UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER'S THESIS

THE USABILITY OF GOOGLE TRENDS DATA IN EQUITY INDEX VOLATILITY PREDICTION - A COMPARISON OF THE SELECTED EUROPEAN LANGUAGES

Ljubljana, May 2023

NIKO JARMO KALEVI JÄRNBERG

AUTHORSHIP STATEMENT

The undersigned Niko Jarmo Kalevi Järnberg, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this written final work of studies with the title Can Google Trends data be used to predict equity index volatility? Multilingual evidence from Europe, prepared under supervision of Prof. Dr. Igor Lončarski

DECLARE

1. this written final work of studies to be based on the results of my own research;

2. the printed form of this written final work of studies to be identical to its electronic form;

3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB LU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the SEB LU's Technical Guidelines for Written Works;

4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;

5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;

6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;

7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;

8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System

9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;

10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, ________________________ Author's signature: ________________________

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ARCH	Autoregressive Conditional Heteroscedasticity
DAX	Deutscher Aktienindex (German stock Index)
EMH	Efficient Market Hypothesis
ETLA	Research Institute of Finnish Economy
EWMA	Exponentially Weighted Moving Average
FTSE 100	The Financial Times Stock Exchange 100 Index
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
HAR	Heterogenous Autoregressive
IBEX	Iberian Index
iid	Independent and Identically Distributed
IV	Integrated Variance
LT	Long Term

MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MEM	Multiplicative Error Model
MSE	Mean Squared Error
MT	Medium Term
MZ	Mincer-Zarnowitz Regressions
OLS	Ordinary Least Squares
OMXHPI	OMX Helsinki All-Share Index
OMXSPI	OMX Stockholm Performance Index
RV	Realized Variance
RVol	Realized Volatility
ST	Short Term

1 INTRODUCTION

Predicting stock market dynamics is perhaps the most researched topic in finance, and major advancements in the field are relatively infrequent. Despite this, the use of online data about the behavior of large groups of people has opened a new frontier in the study of stock markets. The theoretical underpinning of using online data can be traced back to the classic Efficient Market Hypothesis (EMH). According to the EMH, new information is quickly and efficiently embedded in the stock price, and due to the random arrival of new information, stock prices also follow a random path (Fama, 1965). In this view, it is the information that drives stock price changes, and therefore, one should be able to use newly created or new types of information to predict market dynamics to a certain extent.

The advent of wider internet use has created a new stream of online information that should also be reflected in stock dynamics, as attested by information efficiency. The information used in the thesis is the statistics about the frequency of online search words in the Google searches. The reason online searches are especially interesting dataset to explore as a financial leading indicator can be explained by the consumer and financial decision process models, where a purchase decision, such as participation in financial markets, is preceded by information search (Engel et al., 1971; Simon, 1955). Therefore, information search generally precedes any other observable market action and should, as such, be predictive.

Google search data and other crowd-generated online datasets, such as Wikipedia edits and Twitter posts, have been applied numerous times to predict stock returns in the recent years, with varied success. Examples of such research include Market sentiment (Tetlock, 2007), Twitter (Bollen et al., 2011), Wikipedia (Moat et al., 2013), and Google Trends (Preis et al., 2013).

Many authors have noted and superficially analyzed the apparent relationship of volatility (squared return) and online activity, but directly using realized volatility and an appropriate benchmark volatility model to investigate the issue more in detail is rare.

The only directly comparable research (as far as I am aware) investigating the relationship of realized volatility and search queries comes from Dimpfl et al. (2012) and Hamid et al. (2015). In these research studies, they used the search frequency of the index name or closely related term as a proxy for investor attention and then used it as a predictor for realized volatility. In both cases, the researchers have found that Google Trends is useful in volatility prediction.

My research is aimed at expanding the knowledge base in this area, by expanding the types of keywords used and expanding the research to cover more regional markets. The following chapter will describe the research question in more detail.

In essence, this work aims to investigate the question of whether Google search volume can predict future volatility in the stock market. As is immediately evident, such question by all account is too broad to be tackled comprehensively, and therefore, this work only aims to assess the topic partially.

The more operational formulation can be articulated as whether or not the search volume of words frequently used in the financial context, in the majority language of the location where the stock market index originates, improves the step-ahead prediction of realized volatility of a stock market index when it is used as an additional predictor in Heterogenous Autoregressive (HAR) model.

To clarify to the reader, below are listed the key aspects of the operational 'hypothesis' which narrow down the broader aim that the work has.

- Word selection: Words whose search volume is used as predictors are words commonly used in the financial context as opposed to alternative keywords that could be chosen. Alternatively, there are numerous words that have certain theoretical validity. Examples of possible alternatives are words associated with crisis sentiment such as the word "crisis" and "recession", words referring to the name of the index such as "FTSE" and "DAX", words relating to company names or brands of the company being part of the index such as "Volkswagen" and "VW". In summary, there are numerous possibilities of selecting the predictor words some of which might be significantly better predictors.
- Language: The financially significant words are always in the majority language of the index 'Domicile', and possible minority language or foreign languages are not used for prediction.
- Index versus individual stocks: The work is aiming to predict only the volatility of the whole index while it might be more fruitful to focus on the volatility of individual stocks.
- Prediction window: predictions are done only on a 1-step (week) ahead, and any possible improvement on longer forecast horizon is not investigated.
- Improvement in prediction over a benchmark model: It would be fully valid to investigate if Google search volume predicts volatility in isolation, but due to the limited practical value of such investigation, I have decided to focus on the improvement over a benchmark that more realistically reveals the possible practical utility of using search volume as a predictor.
- HAR model as the benchmark: There would be numerous alternatives to be used as the benchmark model, but the HAR model (and its log version) is used largely due to it being easily expanded with added predictors and being a sufficiently good model to be used as a benchmark.

Note that the hypothesis is explicitly aimed at understanding whether *prediction* can be improved and is by no means meant to estimate any causal or other explanatory relationship. Similarly, the work is not intended to gauge the true properties of the volatility process, and thus, the possible bias in the estimated parameters is not of any interest. Excellent discussion about this distinction can be found in Shmueli (2010). In summary, based on Shmueli: "In explanatory modeling the focus is on minimizing bias to obtain the most accurate representation of the underlying theory. In contrast, predictive modeling seeks to minimize the combination of bias and estimation variance, occasionally sacrificing theoretical accuracy for improved empirical precision" (Shmueli, 2010, p. 6).

Also, the hypothesis is not tested in terms of statistical significance, and thus, the hypothesis is not formulated in terms of falsification. Instead, the improvement in prediction is simply observed in (pseudo) out-of-sample setting. To determine if a hypothesis is true, two things need to be considered. First, are the improvements in prediction accuracy big enough to be practically useful? Second, do these improvements occur consistently across different keywords, languages, and model specifications?

As noted, the Google Trends has already been applied successfully in volatility predictions by previous authors. This work expands prior research by not directly using the index name but using search volumes of "financially significant" words instead; an approach directly adopted from Preis et al. (2013). This approach allows for the extension of Google Trends' usefulness in volatility forecasting to cover markets and indices that do not have appropriate names for direct use of the search volume for the index name.

There are two reasons that a given index might not be suitable to retrieve Google search data. The first reason is that the search volume for the index name is not sufficient, which occurs frequently with more local indices. The second reason is that the index name might get a different meaning in different languages, regions, or contexts. For instance, "S&P" might refer to credit ratings or the market index. Similarly, "FTSE" might be used in Italy to refer to the local market index instead to its more famous London namesake. Another addition to the existing research is the comparison of several regions and languages in the context of Google Trends data.

Before tackling the research question head on, I will lay the theoretical foundations of the theme by performing an overview of volatility modelling history, as well as the work of other authors that have used Google Trends, or similar data in their research.

2 THEORETICAL FOUNDATIONS AND PREVIOUS WORK ON THE TOPIC

2.1 Efficient market hypothesis and modeling financial price dynamics as random walk

The concept of volatility and its measurement or estimation is inherently linked to the idea that there is a degree of randomness and dispersion in the dynamics of market prices. Although the foundational ideas of modeling financial prices as random walks were presented much earlier by Louis Bachelier in his Ph.D thesis "Theory of Speculation" (Mandelbrot, 1963, p. 43-58), their mainstream acceptance can be attributed to Eugene Fama's seminal work "The Behavior of Stock-Market Prices" in which Fama put forth what is now called the Efficient Market Hypothesis (EMH). In summary, the EMH states that "current stock prices fully reflect available information about the value of the firm, and there is no way to earn excess profits (more than the market overall) by using this information" (Clarke et al., 2001, p. 1). While there are several empirical observations that do not fully conform to the EMH, it is still considered largely to be a good approximation of empirical reality. The EMH is commonly subdivided into three forms based on which information is incorporated into market prices: weak-form efficiency, semi-strong-form efficiency, and strong-form efficiency (Fama, 1970):

- Weak-form efficiency roughly states that current price reflects all information about the past prices only, meaning that no excess profit can be earned by using past price series to predict future prices.
- Semi-strong-form efficiency roughly states that all publicly available information is incorporated in the stock price.
- Strong-form efficiency states that all information, public or private, is incorporated into the prices.

In most instances, the semi-strong form is thought to hold relatively well, but full strong from efficiency is not widely or fully supported.

As an extension of the EMH being true, one can expect that the market prices fluctuate seemingly randomly, a fact which was proven by Samuelson (1965) around the same time as Fama published his work.

As one can observe, the central theme of EMH is how information is incorporated into market prices. In effect, the theory is that market prices change upon the arrival of new information. This in turn leads to the foundational idea that justifies using Google search volume as predictor of volatility; when people are searching financially significant terms, in large volumes, this indicates that new information is arriving and being incorporated into market prices.

