.1505
$T^{n_1,-} + S^{n_1}$	14.83%	17.78%	18.33%	6.8821	8.5432	8.5378
$T^{n_2,+} + S^{n_1}$	27.42%	14.50%	13.69%	9.5235	6.6661	6.6609
$T^{n_1,+} + S^{n_1}$	22.94%	22.90%	22.71%	8.6738	9.5635	9.5580
$T^{n_2,-} + T^{n_3,-}$	17.22%	5.15%	4.94%	7.6577	4.1684	4.1628
$T^{n_1,-} + T^{n_3,-}$	17.17%	8.80%	8.97%	7.8347	5.5699	5.5646
$T^{n_2,+} + T^{n_3,-}$	34.04%	13.43%	12.43%	12.1210	6.3662	6.3609

Model	Average error rate			Total error rate		
	RMSE	MASE	MAE	RMSE	MASE	MAE
$T^{n_1,+} + T^{n_3,-}$	31.92%	20.05%	20.19%	11.9231	8.6891	8.6837
$T^{n_2,-} + T^{n_3,+}$	26.51%	8.04%	7.86%	10.1567	4.9941	4.9885
$T^{n_1,-} + T^{n_3,+}$	23.27%	15.56%	15.61%	9.1316	7.6096	7.6042
$T^{n_2,+} + T^{n_3,+}$	29.14%	13.32%	12.10%	10.0398	6.2439	6.2386
$T^{n_1,+} + T^{n_3,+}$	24.27%	18.36%	18.49%	8.9315	7.9996	7.9942
$T^{n_3,-} + T^{n_3,+} + S^{n_1}$	26.36%	5.71%	5.83%	10.1293	4.2223	4.2168
$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	26.47%	4.98%	5.02%	10.0983	3.9134	3.9078
$T^{n_3,-} + T^{n_3,+} + S^{n_3}$	27.31%	6.16%	6.09%	10.4199	4.3316	4.3262
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + S^{n_1}_{t-1}$	26.39%	5.72%	5.84%	10.1453	4.2284	4.2229
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+}$	41.55%	27.27%	27.50%	15.0515	11.2763	10.9331
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_1}$	32.25%	9.69%	10.01%	12.2475	5.7947	5.7896
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_2}$	32.07%	8.38%	8.64%	12.0947	5.2451	5.2399
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_3}$	32.49%	9.19%	9.53%	12.3229	5.6179	5.6130
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + T^{n_3,0}_{t-1} + S^{n_1}_{t-1}$	31.96%	8.86%	8.64%	11.9402	5.3106	4.9891
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+} + S^{n_1}$	42.51%	28.46%	29.10%	15.3834	11.6892	11.3453
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + T^{n_1,-} + T^{n_1,+} + T^{n_1,0} + S^{n_3}$	56.56%	44.24%	44.10%	19.6035	16.3099	15.9945
Complete cascade	66.89%	47.28%	47.72%	22.7639	17.4464	17.1355

Table A.1: Average and total error rates, Original data, Source: own analysis

Model	Average error rate			Total error rate		
	RMSE	MASE	MAE	RMSE	MASE	MAE
HAR basic	0.10%	0.99%	0.71%	1.6850	2.5868	2.5814
$T^{n_3,+}$	2.99%	2.31%	1.98%	2.4725	2.9301	2.9247
$T^{n_3,-}$	0.07%	0.66%	0.46%	1.7110	2.5269	2.5215
$T^{n_3,0}$	7.17%	7.76%	7.47%	4.0410	4.8993	4.8941
S^{n_3}	0.00%	0.31%	0.40%	1.6596	2.3803	2.3750
S^{n_2}	0.36%	0.21%	0.21%	1.7084	2.2904	2.2848
S^{n_1}	0.58%	1.79%	1.71%	1.8335	2.8321	2.8267
$T^{n_2,-}$	7.03%	6.30%	5.65%	3.6581	4.2069	4.2013
$T^{n_1,-}$	10.10%	14.93%	14.34%	5.0751	7.3042	7.2988
$T^{n_2,+}$	12.61%	10.85%	9.81%	5.0766	5.3775	5.3723
$T^{n_1,+}$	8.94%	13.30%	13.14%	4.4181	6.4737	6.4683
$T^{n_3,+} + S^{n_1}$	3.31%	2.91%	2.80%	2.5735	3.1298	3.1244
$T^{n_3,-} + S^{n_1}$	0.16%	0.85%	0.92%	1.7552	2.6278	2.6224
$T^{n_3,-} + T^{n_3,+}$	2.68%	1.77%	1.58%	2.4465	2.8380	2.8326
$T^{n_3,+} + S^{n_2}$	3.33%	1.83%	1.64%	2.5419	2.7369	2.7313
$T^{n_3,-} + S^{n_2}$	0.08%	0.00%	0.00%	1.6655	2.2687	2.2632
$T^{n_3,-} + S^{n_3}$	0.21%	0.70%	0.70%	1.7659	2.5475	2.5421
$T^{n_3,-} + S^{n_3}$	3.18%	2.31%	2.12%	2.5533	2.9454	2.9400
$T^{n_2,-} + S^{n_3}$	6.88%	5.61%	5.28%	3.6168	3.9949	3.9894
$T^{n_1,-} + S^{n_3}$	9.41%	13.63%	13.75%	4.9130	6.9671	6.9617
$T^{n_2,+} + S^{n_3}$	12.35%	10.57%	9.62%	5.0067	5.2715	5.2662
$T^{n_1,+} + S^{n_3}$	8.20%	11.22%	11.65%	4.2355	5.8867	5.8813
$T^{n_2,-} + S^{n_2}$	6.51%	4.73%	4.29%	3.4520	3.6441	3.6384
$T^{n_1,-} + S^{n_2}$	8.35%	11.66%	11.76%	4.4644	6.1702	6.1647
$T^{n_2,+} + S^{n_2}$	12.97%	10.42%	9.24%	5.1527	5.1580	5.1526
$T^{n_1,+} + S^{n_2}$	8.50%	11.34%	11.63%	4.2461	5.8040	5.7984
$T^{n_2,-} + S^{n_1}$	7.49%	6.48%	6.24%	3.8109	4.3032	4.2976
$T^{n_1,-} + S^{n_1}$	9.80%	14.79%	15.16%	5.0346	7.3715	7.3661
$T^{n_2,+} + S^{n_1}$	13.41%	11.51%	10.70%	5.3129	5.6027	5.5975
$T^{n_1,+} + S^{n_1}$	9.25%	13.90%	13.72%	4.5713	6.7254	6.7200
$T^{n_2,-} + T^{n_3,-}$	7.39%	6.40%	5.70%	3.7797	4.2181	4.2125
$T^{n_1,-} + T^{n_3,-}$	5.24%	7.26%	7.29%	3.5988	4.9123	4.9070
$T^{n_2,+} + T^{n_3,-}$	11.01%	9.55%	8.49%	4.7097	5.0489	5.0436
$T^{n_1,+} + T^{n_3,-}$	7.27%	9.84%	9.89%	3.9726	5.4540	5.4486
$T^{n_2,-} + T^{n_3,+}$	8.56%	6.53%	5.92%	4.0971	4.2338	4.2282

Model	Average error rate			Total error rate		
	RMSE	MASE	MAE	RMSE	MASE	MAE
$T^{n_1,-} + T^{n_3,+}$	8.78%	11.44%	11.35%	4.5911	6.1615	6.1562
$T^{n_2,+} + T^{n_3,+}$	13.00%	10.77%	9.52%	5.1885	5.3252	5.3200
$T^{n_1,+} + T^{n_3,+}$	8.63%	11.02%	11.10%	4.2335	5.6688	5.6634
$T^{n_3,-} + T^{n_3,+} + S^{n_1}$	2.71%	1.84%	1.88%	2.4740	2.9018	2.8964
$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	2.78%	1.09%	1.05%	2.4449	2.5860	2.5804
$T^{n_3,-} + T^{n_3,+} + S^{n_3}$	3.14%	2.20%	2.06%	2.6129	2.9893	2.9839
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + S^{n_1}_{t-1}$	26.39%	5.95%	6.07%	10.0433	4.2042	4.1987
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+}$	41.55%	27.55%	27.78%	14.9469	11.2659	10.9227
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_1}$	9.23%	6.00%	6.21%	4.6891	4.5188	4.5137
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_2}$	9.14%	4.69%	4.85%	4.5779	3.9732	3.9680
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_3}$	9.07%	5.43%	5.67%	4.6485	4.3281	4.3231
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + T^{n_3,0}_{t-1} + S^{n_1}_{t-1}$	31.96%	9.10%	8.87%	11.8376	5.2873	4.9658
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+} + S^{n_1}$	42.51%	28.74%	29.38%	15.2768	11.6768	11.3330
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + T^{n_1,-} + T^{n_1,+} + T^{n_1,0} + S^{n_3}$	56.56%	44.55%	44.41%	19.4924	16.2973	15.9819
Complete cascade	66.89%	47.60%	48.04%	22.6612	17.4673	17.1563

Table A.2: Average and total error rates, $\mu + 3\sigma$ winsorized data, Source: own analysis

Model	MASE	RMSE	MAE
HAR basic	0.99931	0.00078	0.00058
S^{n_2}	0.99437	0.00078	0.00058
S^{n_3}	0.99618	0.00078	0.00058
$T^{n_3,-}$	1.01404	0.00088	0.00059
S^{n_1}	1.00919	0.00078	0.00059
$T^{n_3,+}$	1.03747	0.00090	0.00060
$T^{n_2,-}$	1.04319	0.00089	0.00061
$T^{n_3,0}$	1.06731	0.00084	0.00062
$T^{n_2,+}$	1.12138	0.00099	0.00066
$T^{n_1,+}$	1.21609	0.00096	0.00071
$T^{n_1,-}$	1.16600	0.00089	0.00068
$T^{n_3,-} + S^{n_2}$	1.00929	0.00087	0.00059
$T^{n_3,-} + S^{n_3}$	1.01650	0.00088	0.00059
$T^{n_3,-} + S^{n_1}$	1.01862	0.00088	0.00059
$T^{n_3,-} + T^{n_3,+}$	1.04934	0.00099	0.00061
$T^{n_3,+} + S^{n_2}$	1.03425	0.00090	0.00060
$T^{n_3,-} + S^{n_3}$	1.03890	0.00090	0.00060
$T^{n_3,+} + S^{n_1}$	1.04562	0.00090	0.00061
$T^{n_2,-} + S^{n_2}$	1.03180	0.00088	0.00060
$T^{n_2,-} + S^{n_3}$	1.04101	0.00089	0.00060
$T^{n_2,-} + T^{n_3,-}$	1.04349	0.00091	0.00061
$T^{n_2,-} + T^{n_3,+}$	1.07256	0.00099	0.00062
$T^{n_2,-} + S^{n_1}$	1.04523	0.00089	0.00061
$T^{n_1,-} + T^{n_3,-}$	1.08351	0.00091	0.00063
$T^{n_2,+} + T^{n_3,-}$	1.11801	0.00105	0.00066
$T^{n_2,+} + S^{n_2}$	1.12083	0.00100	0.00066
$T^{n_2,+} + T^{n_3,+}$	1.11472	0.00101	0.00065
$T^{n_2,+} + S^{n_3}$	1.11912	0.00098	0.00066
$T^{n_1,+} + T^{n_3,-}$	1.19515	0.00103	0.00069
$T^{n_2,+} + S^{n_1}$	1.13046	0.00099	0.00066
$T^{n_1,+} + T^{n_3,+}$	1.17818	0.00097	0.00068
$T^{n_1,-} + T^{n_3,+}$	1.14954	0.00096	0.00067
$T^{n_1,+} + S^{n_2}$	1.20096	0.00095	0.00070
$T^{n_1,+} + S^{n_3}$	1.19927	0.00095	0.00069
$T^{n_1,-} + S^{n_2}$	1.13072	0.00087	0.00066
$T^{n_1,+} + S^{n_1}$	1.22014	0.00096	0.00071