2.2 Model of financial and consumer decision relating to information search

The decision-making related to financial investments can be thought within the frame of consumer purchase decision-making. One way to depict such process is to divide it into distinct steps that include problem recognition, information search, evaluation of alternatives, purchase decision, and post-purchase behavior (Kotler, 2016, p.166) as depicted in Figure 1.

Figure 1: Five-stage model of the consumer buying process

Adapted from Kotler (2016).

The consumer decision-making process has been applied to investment decisions by Lin and Lee (2002) and subsequently extended by her thesis supervisor. In this model, the general view is that whenever consumers (or investors) make a choice to invest in securities, which, according to Lee and Lin (2004), is analogous to purchasing intangible goods, the purchase decision is preceded by some form of information search. Therefore, information search plays a key role in the customer purchase process, and nowadays, Google is one of the leading ways people search for information. The connection between information search and Google search volume in the consumer decision process is extensively discussed in 'Quantitative Trendspotting' by Du and Kamakura (2012).

Based on this framework, prior to making an investment decision in the market, investors generate additional Google search volume, followed by incorporating their view into the market price by market participation (or lack thereof). Also, as one can observe, for most market participants, there is supposed to be a degree of time lag between information search and purchase decision that could enable one to use the search as a leading indicator for future purchase (or sales) actions.

While the above makes a compelling chronology of events, I suspect that there are other forms of information searches involved. For instance, one can easily conceive a situation where the chronology is reversed. Consider, for instance, a situation where another information source causes market volatility, which in turn triggers information search behavior and subsequent search volume increase that considerably lags the actual increased volatility. Considering that the level of volatility tends to be persistent, this reversed chronology might lead researchers astray in believing search volume is predictive when it is, in fact, reactive. This is why it is essential to fully account for the autocorrelation in the level of volatility before one concludes search is in any way predictive.

2.3 Overview of volatility and volatility modeling

2.3.1 Defining volatility

Prior to initiating the discourse on various volatility model, one needs to define the concept of volatility in greater detail. In this, I follow Hansen and Lunde's (2001) paper which evaluates the performance of different volatility models.[1] The starting point is to define the variable interest, which is the continuously compounded return on some time interval from t-1 to t on an asset price or index:

$$\boldsymbol{r_t = log(p_t) - log\ (p_{t-1})} \tag{1}$$

Then, the volatility is defined as the standard deviation of the conditional density of $\mathrm{r_t}$. Note that this already implicitly assumes the existence of such standard deviation, which is not a trivial assumption in the short time horizon, since there is significant evidence that returns are derived from Paretian distribution with tail exponent less than 2, meaning that the variance and, by extension, standard deviation are undefined as noted by Mandelbrot (2010) and Fama (1965). In addition to this, the standard deviation of the conditional density is assumed to be time-changing in a fashion that is not fully predictable, meaning that the models discussed here fall in the class of stochastic volatility models.

Before starting the discussion of various volatility models, I want to highlight a fact that is often lost in discussions of 'volatility' and its modeling. This fact is that volatility is a latent variable of the price-generating process that cannot directly be observed from the price data. Even more, the observed proxy for volatility, usually squared return, is very noisy as is noted by Andersen and Bollerslev (1997) when discussing the problematic of 'low explanatory power of volatility models': "However the realized squared returns are poor estimators of the day-by-day movements in volatility as the idiosyncratic component of daily returns is large" (Andersen and Bollerslev, 1997, p. 8). This causes several difficulties in modeling and evaluating the models, especially historically. To remedy this noise in observing the latent variable, the concept of measuring realized volatility from intraday returns plays a part, as I shall discuss in the chapter 2.4.

2.3.2 Historical origins of volatility modeling

The modeling of volatility has its roots in the empirical observations that the magnitude of price changes tends to cluster to cluster in time, so that "large changes tend to be followed by large changes -of either sign- and small changes tend to follow small changes" (Mandelbrot, 1963, p. 418). The first model to garner serious attention was the

---

[1] This section relies heavily on mentioned paper, and therefore, I want to fully acknowledge its significance as a major influence.

Autoregressive Conditional Heteroscedasticity (ARCH) model by Engle (1982). The ARCH model puts forth a simple autoregressive structure for the errors. The model was subsequently extended by Bollerslev (1986) with a moving average component to form a Generalized ARCH (GARCH) model. These two models, and especially GARCH, are the bedrock of the volatility modeling literature that much of the subsequent work is directly based on. In fact, GARCH is still widely used today despite the numerous improvements that have been proposed. Despite its success and lasting popularity, the models are far from perfect, and below are highlighted some aspects of volatility dynamics that are, in some respects, insufficiently captured by the GARCH model.

2.3.3 Characteristics of volatility

Volatility modeling has advanced in various ways since the original ARCH and GARCH models. In most cases, new models or extensions to the older ones have been developed to capture certain aspects of the volatility dynamics that could not be captured by the original ARCH or GARCH. Below are some stylized facts about volatility that ought to be captured by a good volatility model. The list below largely summarizes the points from Engle and Patton (2000), with certain points emphasized and expanded by me.

2.3.3.1 Volatility exhibits (long) persistence

The cornerstone of volatility modeling is to capture the so-called volatility clustering, usually with some form of autoregressive model. Here, the important aspect that Engle and Patton did not emphasize is that the persistence in volatility is rather long, which poses challenges to models that are inherently short memory.

2.3.3.2 Volatility is mean-reverting

Despite volatility having relatively long persistency the volatility still tends to be mean-reverting. This effectively means that current volatility should not affect the forecast of volatility after a sufficiently long-time horizon, which rules out using any infinitely persistent process.

2.3.3.3 Innovations may have an asymmetric impact on volatility

It has been documented already by Black (1976) that large declines in the stock market tend to increase the subsequent volatility more than large positive returns. This effect has then been dubbed the "leverage effect." This means that a volatility model optimally would, in some way, also take into account the direction of the returns when forecasting future volatility. How the volatility forecast is affected by the most recent return is often depicted using the news impact curve introduced by Pagan and Schwert (1990) and named by Engle and Ng (1993).

2.3.3.4 Exogenous variables may influence volatility

As it is obvious, financial asset prices do not develop in isolation from other markets around it, and therefore, other variables, besides past volatility history of an asset itself, should be relevant for future volatility. These exogenous influences may also be deterministic, such as scheduled earnings and macroeconomic announcements. Optimally, a volatility model would be flexible enough to allow the incorporation of external variables.

2.3.3.5 Fat tails

Stochastic volatility model produces excess kurtosis in unconditional density, even if the conditional return distribution is assumed to be Gaussian. This is because the stochastic volatility results in a mixture of conditional Gaussian distributions with different volatilities. This is true even if the future volatility is considered to have a conditional Gaussian distribution. Still, there are reasons to assume, due to the significant excess kurtosis of the return distribution, that even the conditional density of future volatility could be non-Gaussian. Thus, an optimal volatility model fits the tail properties of the return distribution.

2.3.3.6 Realized volatility is a noisy estimate of 'true' volatility

As we have already mentioned in section 2.3.1, volatility is a latent variable, and all ex-post measurements of volatility are always noisy. When observing the realized volatility, let's say in daily intervals, the observed value is based on only one realized return that is drawn from a distribution defined by the conditional standard deviation, and this realization can be very far from the 'true volatility'. This issue is often remedied in volatility models by using realized volatility estimated from intraday data. Theoretically, volatility can be estimated infinitely precisely by observing realized volatility in infinitesimal time intervals, but this approach has practical limitations that will be discussed further in section 2.4.

All in all, this fact has important implications for how accurate the models can be in the first place and what methods are used when evaluating the accuracy of models.

2.3.4 Overview of volatility models

As one can see from above, there are various aspects that an optimal volatility model should capture. Below is a brief summary of the most prominent volatility models in the literature. This section follows the notation conventions of Hansen and Lunde (2001), which is an important reference material along with Bollerslev (2008). When there are competing abbreviations and model names, the names are presented as they occur in

Bollerslev's work, since his work provides the widest account of various models and explicitly attempts to clarify the naming confusion.

2.3.4.1 ARCH

ARCH was introduced by Robert Engle (1982) to describe the volatility of financial returns. The basic ARCH model can be represented by the following equation:

$$\boldsymbol{\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2} \tag{2}$$

2.3.4.2 GARCH

GARCH was introduced by Tim Bollerslev in his doctoral thesis under the supervision of Engle (1986) as an extension to ARCH by adding a moving average component. The basic GARCH model can be represented by the following equation:

$$\boldsymbol{\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-j}^2} \tag{3}$$

2.3.4.3 Integrated GARCH

I-GARCH was introduced by Engle and Bollerslev (1986). Effectively integrated I-GARCH is simply a special case of the GARCH model with infinite persistency property, since the coefficients sum to 1, meaning that the process is a unit root process. I- GARCH model can be represented by familiar GARCH equation with an added constraint.

$$\boldsymbol{\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-1}^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-j}^2} \tag{4}$$

$$\boldsymbol{\textit{Where: } \sum_{i=1}^{q} \beta_i + \sum_{j=1}^{p} \alpha_i = 1} \tag{5}$$

2.3.4.4 Exponentially Weighted Moving Average

Another notable variation of GARCH is the Exponentially Weighted Moving Average model (EWMA), developed by Longerstaey and Spencer (1996), which is again just a special case of I-GARCH. This model deserves a special mention because of its extremely wide usage in practice. Its wide popularity among practitioners can arguably be traced to its origins in JP Morgan, and inclusion in the 'Risk metrics' framework, that can be said to have spearheaded financial risk management as it is understood today. In fact, the approach is so popular that, based on my own experience, many portfolio management software come with inbuilt Value at Risk (VaR) estimation tool that uses the multivariate version of EWMA as the main way to estimate the time-varying variance-covariance matrix.