Model	**MASE**	**RMSE**	**MAE**
$T^{n_1,-} + S^{n_3}$	1.15905	0.00089	0.00067
$T^{n_1,-} + S^{n_1}$	1.17659	0.00090	0.00068
$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	1.04427	0.00099	0.00061
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + S^{n_1}_{t-1}$	1.05231	0.00099	0.00061
$T^{n_3,-} + T^{n_3,+} + S^{n_1}$	1.05231	0.00099	0.00061
$T^{n_3,-} + T^{n_3,+} + S^{n_3}$	1.05495	0.00099	0.00061
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + T^{n_3,0}_{t-1} + S^{n_1}_{t-1}$	1.09542	0.00104	0.00064
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+}$	1.27121	0.00111	0.00074
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_2}$	1.08028	0.00103	0.00063
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_3}$	1.08914	0.00103	0.00063
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_1}$	1.09386	0.00103	0.00063
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+} + S^{n_1}$	1.28999	0.00111	0.00075
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + T^{n_1,-} + T^{n_1,+} + T^{n_1,0} + S^{n_3}$	1.43532	0.00122	0.00083
Complete cascade	1.49279	0.00132	0.00086

Table A.3: Average MASE, RMSE and MAE, Original data, Source: own analysis

Model	**MASE**	**RMSE**	**MAE**
HAR basic	0.99931	0.000781	0.000582
S^{n_2}	0.99437	0.000783	0.000578
S^{n_3}	0.99618	0.00078	0.000578
$T^{n_3,-}$	0.99675	0.000781	0.00058
S^{n_1}	1.00919	0.000785	0.000587
$T^{n_3,+}$	1.01191	0.000804	0.00059
$T^{n_2,-}$	1.04832	0.000835	0.000613
$T^{n_3,0}$	1.06634	0.000836	0.000621
$T^{n_2,+}$	1.08958	0.000879	0.000639
$T^{n_1,+}$	1.12260	0.00085	0.000653
$T^{n_1,-}$	1.13450	0.000859	0.000662
$T^{n_3,-} + S^{n_2}$	0.99223	0.000781	0.000576
$T^{n_3,-} + S^{n_3}$	0.99918	0.000782	0.00058
$T^{n_3,-} + S^{n_1}$	1.00135	0.000782	0.000581
$T^{n_3,-} + T^{n_3,+}$	1.00787	0.000801	0.000587
$T^{n_3,+} + S^{n_2}$	1.00854	0.000806	0.000587
$T^{n_3,-} + S^{n_3}$	1.01329	0.000805	0.00059
$T^{n_3,+} + S^{n_1}$	1.02004	0.000806	0.000593
$T^{n_2,-} + S^{n_2}$	1.03481	0.000831	0.000604
$T^{n_2,-} + S^{n_3}$	1.04459	0.000834	0.000609
$T^{n_2,-} + T^{n_3,-}$	1.04878	0.000838	0.000613
$T^{n_2,-} + T^{n_3,+}$	1.05101	0.000847	0.000614
$T^{n_2,-} + S^{n_1}$	1.05414	0.000839	0.000614
$T^{n_1,-} + T^{n_3,-}$	1.06459	0.000821	0.000618
$T^{n_2,+} + T^{n_3,-}$	1.07650	0.000866	0.000631
$T^{n_2,+} + S^{n_2}$	1.08394	0.000882	0.000636
$T^{n_2,+} + T^{n_3,+}$	1.08671	0.000882	0.000638
$T^{n_2,+} + S^{n_3}$	1.08773	0.000877	0.000637
$T^{n_1,+} + T^{n_3,-}$	1.09032	0.000837	0.000633
$T^{n_2,+} + S^{n_1}$	1.09842	0.000885	0.000643
$T^{n_1,+} + T^{n_3,+}$	1.10232	0.000848	0.00064
$T^{n_1,-} + T^{n_3,+}$	1.10484	0.000849	0.000642
$T^{n_1,+} + S^{n_2}$	1.10759	0.000847	0.000642
$T^{n_1,+} + S^{n_3}$	1.10784	0.000844	0.000641
$T^{n_1,-} + S^{n_2}$	1.10891	0.000846	0.000644
$T^{n_1,+} + S^{n_1}$	1.12842	0.000853	0.000656

Model	**MASE**	**RMSE**	**MAE**
$T^{n_1,-} + S^{n_3}$	1.12868	0.000854	0.000655
$T^{n_1,-} + S^{n_1}$	1.14267	0.000857	0.000662
$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	1.00266	0.000802	0.000583
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + S^{n_1}_{t-1}$	1.01001	0.000805	0.000588
$T^{n_3,-} + T^{n_3,+} + S^{n_1}$	1.01093	0.000802	0.000587
$T^{n_3,-} + T^{n_3,+} + S^{n_3}$	1.01265	0.000805	0.000589
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + T^{n_3,0}_{t-1} + S^{n_1}_{t-1}$	1.05544	0.000858	0.000614
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+}$	1.16590	0.000893	0.000678
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_2}$	1.04037	0.000852	0.000603
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_3}$	1.04848	0.000851	0.000608
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_1}$	1.05388	0.000852	0.000611
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+} + S^{n_1}$	1.17568	0.000897	0.000681
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + T^{n_1,-} + T^{n_1,+} + T^{n_1,0} + S^{n_3}$	1.29954	0.000998	0.000758
Complete cascade	1.43316	0.001141	0.000833

Table A.4: Average MASE, RMSE and MAE, $\mu+3\sigma$ winsorized data, Source: own analysis

MASE	HAR basic	$T^{n_3,-}$ + S^{n_2}	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-}$ + S^{n_3}	$T^{n_3,-}$ + $T^{n_3,+}$ + S^{n_2}	$T^{n_3,-}$ + S^{n_1}
AXP	0.79059	0.82013	0.79420	0.79475	0.80876	0.81386	0.80355	0.83912
BA	0.90672	0.91787	0.93717	0.88616	0.90242	0.89677	0.89575	0.91763
CAT	0.91788	1.03079	1.06450	0.93877	0.89803	0.89889	1.07252	0.97495
CSCO	0.96921	0.76977	0.78333	0.88002	0.93260	0.89442	0.76000	0.91491
CVX	0.87690	0.86443	0.86016	0.88480	0.87753	0.90214	0.87547	0.91598
DD	0.82180	0.92469	0.84983	0.85348	0.87843	0.89167	0.89424	0.96857
DIS	0.67121	0.65868	0.67463	0.67330	0.65882	0.65830	0.65320	0.65067
GE	0.93387	0.96889	0.96827	0.92867	0.93772	0.92889	0.97280	0.94088
GS	0.93069	0.92160	0.93069	0.92780	0.92160	0.92162	0.91677	0.91971
HD	0.97928	0.91502	0.83082	0.94090	1.08498	1.08091	0.91660	1.08826
IBM	1.06863	1.08739	1.22319	1.11634	0.95256	0.99508	1.04991	1.10882
INT	1.13783	1.14319	1.15117	1.14886	1.11551	1.13984	1.14092	0.89229
JNJ	1.10808	1.05985	1.08088	1.08763	1.08232	1.07685	1.04826	1.05019
JPM	0.83853	0.85395	0.85531	0.82838	0.83671	0.82907	0.85469	0.83822
KO	1.43971	1.44546	1.43989	1.44160	1.44507	1.44817	1.46031	1.44704
MCD	1.17647	1.20758	1.19672	1.17318	1.18771	1.18414	1.21240	1.18392
MMM	0.65880	0.65767	0.65880	0.66254	0.65792	0.66000	0.65182	0.66249
MRK	1.02958	1.05066	1.01780	1.04425	1.05143	1.05604	1.29176	1.05419
MSFT	0.54380	0.49409	0.47914	0.48504	0.54073	0.51191	0.53431	0.54033
NKE	1.15682	1.13618	1.16093	1.15762	1.13563	1.13294	1.13050	1.12793
PFE	0.98564	1.07638	0.98381	0.99592	1.08813	1.10773	1.17243	1.04007
PG	1.28829	1.31544	1.30387	1.31084	1.28269	1.32159	1.31200	1.28351
T	1.36575	1.27515	1.34010	1.36849	1.31908	1.31676	1.31935	1.34213
TRV	1.12575	1.13244	1.13629	1.13762	1.11647	1.12413	1.13537	1.16535
UNH	0.98367	0.92909	0.96443	0.96763	0.95574	0.95952	0.92408	0.94928
UTX	1.05210	1.01923	1.01843	1.03765	1.05255	1.03715	1.01907	0.98347
V	0.96519	0.96456	0.96876	0.96411	0.95859	0.95847	0.96621	0.95658
VZ	1.08487	1.07517	1.07693	1.08528	1.08532	1.09227	1.07133	1.08456
WMT	1.39228	1.35229	1.37558	1.40343	1.36804	1.37931	1.33973	1.46134
XOM	0.77927	0.69935	0.70534	0.76020	0.76948	0.75699	0.68451	0.73804
Average	0.9993	0.9922	0.9944	0.9962	0.9968	0.9992	1.0027	1.0013
Median	0.9815	0.9941	0.9763	0.9659	0.9572	0.9773	0.9959	0.9718
Max	1.4397	1.4455	1.4399	1.4416	1.4451	1.4482	1.4603	1.4613
Min	0.5438	0.4941	0.4791	0.4850	0.5407	0.5119	0.5343	0.5403
Stdev	0.2124	0.2166	0.2265	0.2208	0.2085	0.2144	0.2268	0.2148