$$\boldsymbol{\sigma_t^2 = \omega + \sum_{1=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-j}^2} \tag{6}$$

$$\textbf{\textit{Where: }} \boldsymbol{\sum_{i=1}^{q} \beta_i + \sum_{j=1}^{p} \alpha_i = 1 \;\&\; \omega = 0} \tag{7}$$

2.3.4.5 Taylor-Schwert GARCH

The Taylor-Schwert GARCH (TS-GARCH) Taylor (1986) and Schwert (1989) is again a relatively straightforward variation of the original GARCH that simply dampens the effect of large price swings on future volatility by not operating with squared values but with absolute values instead. TS-GARCH model can be represented by the following equation:

$$\boldsymbol{\sigma_t = \omega + \sum_{i=1}^{p} \alpha_i |\varepsilon_{t-i}\,| + \sum_{j=1}^{p} \beta_i \sigma_{t-j}} \tag{1}$$

2.3.4.6 Exponential GARCH

Exponential GARCH (EGARCH) is another formulation that was explicitly designed to capture the asymmetric nature of positive and negative shocks (Nelson, 1991). The model achieves this by including both the absolute value of a normalized return as well as the normalized return itself. Thus, the direction of normalized return can affect the future volatility. EGARCH model can be represented by the following equation:

$$\boldsymbol{ln(\sigma_t^2) = \omega + \sum_{i=1}^{p} [\alpha_i \varepsilon_{t-i} / \sigma_{t-j}^2 + \gamma_i (|\varepsilon_{t-i} / \sigma_{t-i}^2| - E|\varepsilon_{t-i} / \sigma_{t-i}^2|)] + \sum_{j=1}^{q} \beta_i \sigma_{t-j}^2} \tag{9}$$

2.3.4.7 Glosten, Jagannathan, and Runkle - GARCH

Glosten, Jagannathan, and Runkle - GARCH (GJR - GARCH) is one of the most influential generalizations of the original GARCH introduced by the named authors (Glosten et al., 1993). The main addition to GARCH is to add an extra term that combines a binary indicator variable for the direction of the return with the actual squared deviation in the previous period. This, in turn, means that the model is able to capture the leverage effect, unlike the models discussed above. GJR-GARCH model can be represented by the following equation:

$$\boldsymbol{\sigma_t^2 = \omega + \sum_{1=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{1=1}^{p} \gamma_i I(\varepsilon_{t-i} < 0) \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-j}^2} \tag{10}$$

Where: **I** represents the indicator variable

2.3.4.8 Asymmetric -GARCH

Asymmetric GARCH (AGARCH) is yet another model capturing the asymmetry. Here it is achieved by simply augmenting the original GARCH with a non-squared return term (Engle and Ng, 1993). AGARCH model can be represented by the following equation:

$$\boldsymbol{\sigma_t^2 = \omega + \sum_{i=1}^{p}[\alpha_i\varepsilon_{t-i}^2 + \gamma_i\varepsilon_{t-i}] + \sum_{j=1}^{q}\beta_i\sigma_{t-j}^2} \tag{11}$$

2.3.4.9 Threshold GARCH

Threshold GARCH (Zakoian, 1994) is closely related to the GJR GARCH model and can be considered essentially the same model, except that it is for conditional standard deviation instead of conditional variance. The Threshold GARCH model can be represented by the following equation:

$$\boldsymbol{\sigma_t = \omega + \sum_{1=1}^{p}\alpha_i\varepsilon_{t-i} + \sum_{1=1}^{p}\gamma_i I(\varepsilon_{t-i} < 0)|\varepsilon_{t-i}| + \sum_{j=1}^{q}\beta_i\sigma_{t-j}} \tag{12}$$

2.3.4.10 Heterogeneous ARCH

Especially for this paper, Heterogeneous ARCH (HARCH) is very important (Müller et al., 1997). The model marks the first instance where the conditional variance is modeled to be based on different timescale volatility components, which are formed as a sum on various time scales. One could even say that Corsi's HAR-RV (2009) is effectively the same model applied to realized volatility. HARCH model can be represented by the following equation:

$$\boldsymbol{\sigma_t^2 = \omega + \sum_{i=1}^{n}\gamma_i\left(\sum_{j=1}^{p}\varepsilon_{t-i}\right)^2} \tag{13}$$

2.3.4.11 Nonlinear GARCH

Nonlinear GARCH (NGARCH) model introduced by Engle and Bollerslev (1986), can be viewed as a generalization of TSGARCH. The generalization allows the use of an arbitrary exponent for the error and volatility, allowing the model to more flexibly capture different properties of the volatility process, most notably asymmetry. NGARCH model can be represented by the following equation:

$$\boldsymbol{\sigma_t^\delta = \omega + \sum_{i=1}^{p}\alpha_i(|\varepsilon_{t-i}|)^\delta + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^\delta} \tag{14}$$

2.3.4.12 Asymmetric power ARCH

Asymmetric power ARCH (APARCH) model is a further generalization that nests several previously mentioned models, such as GARCH, TS-GARCH, GJR-GARCH and TGARCH, as specific parameterizations (Ding, Granger and Engle, 1993). This makes the model very flexible. APARCH model can be represented by the following equation:

$$\boldsymbol{\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta} \tag{15}$$

2.3.4.13 Fractionally Integrated GARCH

The development of Fractionally Integrated GARCH (FIGARCH) is probably the most significant advancement in the original GARCH model to capture the long-term persistence (geometric instead of exponential decay) observed in volatility (Baillie, Bollerslev and Mikkelsen, 1996). This model applies the autoregressive fractionally integrated moving average representation to volatility. FIGARCH model can be represented by the following equation:

$$\boldsymbol{\phi(L)(1 - L)^d \varepsilon_t^2 = \omega + (1 - \beta(L))v_t} \tag{16}$$

Where:

- $\mathbf{v_t}$ is restricted to be equal to $\varepsilon_t^2 - \sigma_t^2$
- **L** stands for the lag operator
- **d** stands for the order of fractional differentiation, $0 < d < 1$
- $\boldsymbol{\phi}(\mathbf{L})$ and $\boldsymbol{\beta}(\mathbf{L})$ stand for autoregressive lag polynomials

2.4 Realized volatility

Volatility, as defined above, is a latent variable for the price process (Andersen et al., 2005). The fact that the variable is latent poses significant difficulties form modeling volatility and especially for evaluating the performance of the models. This led to now resolved dilemma of why volatility models generally had high statistical significance, while still having such low explanatory power. This situation was improved after the introduction of realized volatility.

The basic idea of realized volatility is that, if we observe returns multiple times a day instead of observing the realized return once a day, we can better identify the true volatility and reduce the effect of random noise. This results in the situation where we can effectively treat the volatility as fully observed and model it as such (Andersen et al., 2001). In fact, volatility can be observed with infinite precision if stock price follows a true continuous time diffusion. The theoretical foundations for realized variance (and volatility) and how it approximates the integrated variance of a continuous time diffusion were laid out by Andersen et al. in 2001. This work sparked a wide literature on estimating realized volatility and using it in modeling.

Below, I present a similar non-rigorous sketch of how realized volatility relates to continuous time diffusion of asset price as Anderson et al. (2001) presented in their original paper.

The starting point is to assume the following price process:

$$\boldsymbol{dS_t = \sigma_t W_t = r_t} \tag{17}$$

Where S_t is the logarithm of continuous price, μ_t is the drift process and W_t is the standard Weiner process. σ_t is a latent variable of independent stochastic process, which is of the prime interest in the model.

For this diffusion process, the Integrated Variance (IV) for a day is the Ito integral of instantaneous variances over the day.

$$\boldsymbol{IV_t = \int_{t-1}^{t} \sigma_\omega^2 W_\omega} \tag{18}$$

And the integrated volatility is then given by:

$$\boldsymbol{(IV_t)^{1/2} = \sigma_t} \tag{19}$$

The integrated variance for a day can be approximated by sum of intraday squared returns from many short time intervals:

$$\boldsymbol{RV_t = \sum_{i=0}^{T-1} r_{t-i*\Delta}^2} \tag{20}$$

Where: Δ = period length/M, and $r_{t-i*\Delta} = S_{(t-i*\Delta)} - S_{(t-(i+1)*\Delta)}$.

The weekly realized variance, in turn, can be approximated by summing several daily realized variances. By square root we get the weekly realized volatility.

$$\boldsymbol{RVol_t^W = \sqrt{\sum_{i=1}^{T-1} RV_{t-i}}} \tag{21}$$

2.4.1 Difficulties in estimating realized volatility

As noted, if a price process were truly a continuous diffusion, we could reduce the time interval in equation 20 above to be arbitrary small to make the estimate of volatility infinitely precise. However, this is not the case, meaning that realized volatility is still just a (very precise) estimate of the latent volatility. The main issue with reducing the time interval in equation 20 below approximately 1 minute is the presence of market microstructure noise. Microstructure noise tends to cause the realized variance to be significantly overestimated when the sampling frequency is very short. This effect is often observed from volatility signature plots that compare the estimated realized volatility between different sampling frequencies. An excellent treatment of market microstructure noise and its implications for realized volatility can be found from Hansen and Lunde (2006).

Bid-ask bounce is probably the most widely discussed source of market microstructure noise affecting the estimation of realized volatility. Bid-ask bounce is the tendency of transaction prices to bounce between prevailing bid and ask prices when observed in very short intervals. This, in turn, appears to be part of volatility when naively measured, while in reality, the "true" market price remains constant. All of this, in effect, causes negative autocorrelation for the price in very short time intervals and thus the sum of squared return estimator of variance becomes biased, as is the case in other time series settings.

Another major cause of noise is the discrete tick sizes that cause the price to always change by at least the minimum tick size, which may be more than in the absence of a minimum tick size.

In their work on appropriate sampling frequency, Aït-Sahalia et al. (2005) list numerous microstructure effects that can be present: "…bid-ask spread and the corresponding bounces, the differences in trade sizes and the corresponding differences in representativeness of the prices, the different informational content of price changes owing to informational asymmetries of traders, the gradual response of prices to a block trade, the strategic component of the order flow, inventory control effects, the discreteness of price changes in markets that are not decimalized etc." (Aït-Sahalia et al., 2005, p. 355).

In addition to actual microstructure effects in the financial sense, high frequency data, due to its sheer volume, is more prone to errors, which may also exacerbate the errors in realized volatility measurement. The challenges related to this, as well as some good practices related to the data management and cleaning, are presented by Brownlees and Gallo (2006).

2.4.2 Realized volatility models

2.4.2.1 GARCH X

The earliest use of realized volatility/variance measures in forecasting and modeling volatility that I have found is Engle's GARCH X model (2002), which is simply a regular GARCH structure with an added component for realized variance. The problem of this model is that there is no structure defined for the dynamics of RV, making it incomplete and thus impractical for multistep forecasting or for the purpose of simulation. GARCH X model can be represented by the following equation:

$$\boldsymbol{\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_i \sigma_{t-j}^2 + \sum_{k=1}^{s} \gamma_i \, RV_{t-j}} \tag{22}$$

2.4.2.2 Realized GARCH

Realized GARCH model by Hansen, Huang and Shek (2012) corrects the significant shortfall of GARCH X by completing the system of equations while dropping the term for squared return as unnecessary. A notable feature of the model is the flexibility allowed by the measurement function, which enables the model to capture the leverage effect. Also, this feature nests many of the earlier GARCH models as special cases. For instance, regular GARCH is a case where the realized volatility measure is directly the squared return. The model can also be applied in its log-linear form, where all the components are replaced by their respective logarithms. Since the model relies on realized volatility, the problem with (close to) zero returns is eliminated compared to the log-linear configuration of regular GARCH. Realized GARCH model can be represented by the following set of equations:

$$\boldsymbol{r_t = (\sigma_t^2)^{1/2} * W_t} \tag{23}$$

$$\boldsymbol{\sigma_t^2 = + \sum_{i=1}^{p} \alpha_i \sigma_{t-t}^2 + \sum_{j=1}^{q} \beta_j \, RV_{t-j}} \tag{24}$$

$$\boldsymbol{RV_t = m(\sigma_t^2, W_t, \varepsilon_t)} \tag{25}$$

Where: $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$ and m() signifies the measurement function that specifies the structure of the 'error' based on which the integrated variance is estimated by the realized variance measure.

2.4.2.3 Use of general time series models

The advent of realized volatility has also meant that one does not need to necessarily make a distinction between volatility models and time series models in general. This derives from the fact that realized volatility can be treated as fully observed. Thus, any timeseries model specification can be utilized in modeling realized volatility. This opens up the possibility of using a wide range of techniques. For example, Corsi (2009), in the paper where he introduced HAR-RV model, compared its performance to Autoregressive Fractionally Integrated Moving Average (ARFIMA) model, which is not a dedicated volatility model.

2.4.3 HAR-RV model

The volatility model that I have chosen to use in this research is HAR. It was proposed by Corsi (2009), and it extends the traditional ARCH and GARCH models by incorporating the concept of realized volatility. The basic idea of HAR is to model the latent integrated volatility using three historical realized volatility components, Short-

Term (ST), Medium-Term (MT), and Long-Term (LT). This introduces, in a parsimonious way, the empirically observed long persistence while maintaining a simple structure.

$$\boldsymbol{\sigma_{t+1}^{ST} = c + \beta RVol_t^{ST} + \beta RVol_t^{MT} + \beta RVol_t^{LT} + \varepsilon_{t+1}^{ST}} \tag{26}$$

The main advantages that lead to the selection of the model as the base model are:

– HAR model is a parsimonious model that is easy to estimate, which is an advantage since the model needs to be re-estimated for each step of the forecast.
– HAR model can provide more accurate volatility forecasts than traditional ARCH and GARCH models.
– HAR is a realized volatility-based model, which reduces the measurement error problem associated with latent volatility models. Eliminating this source of uncertainty makes it easier to compare competing models.
– HAR model is more flexible than the traditional ARCH and GARCH models, making it very easy to incorporate exogenous variables to the model.
– HAR model can capture the long memory of volatility.

HAR is by no means conclusively the best option for volatility modeling. One of its most notable issues is its lack of asymmetric impact from innovations. Additionally, if the model is augmented with other factors, it is not a complete model, which makes multistep forecasting cumbersome.

An overview of the volatility models and their properties, in terms of the features described in section 2.3.3, is summarized in Table 1. As one can observe, the key advantages of facilitating exogenous variables and relying on realized measures separate HAR from many of the well-known volatility models. However, note that the binary classification of properties into “yes” or “no” is an oversimplification. Also, note that for each of the models, there are numerous further variations and extensions that may address any of the missing properties.

HAR model does not dictate the length of the short-term, medium-term, and long-term components. Because the Google Trends data has a weekly frequency, in the used model, I have chosen to use weekly realized volatility as the short-term component, monthly as the medium-term component, and quarterly as the long-term component.

$$\boldsymbol{\sigma_{t+1}^{W} = c + \beta RVol_t^{W} + \beta RVol_t^{M} + \beta RVol_t^{Q} + \varepsilon_{t+1}^{W}} \tag{27}$$

The research question then centres around whether augmenting this model with search query terms improves the volatility prediction. A more detailed discussion of the methodology employed follows in section 4.

Table 1: Overview of volatility from the perspective of highlighted properties

Model	**Long persistence**	**Mean reverting**	**Asymmetric**	**Exogenous variables**	**Fat Tails**	**Non-latent**
ARCH	No	Yes	No	No	No	No
GARCH	Yes	Yes	No	No	No	No
Integrated GARCH	Yes	No	No	No	No	No
EWMA	Yes	No	No	No	No	No
Taylor-Schwert GARCH	No	Yes	No	No	No	No
Exponential GARCH	No	Yes	Yes	No	No	No
GJR - GARCH	No	Yes	Yes	No	No	No
Asymmetric -GARCH	No	Yes	Yes	No	No	No
Threshold GARCH	No	Yes	Yes	No	No	No
Heterogenous ARCH	Yes	Yes	No	No	No	No
Nonlinear GARCH	No	Yes	Yes	No	No	No
Asymmetric power ARCH	No	Yes	Yes	No	No	No
Fractionally Integrated GARCH	Yes	Yes	Yes	No	No	No
GARCH X	No	Yes	No	No	No	Yes
Realized GARCH	No	Yes	Yes	Yes	No	Yes
HAR-RV	Yes	Yes	No	Yes	No	Yes

Source: Own work.

2.5 Use of Google Trends data in economic modeling

Apart from relying on the volatility modeling history, this work draws cues from literature that concerns using Google Trends data to predict various economic phenomena. More widely, one can include the usage of all crowd-generated online data in economic modeling. In this context, crowd-generated data refers to any data that is created by a very large group of people either directly with the intention of creating data or indirectly by their actions being measured by another party. The classification of such data, types of sources, and modeling approaches have been summarized by Blazquez and Domenech (2018). The most commonly used examples of such data in the literature apart from Google Trends data are data related to Wikipedia views and edits, and data about Tweets.

The earliest usage of such data that I am aware of is the use of search query data to predict unemployment-related statistics by Ettredge et al. (2005). This was still on a time before Google Trends data was available. Google Trends data (then also called Google Insights) was then used in an article by Google's chief economist Hal Varian and Hyunyoung Choi. They initially published their work predicting the present with Google Trends in 2009 in a Google technical report, and later officially in 2012. They also showed that Google Trends data improves predictions related to unemployment claims and auto parts sales. Similarly, unemployment claims data have been modeled with the help of Google Trends data by Askitas and Zimmermann (2009) and Choi and Varian (2009). Related to unemployment, Castelnuovo and Tran (2017) investigate connection of Google search data to proxies of economic uncertainty, among which is also unemployment. They found Google Trends to be a reliable and timely indicator for forecasting.

In all cases, the authors found that Google Trends is useful or showing promise in predicting unemployment claims. Later, Joonas Tuhkuri from the research institute of the Finnish economy (ETLA) published a study that investigated this in a Europe-wide study. The main finding of the study is the following: "Google searches are associated with the unemployment rate in the EU — even after controlling for the country level, lagged, and seasonal effects" (Tuhkuri, 2014, p. 13).

This research was part of the ETLAnow project, and the live forecasts can be found from the project website (https://www.etla.fi/en/etlanow/). This research is especially notable due to the fact that it explicitly deals with Google searches in a multilingual setting in European countries, which is also what my research is aiming to do.

Wu and Brynjolfsson (2009) use Google Trends data for the housing market in the United States. They find that future housing purchases are foreshadowed by Google search volume. Goel et al. (2010) investigate the predictive ability of Google searches in the setting of consumption of games, music, and movies. They find that Google Trends data is useful in prediction but only to a marginal degree. Koop and Onorante (2016) use Google Trends to nowcast several macroeconomic variables and find Google Trends to improve the nowcasting accuracy. Overall, the published research points towards the usefulness of Google Trends as a predictor in various economic settings.

Google Trends data has also been used to investigate issues related to specific asset classes and industries. In the current environment, where the Russian-Ukraine conflict has caused havoc in the European energy markets, the use of Google Trends in the energy market context is particularly interesting. Yu et al. (2019) found that Google Trends data shows promise in predicting oil consumption.

Since 2022, inflation has made a resurgence in the Western world after a long period of relative calm. This has also brought renewed attention to the issue by policymakers. A recent European Commission discussion paper (Buelens, 2023) investigates the public's attention to inflation with the help of Google Trends data. Again, this shows the power of crowd data, as the dreaded inflation spiral is inherently linked to people's inflation expectations, which should be reflected in their interest in inflation-related topics online. An interesting overview of the types of settings where Google Trends data has been applied can be found in Jun et al. (2018). They analyzed 657 papers to understand the research trends around the topic.