Table A.5: MASE values of the top performing models, $\mu + 3\sigma$ winsorized data, Source: own analysis

RMSE	HAR basic	$T^{n_3,-}$ + S^{n_2}	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-}$ + S^{n_3}	$T^{n_3,-}$ + $T^{n_3,+}$ + S^{n_2}	$T^{n_3,-}$ + S^{n_1}
AXP	0.00057	0.00059	0.00058	0.00058	0.00059	0.00059	0.00058	0.00060
BA	0.00091	0.00100	0.00103	0.00093	0.00092	0.00092	0.00100	0.00094
CAT	0.00096	0.00103	0.00104	0.00098	0.00096	0.00096	0.00108	0.00098
CSCO	0.00132	0.00120	0.00121	0.00127	0.00130	0.00127	0.00123	0.00128
CVX	0.00055	0.00055	0.00055	0.00056	0.00055	0.00057	0.00056	0.00057
DD	0.00075	0.00080	0.00074	0.00072	0.00079	0.00075	0.00077	0.00083
DIS	0.00059	0.00060	0.00061	0.00061	0.00058	0.00060	0.00059	0.00060
GE	0.00076	0.00080	0.00080	0.00077	0.00076	0.00077	0.00080	0.00077
GS	0.00089	0.00089	0.00089	0.00089	0.00089	0.00089	0.00089	0.00089
HD	0.00083	0.00083	0.00079	0.00082	0.00089	0.00089	0.00083	0.00089
IBM	0.00092	0.00092	0.00099	0.00095	0.00086	0.00089	0.00090	0.00092
INT	0.00113	0.00113	0.00114	0.00114	0.00111	0.00113	0.00111	0.00099
JNJ	0.00051	0.00050	0.00050	0.00050	0.00050	0.00050	0.00049	0.00049
JPM	0.00083	0.00084	0.00084	0.00082	0.00083	0.00082	0.00084	0.00083
KO	0.00096	0.00097	0.00096	0.00096	0.00097	0.00097	0.00098	0.00097
MCD	0.00070	0.00068	0.00070	0.00070	0.00069	0.00069	0.00069	0.00069
MMM	0.00045	0.00044	0.00044	0.00045	0.00045	0.00045	0.00044	0.00045
MRK	0.00079	0.00080	0.00079	0.00079	0.00080	0.00080	0.00122	0.00080
MSFT	0.00070	0.00064	0.00064	0.00067	0.00069	0.00066	0.00071	0.00069
NKE	0.00087	0.00085	0.00087	0.00087	0.00085	0.00085	0.00085	0.00085
PFE	0.00078	0.00086	0.00078	0.00078	0.00087	0.00089	0.00097	0.00085
PG	0.00054	0.00055	0.00055	0.00056	0.00054	0.00056	0.00055	0.00054
T	0.00085	0.00079	0.00084	0.00085	0.00080	0.00080	0.00082	0.00081
TRV	0.00077	0.00075	0.00077	0.00077	0.00075	0.00075	0.00076	0.00076
UNH	0.00069	0.00072	0.00069	0.00069	0.00074	0.00075	0.00072	0.00074
UTX	0.00084	0.00083	0.00083	0.00084	0.00085	0.00084	0.00082	0.00081
V	0.00085	0.00085	0.00085	0.00085	0.00085	0.00085	0.00086	0.00084
VZ	0.00078	0.00078	0.00077	0.00078	0.00078	0.00079	0.00077	0.00078
WMT	0.00087	0.00083	0.00087	0.00088	0.00084	0.00084	0.00084	0.00088
XOM	0.00044	0.00040	0.00041	0.00043	0.00044	0.00043	0.00040	0.00042
Average	0.00078	0.00078	0.00078	0.00078	0.00078	0.00078	0.00080	0.00078
Median	0.00079	0.00080	0.00079	0.00079	0.00080	0.00080	0.00082	0.00081
Max	0.00132	0.00120	0.00121	0.00127	0.00130	0.00127	0.00123	0.00128
Min	0.00044	0.00040	0.00041	0.00043	0.00044	0.00043	0.00040	0.00042
Stdev	0.00019	0.00019	0.00020	0.00019	0.00019	0.00019	0.00021	0.00018

Table A.6: RMSE values of the top performing models, $\mu + 3\sigma$ winsorized data, Source: own analysis