2.6 Use of Google Trends and similar data in predicting stock market dynamics

The use of Google Trends data in stock market analysis is largely believed to originate from the analysis of market sentiment. The most influential paper in this field is by Tetlock (2007), who used the frequency of words in media articles to successfully predict market sentiment and the direction of returns.

The first instance that Google Trends was applied to the prediction of stock market dynamics was by Preis et al. (2010). In their paper, they used the search volume of names of S&P 500 index constituent companies and found a linkage to trading volume. Trading volume, in turn, has been linked to volatility by a multitude of authors such as Andersen (1996), making this a very relevant finding for volatility modeling.

The most important research, which also garnered major media attention, is the study by Preis et al. (2013), where the authors put forth a profitable (in back-testing) trading strategy based on the changes in search volume of financially important terms. The authors conclude: “We detect increases in Google search volumes for keywords relating to financial markets before stock market falls. Our results suggest that these warning signs in search volume data could have been exploited in the construction of profitable trading strategies” (Preis et al., 2013, p. 5). This would again indicate that realized volatility would also be predicted by increased search volumes.

The mentioned research has also been a target of criticism from various angles by Challet and Ayed (2014). The most important criticism is that Preis et al. selected the financially significant words (for which Google Trends data is used) by their frequency of usage in financial publications during the testing window. This, in effect, means that one selects ex-post the words that turned out to be most important (thus having also the highest information value), and ex-ante selection of words would have likely been different. Note that since this work shares the same set of keywords, the critique is also partially valid for this research. The difference is that Preis et al. used publication frequency during the 2004-2011 interval, which only partially overlaps with the testing window of this research, while in their case the overlap was for the full period.

Other notable research, especially for this work, are Dimpfl and Jank (2012) and Hamid and Heiden (2015). In both works, the HAR-RV model (or its log version) was augmented with Google Trends components, an approach identical to this work. One of the works found that Google Trends data improved the volatility prediction, and the other found that it did not. The differentiating aspect of their research and this work is that they focused on the search volume of index name and closely related words instead of general financially significant words. Also, both works focused on a single language and index, unlike this work. Thus, one might say that the methodology of this work is a synthesis of the approach used in Preis et al. (2013) and by Dimpfl and Jank (2012) as well as Hamid and Heiden (2015), with an added layer of multilingual approach.

Another more recent example that has direct implications for this work is by Audrino et al. (2020), who used search words relevant to the stock market as proxies for investor attention. What makes their work especially applicable is that they use the same HAR as the benchmark and base model to which Google search information is incorporated. Furthermore, they use realized volatility as the target. They found that: "…attention and sentiment variables are able to improve volatility forecasts significantly, although the magnitudes of the improvements are relatively small from an economic point of view" (Audrino et al., 2020, p. 334).

In addition to the works directly related to this research, Google Trends has been widely investigated in connection with stock market data. For instance, Beer, Hervé and Zouaoui (2012) found that the search frequency of negative search words is heightened during crises, and that investor sentiment, gauged using Google Trends, contributes to predicting short-term market returns. Notably, their research was conducted using French keywords and analyzing the French stock market, which is a relatively rare example of using non-English language keywords. Irresberger et al. (2015) found that the search volume of terms related to the financial crisis, along with the bank name and ticker, helped explain the performance of bank stocks during the financial crisis.

Bank et al. (2011) studied the correlation between Google search volume of a company name and stock dynamics. They concluded that trading volume and return dynamics had a connection to search volume.

Closely related to the use of Google search terms in predicting and explaining stock market dynamics is the use of other crowd-generated online data. Preis et al. (2013) found evidence that Wikipedia site edits of listed companies tend to increase before a major fall in stock prices. Bollen et al. (2011) found that Twitter mood, determined from a sample of tweets using machine learning techniques, does Granger-cause index returns and, in some cases, improve prediction accuracy. Tirunillai and Tellis (2012) found that reviews and other online "chatter" predict the stock price dynamics of individual companies: "...chatter predicts returns and trading volume. The impact of chatter prevails even after controlling for analysts' forecasts, media citations, advertising, and new product announcement" (Tirunillai and Tellis, 2011, p. 33).

An important strand of research is also indirectly related to stock market dynamics. Research in this area evaluates the relationship between Google Trends and market sentiment. This relationship is also closely related to volatility since uncertainty and crisis sentiment are sometimes proxied with stock market volatility metrics like the VIX implied volatility index. An example of such research is Nikkinen and Peltomaki's (2020) investigation of the connection between searches about "market crash" and the VIX.

2.7 COVID-19 and use of Google Trends

The use of online data, such as Google Trends, gained major prominence during the COVID-19 pandemic. The key feature making this data very useful, is its relative timeliness compared to traditional economic statistics. This feature was a very useful source of data in a setting where the COVID-19 and its economic effects appeared relatively suddenly. Before this time, it is also notable that Google Trends data has a rich research history in epidemiology. Most notably, Ginsberg et al. (2009) used Google Trends to track the progress of influenza epidemics across different geographies.

The combination of the need for near real-time data and the proven track record of using Google Trends in an epidemiological context makes Google Trends data a natural candidate in modeling. Among the examples is the work by Papadamou et al. (2020), who found a link between COVID-19-related searches and volatility in a panel setting that included observations from multiple countries around the world.

Another example where Google Trends data was used is from Szczygielski et al. (2021). In their research, the Google search volume was used as a proxy for COVID-19 uncertainty in the process of explaining the influence of COVID-19 uncertainty on stock markets.

The concepts have also been applied to alternative 'assets' such as Bitcoin. Chen et al. (2020) found that: "…market volatility has been exacerbated by fear sentiment as the result of an increase in search interest in coronavirus" (Chen et al., 2020, p. 2).

3 DATA

3.1 Realized volatility data

The research combines two data sources. First one is the realized volatility data of stock indices compiled by the Oxford-Man Institute of Quantitative Finance[2]. The data illustrated in Figure 2 is a series of daily realized volatilities based on a 5-minute returns. This data is widely used by volatility researchers and has already been applied in the context of Google search data by Hamid et al. (2015).

The market indices for which the realized volatility is used, are listed in the Table 2 below. Selection of the markets was based on the availability of index data in the above-mentioned source, aim to limit the research to the European context, and availability of assistance from native speakers in the translation of keywords.

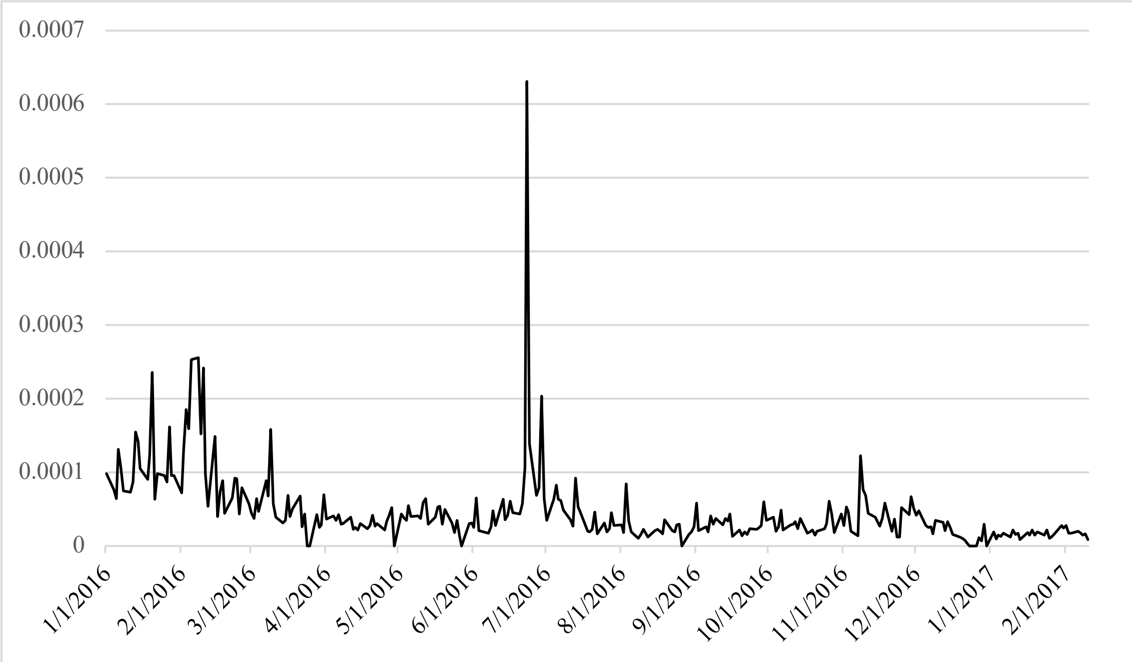
Table 2: Market indices used for the research

Index	Country	Majority language
FTSE 100	United Kingdom	English
DAX	Germany	German
OMXHPI	Finland	Finnish
IBEX	Spain	Spanish
OMXSPI	Sweden	Swedish

Source: Own work.

[2] www.oxford-man.ox.ac.uk

Figure 2: Example of the realized variance data for FTSE 100 index



Source: Own work.

In addition to the market indices, the research uses data about Google search query volumes available through Google Trends. This data runs from January 1, 2004 until the date of the analysis.