MAE	HAR basic	$T^{n_3,-}$ + S^{n_2}	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-}$ + S^{n_3}	$T^{n_3,-}$ + $T^{n_3,+}$ + S^{n_2}	$T^{n_3,-}$ + S^{n_1}
AXP	0.00044	0.00045	0.00044	0.00044	0.00045	0.00045	0.00044	0.00046
BA	0.00068	0.00069	0.00070	0.00067	0.00068	0.00067	0.00067	0.00069
CAT	0.00067	0.00075	0.00077	0.00068	0.00065	0.00065	0.00078	0.00071
CSCO	0.00098	0.00078	0.00079	0.00089	0.00095	0.00091	0.00077	0.00093
CVX	0.00042	0.00041	0.00041	0.00042	0.00042	0.00043	0.00042	0.00044
DD	0.00055	0.00062	0.00057	0.00057	0.00059	0.00059	0.00060	0.00065
DIS	0.00046	0.00045	0.00046	0.00046	0.00045	0.00045	0.00045	0.00045
GE	0.00064	0.00067	0.00067	0.00064	0.00065	0.00064	0.00067	0.00065
GS	0.00066	0.00065	0.00066	0.00066	0.00065	0.00065	0.00065	0.00065
HD	0.00054	0.00051	0.00046	0.00052	0.00060	0.00060	0.00051	0.00061
IBM	0.00068	0.00070	0.00078	0.00071	0.00061	0.00064	0.00067	0.00071
INT	0.00093	0.00093	0.00094	0.00094	0.00091	0.00093	0.00093	0.00073
JNJ	0.00041	0.00039	0.00040	0.00040	0.00040	0.00039	0.00038	0.00038
JPM	0.00059	0.00060	0.00060	0.00058	0.00059	0.00058	0.00060	0.00059
KO	0.00069	0.00069	0.00069	0.00069	0.00069	0.00069	0.00070	0.00069
MCD	0.00045	0.00046	0.00045	0.00045	0.00045	0.00045	0.00046	0.00045
MMM	0.00034	0.00034	0.00034	0.00034	0.00034	0.00034	0.00033	0.00034
MRK	0.00058	0.00059	0.00057	0.00058	0.00059	0.00059	0.00072	0.00059
MSFT	0.00059	0.00054	0.00052	0.00053	0.00059	0.00056	0.00058	0.00059
NKE	0.00068	0.00067	0.00069	0.00068	0.00067	0.00067	0.00067	0.00067
PFE	0.00060	0.00066	0.00060	0.00061	0.00067	0.00068	0.00072	0.00064
PG	0.00043	0.00044	0.00044	0.00044	0.00043	0.00045	0.00044	0.00043
T	0.00057	0.00053	0.00056	0.00057	0.00055	0.00055	0.00055	0.00056
TRV	0.00051	0.00051	0.00051	0.00051	0.00050	0.00051	0.00051	0.00053
UNH	0.00054	0.00051	0.00053	0.00053	0.00053	0.00053	0.00051	0.00052
UTX	0.00061	0.00059	0.00059	0.00060	0.00061	0.00060	0.00059	0.00057
V	0.00060	0.00060	0.00060	0.00060	0.00060	0.00060	0.00060	0.00059
VZ	0.00060	0.00059	0.00060	0.00060	0.00060	0.00060	0.00059	0.00060
WMT	0.00065	0.00063	0.00065	0.00066	0.00064	0.00065	0.00063	0.00069
XOM	0.00036	0.00033	0.00033	0.00035	0.00036	0.00035	0.00032	0.00034
Average	0.00058	0.00058	0.00058	0.00058	0.00058	0.00058	0.00058	0.00058
Median	0.00059	0.00059	0.00058	0.00058	0.00059	0.00060	0.00059	0.00059
Max	0.00098	0.00093	0.00094	0.00094	0.00095	0.00093	0.00093	0.00093
Min	0.00034	0.00033	0.00033	0.00034	0.00034	0.00034	0.00032	0.00034
Stdev	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00014	0.00013

Table A.7: MAE values of the top performing models, $\mu + 3\sigma$ winsorized data, Source: own analysis

RMSE	$T^{n_3,-} + S^{n_2}$	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-} + S^{n_3}$	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	$T^{n_3,-} + S^{n_1}$
AXP	3%	0%	0%	2%	3%	1%	5%
BA	10%	14%	2%	1%	2%	10%	3%
CAT	7%	8%	2%	0%	0%	13%	2%
CSCO	-9%	-9%	-4%	-2%	-4%	-7%	-3%
CVX	-1%	-1%	1%	0%	2%	1%	3%
DD	7%	-1%	-4%	6%	0%	3%	11%
DIS	1%	3%	3%	-2%	1%	-1%	0%
GE	5%	5%	1%	0%	1%	5%	1%
GS	0%	0%	0%	0%	0%	0%	0%
HD	0%	-4%	-1%	7%	7%	0%	7%
IBM	1%	8%	3%	-6%	-3%	-2%	1%
INT	0%	1%	1%	-1%	0%	-2%	-13%
JNJ	-2%	-1%	-1%	-1%	-1%	-3%	-3%
JPM	1%	2%	-1%	0%	-1%	1%	0%
KO	1%	0%	0%	1%	1%	1%	1%
MCD	-2%	0%	0%	-2%	-2%	-2%	-2%
MMM	0%	-1%	0%	0%	1%	-1%	1%
MRK	1%	-1%	0%	1%	1%	54%	1%
MSFT	-9%	-9%	-5%	-3%	-6%	1%	-3%
NKE	-3%	0%	0%	-3%	-3%	-3%	-2%
PFE	10%	0%	1%	12%	14%	24%	9%
PG	2%	1%	3%	0%	4%	2%	0%
T	-7%	-1%	0%	-6%	-6%	-3%	-5%
TRV	-2%	0%	0%	-3%	-3%	-2%	-2%
UNH	5%	0%	-1%	8%	8%	4%	7%
UTX	-2%	-2%	0%	0%	0%	-3%	-4%
V	-1%	0%	0%	-1%	-1%	1%	-1%
VZ	-1%	-1%	0%	0%	1%	-1%	0%
WMT	-5%	-1%	0%	-4%	-4%	-5%	0%
XOM	-9%	-8%	-2%	-1%	-2%	-10%	-5%
Average	0.02%	0.14%	-0.05%	0.11%	0.30%	2.58%	0.31%

Table A.8: The difference between RMSE of the extended model and that of the original model the top performing models, $\mu + 3\sigma$ winsorized data, Source: own analysis