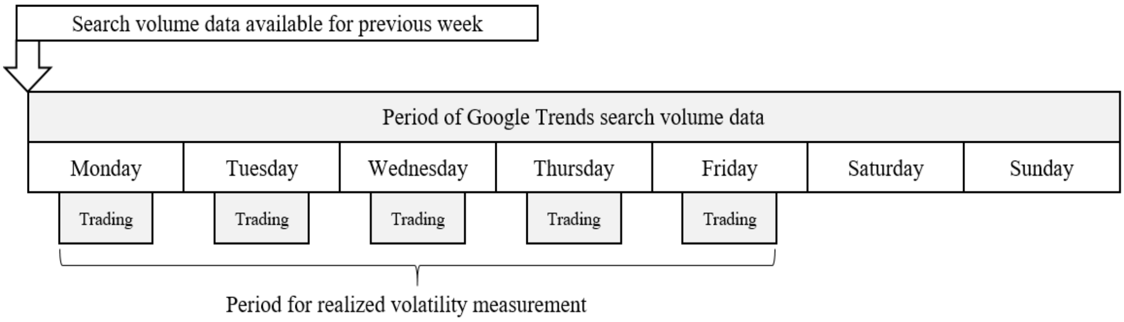
3.2 Google search query data

There are two notable issues that have to be considered when approaching any modeling with Google Trends data. The first issue is that the data is normalized to range from 0 to 100. The highest weekly search volume, compared to total search volume on Google, gets the value 100, and remaining weeks get the value proportional to their relative search volume.

$$\boldsymbol{SQ_{tT}(debt)} = \left(\frac{\boldsymbol{Search\ volume_t\ ("debt")}}{\boldsymbol{Total\ search\ volume_t}} \Big/ \boldsymbol{Max}\left(\frac{\boldsymbol{Search\ volume_T("debt")}}{\boldsymbol{Total\ search\ volume_T}}\right)\right) * \mathbf{100} \tag{28}$$

This implies that we cannot utilize the exact search volume for analysis or compare the levels of different search terms directly. Additionally, there are restrictions on the timing and time interval of the data. The dynamics of this is depicted in Figure 3. Presently, Google releases data for longer intervals solely on a weekly basis, which encompasses all days of the week rather than just from Monday through Friday. Consequently, weekly volatility predictions can only be made on Monday morning before the market opens, and as such maintaining that past information serves as the only input for the forecast.

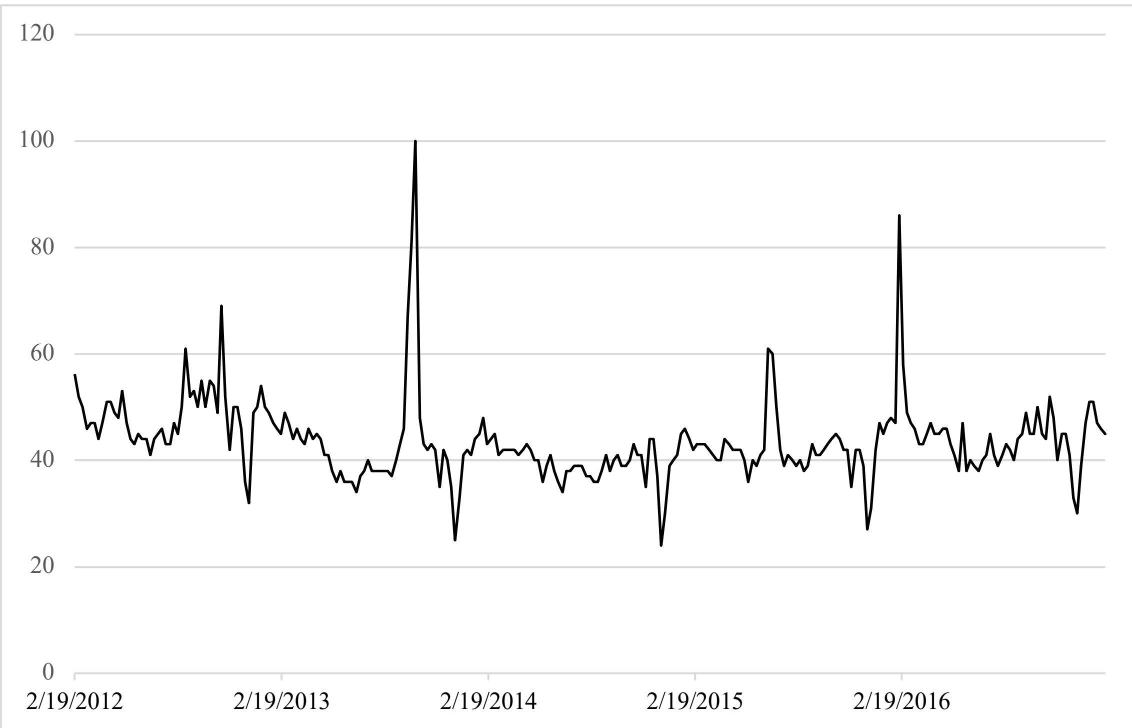
Figure 3: Periodicity of Google Trends and realized volatility data



Source: Own work.

In Figure 4 one can observe an example of raw Google Trends data, from a single download window. The graph clearly illustrates that the data for each separate download has been normalized so that 100 represents the highest value. The practical issues posed by combining data from different downloads are discussed later in section 4.2.

Figure 4: Example of Google search data for the search term "debt" for one download window



Source: Own work.

4 METHODOLOGY

In order to assess whether Google Trends data can serve as a useful and valuable leading indicator for realized volatility, we will compare a Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) introduced by Corsi (2009), to a HAR-RV model extended by a Google search component. The approach is similar to the Dimpfl et al. (2011).

4.1 Selection of search terms and research period

Any discretionary factors by the researcher (me) can be a source of potential bias. Example of this are data fishing or p-hacking. To avoid any possibility of this, the keywords selected are based directly on the research conducted by Preis et al. (2013). The list of words presented in Table 3 is based on their work, where they have developed a measure of a given word's financial relevance. The selected words have the highest financial relevance based on this metric.

In addition, the research period was selected to be as extensive as possible. The period selected corresponds to the period from earliest data available at the Oxford-Mann institute of quantitative finance for the Nordic OMX exchanges (3.10.2005) until the date when the data was downloaded (30.8.2022). This ensured the minimum amount of discretionary factors in the research period selection.

4.1.1 Translation of search terms and other language related issues in Google search data

Table 3: Translated search terms

English	German	Finnish	Spanish	Swedish
hedge	Absicherung	suojaus	cobertura	hedge
dividend	Dividende	osinko	dividendo	utdelning
earnings	Einkünfte	tulot	ganancias	förtjänst
inflation	Inflation	inflaatio	inflación	inflation
markets	Märkte	markkinat	mercados	marknader
bonds	Anleihen	joukkovelkakirja	bonos	obligationer
debt	Schuld	velka	deuda	skuld
gains	Gewinne	voitto	beneficios	vinster
investment	Investition	investointi	inversión	investering

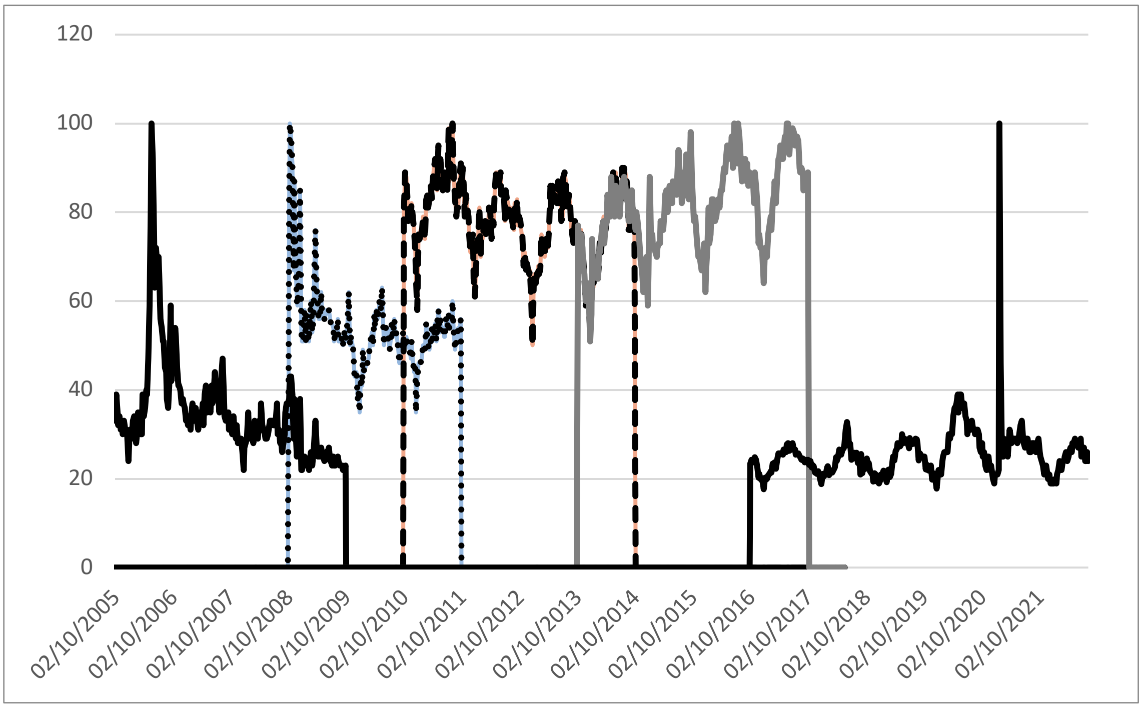
Source: Own work.

4.2 Data preparation

4.2.1 Correcting the inconsistent normalization in Google Trends data

The first thing to note is that if one wants to download data that is on a weekly frequency, one can only download 4 years' worth of data at one time. Additionally, the data has also been normalized so that the highest value in a period (download) gets a value of 100. This means that data from different download instances is inconsistent with data from another download instance, even if the time periods partially overlap. This inconsistency can be observed in the Figure 5 below where differently accented lines originate from different download instances. For example, consider the period from 2/10/2009 to 2/10/2010 where two data sets have dates for the period but the values are significantly different. This issue makes it clear that we cannot use the data without any adjustments.