MAE	$T^{n_3,-} + S^{n_2}$	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-} + S^{n_3}$	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	$T^{n_3,-} + S^{n_1}$
AXP	4%	0%	1%	2%	3%	2%	6%
BA	1%	3%	-2%	0%	-1%	-1%	1%
CAT	12%	16%	2%	-2%	-2%	17%	6%
CSCO	-21%	-19%	-9%	-4%	-8%	-22%	-6%
CVX	-1%	-2%	1%	0%	3%	0%	4%
DD	13%	3%	4%	7%	9%	9%	18%
DIS	-2%	1%	0%	-2%	-2%	-3%	-3%
GE	4%	4%	-1%	0%	-1%	4%	1%
GS	-1%	0%	0%	-1%	-1%	-1%	-1%
HD	-7%	-15%	-4%	11%	10%	-6%	11%
IBM	2%	14%	4%	-11%	-7%	-2%	4%
INT	0%	1%	1%	-2%	0%	0%	-22%
JNJ	-4%	-2%	-2%	-2%	-3%	-5%	-5%
JPM	2%	2%	-1%	0%	-1%	2%	0%
KO	0%	0%	0%	0%	1%	1%	1%
MCD	3%	2%	0%	1%	1%	3%	1%
MMM	0%	0%	1%	0%	0%	-1%	1%
MRK	2%	-1%	1%	2%	3%	25%	2%
MSFT	-9%	-12%	-11%	-1%	-6%	-2%	-1%
NKE	-2%	0%	0%	-2%	-2%	-2%	-2%
PFE	9%	0%	1%	10%	12%	19%	6%
PG	2%	1%	2%	0%	3%	2%	0%
T	-7%	-2%	0%	-3%	-4%	-3%	-2%
TRV	1%	1%	1%	-1%	0%	1%	4%
UNH	-6%	-2%	-2%	-3%	-2%	-6%	-3%
UTX	-3%	-3%	-1%	0%	-1%	-3%	-7%
V	0%	0%	0%	-1%	-1%	0%	-1%
VZ	-1%	-1%	0%	0%	1%	-1%	0%
WMT	-3%	-1%	1%	-2%	-1%	-4%	5%
XOM	-10%	-9%	-2%	-1%	-3%	-12%	-5%
Average	-0.98%	-0.77%	-0.68%	-0.33%	-0.30%	0.10%	-0.14%

Table A.9: The difference between MAE of the extended model and that of the original model the top performing models, $\mu + 3\sigma$ winsorized data, Source: own analysis

MASE	$T^{n_3,-} + S^{n_2}$	S^{n_2}	S^{n_3}	$T^{n_3,-}$	$T^{n_3,-} + S^{n_3}$	$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	$T^{n_3,-} + S^{n_1}$
AXP	4%	0%	1%	2%	3%	2%	6%
BA	1%	3%	-2%	0%	-1%	-1%	1%
CAT	12%	16%	2%	-2%	-2%	17%	6%
CSCO	-21%	-19%	-9%	-4%	-8%	-22%	-6%
CVX	-1%	-2%	1%	0%	3%	0%	4%
DD	13%	3%	4%	7%	9%	9%	18%
DIS	-2%	1%	0%	-2%	-2%	-3%	-3%
GE	4%	4%	-1%	0%	-1%	4%	1%
GS	-1%	0%	0%	-1%	-1%	-1%	-1%
HD	-7%	-15%	-4%	11%	10%	-6%	11%
IBM	2%	14%	4%	-11%	-7%	-2%	4%
INT	0%	1%	1%	-2%	0%	0%	-22%
JNJ	-4%	-2%	-2%	-2%	-3%	-5%	-5%
JPM	2%	2%	-1%	0%	-1%	2%	0%
KO	0%	0%	0%	0%	1%	1%	1%
MCD	3%	2%	0%	1%	1%	3%	1%
MMM	0%	0%	1%	0%	0%	-1%	1%
MRK	2%	-1%	1%	2%	3%	25%	2%
MSFT	-9%	-12%	-11%	-1%	-6%	-2%	-1%
NKE	-2%	0%	0%	-2%	-2%	-2%	-2%
PFE	9%	0%	1%	10%	12%	19%	6%
PG	2%	1%	2%	0%	3%	2%	0%
T	-7%	-2%	0%	-3%	-4%	-3%	-2%
TRV	1%	1%	1%	-1%	0%	1%	4%
UNH	-6%	-2%	-2%	-3%	-2%	-6%	-3%
UTX	-3%	-3%	-1%	0%	-1%	-3%	-7%
V	0%	0%	0%	-1%	-1%	0%	-1%
VZ	-1%	-1%	0%	0%	1%	-1%	0%
WMT	-3%	-1%	1%	-2%	-1%	-4%	5%
XOM	-10%	-9%	-2%	-1%	-3%	-12%	-5%
Average	-0.71%	-0.49%	-0.31%	-0.26%	-0.01%	0.34%	0.20%

Table A.10: The difference between MASE of the extended model and that of the original model the top performing models, $\mu + 3\sigma$ winsorized data, Source: own analysis

Model	**MASE Original**	**MASE** $\mu + 3\sigma$	**MASE** $\mu + 2.5\sigma$	**MASE** $\mu + 2\sigma$	**MASE** $\mu + 1.5\sigma$
HAR basic	0.99931	0.99931	0.99931	0.99931	0.99931
$T^{n_3,+}$	1.03747	1.01191	1.00998	1.00829	1.00648
$T^{n_3,-}$	1.01404	0.99675	0.99616	0.99560	0.99500
$T^{n_3,0}$	1.06731	1.06634	1.06592	1.06560	1.06535
S^{n_3}	0.99618	0.99618	0.99618	0.99618	0.99618
S^{n_2}	0.99437	0.99437	0.99437	0.99437	0.99437
S^{n_1}	1.00919	1.00919	1.00919	1.00919	1.00919
$T^{n_2,-}$	1.04319	1.04832	1.04713	1.04594	1.04480
$T^{n_1,-}$	1.16600	1.13450	1.13311	1.13173	1.13033
$T^{n_2,+}$	1.12138	1.08958	1.08596	1.08238	1.07881
$T^{n_1,+}$	1.21609	1.12260	1.11784	1.11332	1.10889
$T^{n_3,+} + S^{n_1}$	1.04562	1.02004	1.01812	1.01643	1.01462
$T^{n_3,-} + S^{n_1}$	1.01862	1.00135	1.00076	1.00019	0.99960
$T^{n_3,-} + T^{n_3,+}$	1.04934	1.00787	1.00556	1.00356	1.00142
$T^{n_3,+} + S^{n_2}$	1.03425	1.00854	1.00661	1.00493	1.00312
$T^{n_3,-} + S^{n_2}$	1.00929	0.99223	0.99165	0.99109	0.99050
$T^{n_3,-} + S^{n_3}$	1.03890	1.01329	1.01142	1.00979	1.00805
$T^{n_3,-} + S^{n_3}$	1.01650	0.99918	0.99859	0.99802	0.99742
$T^{n_2,-} + S^{n_3}$	1.04101	1.04459	1.04341	1.04223	1.04111
$T^{n_1,-} + S^{n_3}$	1.15905	1.12868	1.12733	1.12600	1.12465
$T^{n_2,+} + S^{n_3}$	1.11912	1.08773	1.08450	1.08128	1.07809
$T^{n_1,+} + S^{n_3}$	1.19927	1.10784	1.10350	1.09940	1.09543
$T^{n_2,-} + S^{n_2}$	1.03180	1.03481	1.03373	1.03265	1.03184
$T^{n_1,-} + S^{n_2}$	1.13072	1.10891	1.10778	1.10667	1.10561
$T^{n_2,+} + S^{n_2}$	1.12083	1.08394	1.08028	1.07670	1.07315
$T^{n_1,+} + S^{n_2}$	1.20096	1.10759	1.10344	1.09953	1.09573
$T^{n_2,-} + S^{n_1}$	1.04523	1.05414	1.05294	1.05175	1.05059
$T^{n_1,-} + S^{n_1}$	1.17659	1.14267	1.14109	1.13953	1.13794
$T^{n_2,+} + S^{n_1}$	1.13046	1.09842	1.09484	1.09127	1.08771
$T^{n_1,+} + S^{n_1}$	1.22014	1.12842	1.12416	1.12000	1.11593
$T^{n_2,-} + T^{n_3,-}$	1.04349	1.04878	1.04815	1.04755	1.04693
$T^{n_1,-} + T^{n_3,-}$	1.08351	1.06459	1.06379	1.06304	1.06228
$T^{n_2,+} + T^{n_3,-}$	1.11801	1.07650	1.07281	1.06914	1.06544
$T^{n_1,+} + T^{n_3,-}$	1.19515	1.09032	1.08573	1.08130	1.07699
$T^{n_2,-} + T^{n_3,+}$	1.07256	1.05101	1.04819	1.04561	1.04296
$T^{n_1,-} + T^{n_3,+}$	1.14954	1.10484	1.10198	1.09936	1.09661
$T^{n_2,+} + T^{n_3,+}$	1.11472	1.08671	1.08230	1.07820	1.07402
$T^{n_1,+} + T^{n_3,+}$	1.17818	1.10232	1.09761	1.09318	1.08880
$T^{n_3,-} + T^{n_3,+} + S^{n_1}$	1.05231	1.01093	1.00868	1.00675	1.00464
$T^{n_3,-} + T^{n_3,+} + S^{n_2}$	1.04427	1.00266	1.00039	0.99841	0.99627
$T^{n_3,-} + T^{n_3,+} + S^{n_3}$	1.05495	1.01265	1.01037	1.00839	1.00624
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + T^{n_3,0}_{t-1} + S^{n_1}_{t-1}$	1.09542	1.05544	1.05311	1.05117	1.04904
$T^{n_3,-}_{t-1} + T^{n_3,+}_{t-1} + S^{n_1}_{t-1}$	1.05231	1.01001	1.00776	1.00582	1.00372
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+}$	1.27121	1.16590	1.16081	1.15612	1.15155
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_1}$	1.09386	1.05388	1.05155	1.04961	1.04748
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_2}$	1.08028	1.04037	1.03804	1.03609	1.03393
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + S^{n_3}$	1.08914	1.04848	1.04613	1.04417	1.04200
$T^{n_3,-} + T^{n_3,+} + T^{n_3,0} + T^{n_1,-} + T^{n_1,+} + T^{n_1,0} + S^{n_3}$	1.43532	1.29954	1.29604	1.29288	1.29024
$T^{n_3,-} + T^{n_3,+} + T^{n_1,-} + T^{n_1,+} + S^{n_1}$	1.28999	1.17568	1.17057	1.16573	1.16117
Complete cascade	1.49279	1.43316	1.42580	1.41905	1.41250

Table A.11: Robustness analysis: average MASE values for various winsorization levels, Source: own analysis