Figure 5: Example of the inconsistent normalization between different datasets from Google Trends



Source: Own work.

There are several ways to make the data compatible given the level shifts between datasets. One way is to first convert the data to % growth rates and then recursively apply the growth rates to the initial values, thus generating a series that is internally consistent. Another way, which I used, is to compare the values from the overlapping periods and then adjust the dataset that extends further in the future with the ratio of the means of different datasets during the overlapping period. The adjustment ratio for dataset 2 is calculated as:

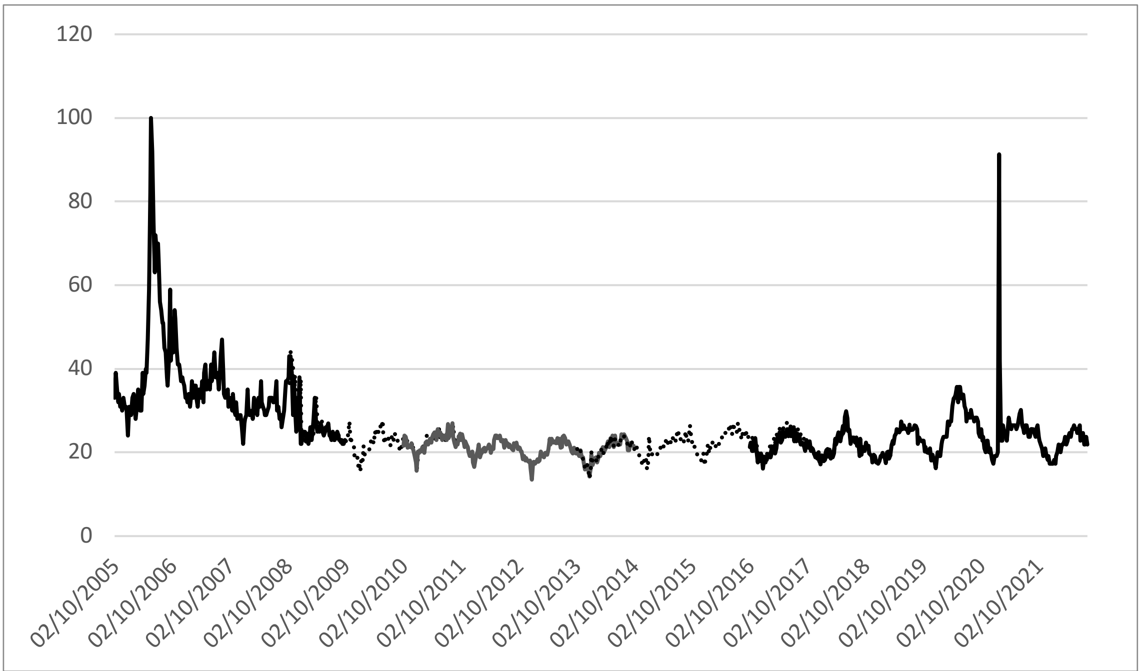
$$Adjustment\ ratio_{2/1} = \frac{\sum_{t=A}^{B} SQ\_dataset2_t}{\sum_{t=A}^{B} SQ\_dataset1_t} \tag{29}$$

Then, each value in dataset 2 (the dataset that has values further to the future) is multiplied by the adjustment ratio:

$$SQ\_dataset2\_adjusted_t = SQ\_dataset2_t * \text{Adjustment ratio}_{2/1} \tag{30}$$

The same issue is present for the period where dataset 2 and dataset 3 overlap, and dataset 3 is not consistent with values from dataset 1 or 2. The adjustment for this is done in an analogous way but by replacing dataset 1 in the formulas with the already adjusted dataset 2, and dataset 3 takes the place of dataset 2. Then, all the subsequent datasets are adjusted in this fashion.

Figure 6: Example of the adjusted datasets resulting from the procedure described above



Source: Own work.

As one can observe from Figure 6, the procedure results in datasets that are relatively continuous in the overlapping period, which was the purpose of the whole adjustment. However, I must note that in the overlapping periods, two datasets never match perfectly, and there is no direct remedy for this. While some of the inconsistency can be attributed to rounding (data is rounded to have no decimals), in certain instances the differences are much larger and thus cannot be attributed to rounding. Furthermore, the discrepancy cannot be attributed to inappropriate normalizing or scaling since there are instances where the data for week A is notably[3] higher than for week B in a given dataset, while in another dataset covering the same period, the order is reversed, proving that a difference in a simple linear transformation cannot be responsible for the error. Since Google has not been too forthcoming about the details of how Google Trends data is collected or generated publicly, I had limited possibilities to correct the inconsistency without the knowledge of the cause. Thus, the inconsistency between datasets is largely ignored and is addressed only by constructing the final dataset as a simple average of two values for the periods where two values are available. The final adjustment that is done is performed by dividing all values in a dataset by its maximum value and multiplying all values by 100. This ensures that we get a dataset that again has the property of having 100 as the largest value (as in the original data), regardless of whether the highest value for the period occurs in the earliest or most recent dataset.

4.2.2 Seasonality in Google Trends data

While doing exploratory analysis for the Google Trends data I noticed that for many search terms one can identify a pronounced seasonality. For an example, see the Figure 5 between years 2009 and 2017 which clearly exhibits a regular seasonality. The search word in question is 'hedge' for which it is hard, in financial context to find a reason for seasonality. But as it turns out in British English 'hedge' is more regularly used to describe a 'bush fence' which explains why in certain times of year the search frequency would increase.

Clear seasonality was also observed in the search frequency for the keyword 'earnings', likely corresponding to quarterly earnings seasons. Similarly, most keywords exhibited a pronounced dip during holiday seasons, possibly due to their frequent use mainly in professional contexts. Although one could attempt to remove some of the seasonality to better isolate the relevant signal, in this study I have opted to leave the seasonality unadjusted since the factors causing it may also be relevant for volatility forecasting.

4.2.3 Observations related to financially related search terms

During the preliminary analysis of the data, I have also made other observations that may or may not be relevant to the results of this analysis or that may help in the future research.

[3] By "notable" I mean differences that are not small enough to be attributed to rounding errors. Although there are instances where the discrepancies between datasets are clearly not due to rounding errors, I have not found serious enough instances of this kind of inconsistency that would be material for the analysis at hand.

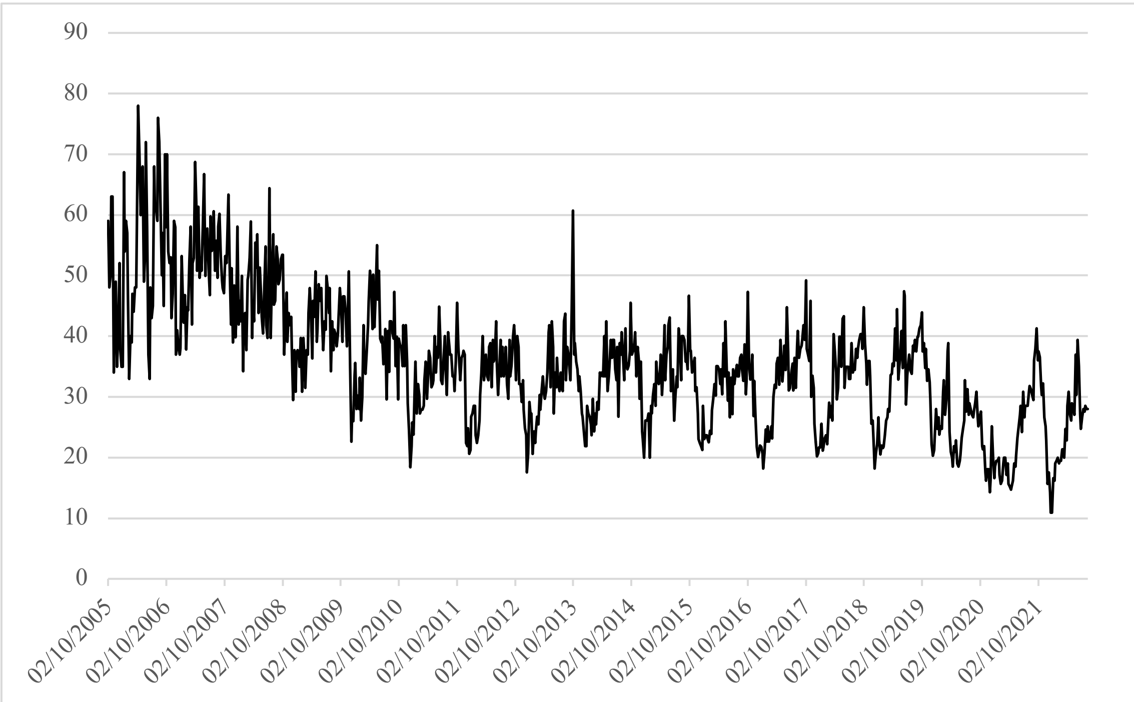
4.2.3.1 Common use of English over the native language

Another key issue impacting the use of keywords from another languages is the widespread use of English as a lingua franca in the Western world, especially in professional settings such as finance. This might be a contribution factor for the observation that data from languages with fewer speakers is exceedingly noisy, especially in the earlier years.

4.2.3.2 Evolution in the characteristics of Google Trends data

There is a clear trend towards smoother data over the years. My hypothesis here is that this trend reflects the increasing adoption of the internet by the wider population. This likely also makes the signal estimation from earlier periods less applicable in the current setting. This dynamic is illustrated in Figure 7.

Figure 7: Example of Google search data where the decreasing noise is visible



Source: Own work.

4.2.3.3 Usage of synonyms and declensions

Another issue that might hinder the use of certain languages is the multitude of alternate forms for the same word. For instance, in the Finish language, the word "store" (shop) has well over two thousand valid declensions, and a similar number of alternate forms can be generated for most nouns (http://www.ling.helsinki.fi/~fkarlsso/genkau2.html). At the same time, Google Trends takes into account only exact matches of the specific word being part of an expression. For instance, searches for "grocery store" are counted as searching the word “store” while searching for “stores” are not. This results in greatly diminished frequency for any expression with multiple commonly used declensions and thus the influence of random noise increases, which likely adversely affects the results.

4.2.4 Matching and aggregating volatility data to the Google Trends data

As already mentioned, the Google Trends data is published with weekly frequency, corresponding always to a Monday to Sunday time horizon, regardless of whether the time period contains a holiday or a change in the year. On the other hand, the Realized Volatility (RV) data corresponds to trading days. Firstly, I ensured that the RV data was correctly matched to the Google Trends data, so that all the trading dates that were after the previous week's Google Trends data publishing date and before the current week's publishing date were matched with the Google Trends data for the current week. Afterwards, all the RV values were averaged using a simple arithmetic average.

Following Corsi (2009), realized volatility can be averaged to a longer time horizon with a simple average, with insignificant convexity error.

$$\mathrm{RVol}_{\mathrm{t}}^{\mathrm{W}} = \frac{1}{4}(\mathrm{RVol}_{\mathrm{t}}^{\mathrm{d}} + \mathrm{RVol}_{\mathrm{t-1}}^{\mathrm{d}} + \mathrm{RVol}_{\mathrm{t-2}}^{\mathrm{d}} + \mathrm{RVol}_{\mathrm{t-3}}^{\mathrm{d}}) \tag{31}$$

The model also includes realized volatility that is a longer-term average (monthly, quarterly), and in these cases, the terms were arrived at by further averaging the weekly components using a simple arithmetic mean. The monthly period included average of 4 lagged values, whereas the quarterly component included 12. Note that averaging did not, in any way, take into account that certain weeks had fewer trading days than others. This approach was adopted purely for analytical convenience.

4.3 Modeling specification and parameter estimation technique

The models used are based on the HAR-RV models specified in equation (32) and the HAR model estimated with the natural logarithm of realized volatility HAR-ln(RV) specified in equation (33).

$$RVol_{t+1}^{W} = c + \beta_1 RVol_t^{W} + \beta_2 RVol_t^{M} + \beta_3 RVol_t^{Q} + \varepsilon_{t+1}^{W} \tag{32}$$

$$ln(RVol_{t+1}^{W}) = c + \beta_1 ln(RVol_t^{W}) + \beta_2 ln(RVol_t^{M}) + \beta_3 ln(RVol_t^{Q}) + \varepsilon_{t+1}^{W} \tag{33}$$

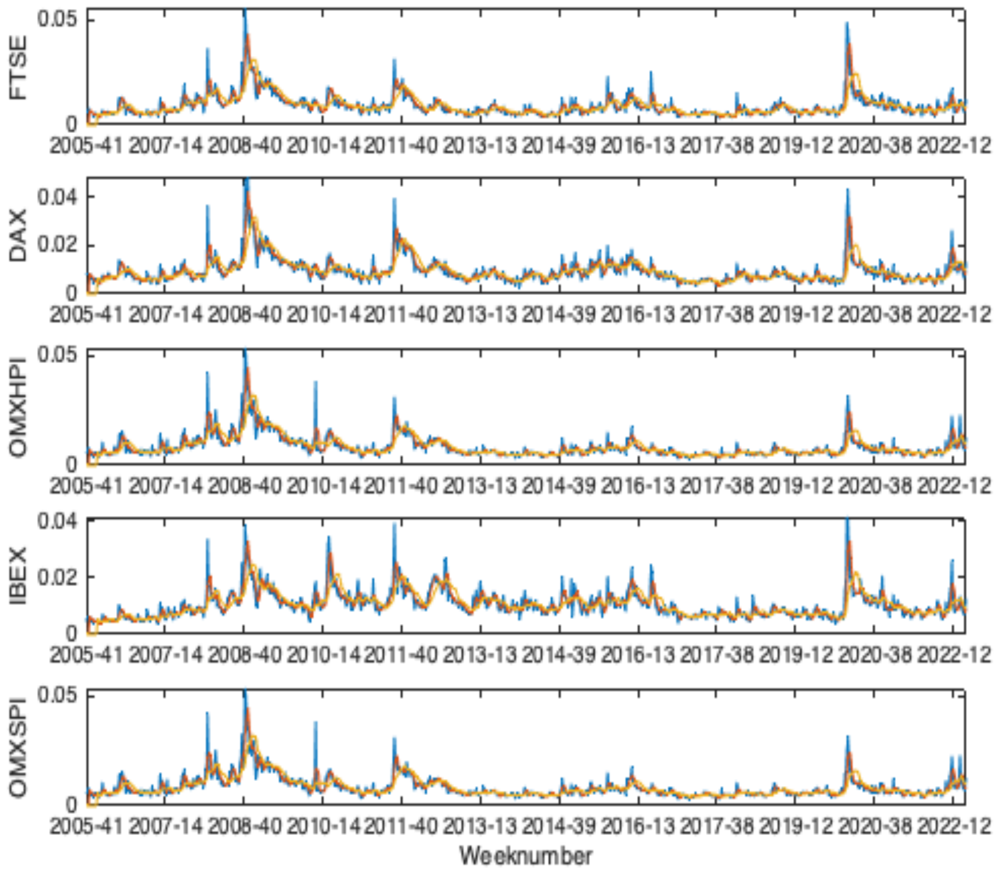
These two models are then augmented (extended) with a single Google Trends component. Here it is important to note that the model could be extended with 3 Google Trends search volume components formed similarly as the 3 volatility components in HAR. This was done to avoid possible overfitting. Note that the model structure was selected prior to any data analysis, and thus there is a valid concern to expect that the ‘signal’ from Google Trends data is very weak, and thus having multiple correlated search parameters may not provide any additional benefit.

$$RVol_{t+1}^{W} = c + \beta_1 RVol_t^{W} + \beta_2 RVol_t^{M} + \beta_3 RVol_t^{Q} + \beta_4 SQ_t^{W} + \varepsilon_{t+1}^{W} \tag{34}$$

$$ln(RVol_{t+1}^{W}) = c + \beta_1 ln(RVol_t^{W}) + \beta_2 ln(RVol_t^{M}) + \beta_3 ln(RVol_t^{Q}) + \beta_4 ln(SQ_t^{W}) + \varepsilon_{t+1}^{W} \tag{35}$$

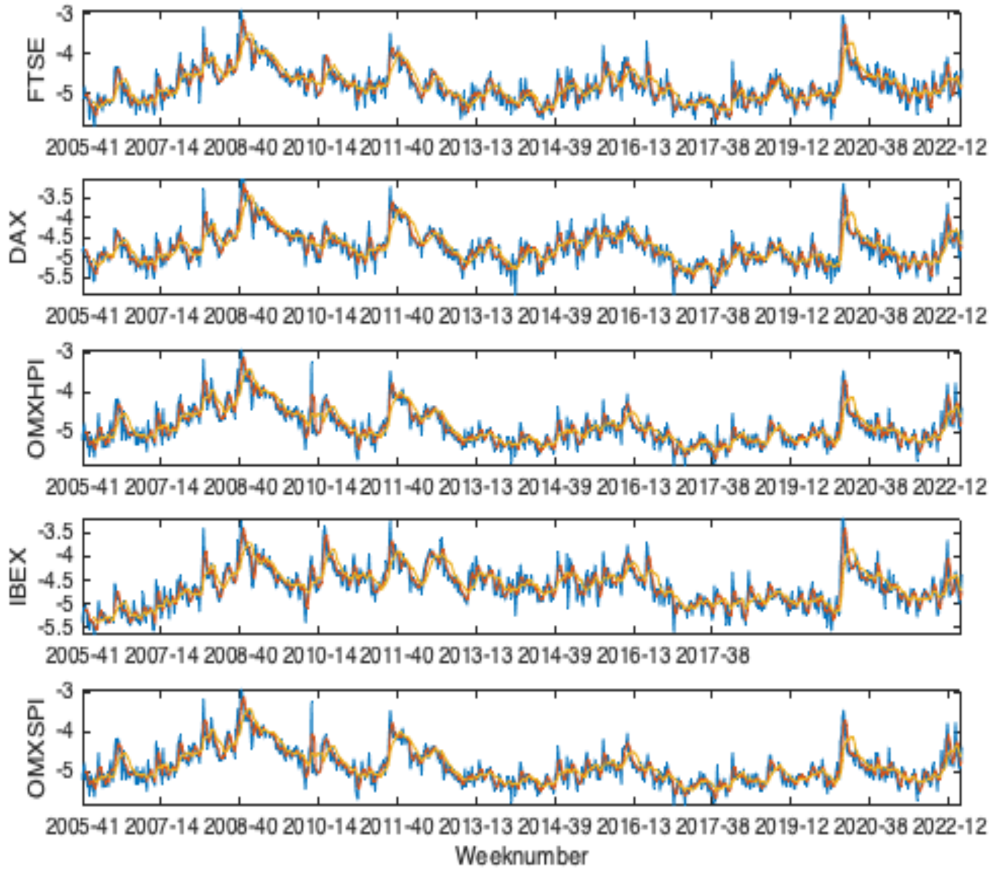
Here, the search query terms are specified as simple period averages, similar to RV terms. Below one can see the values of short-term, medium-term, and long-term realized volatility components throughout the period. The dynamics of each component for all markets are depicted in Figure 8 and Figure 9.

Figure 8: Dynamics of HAR-RV volatility components



Source: Own work.

Figure 9: Dynamics of the logarithm of HAR-RV volatility components



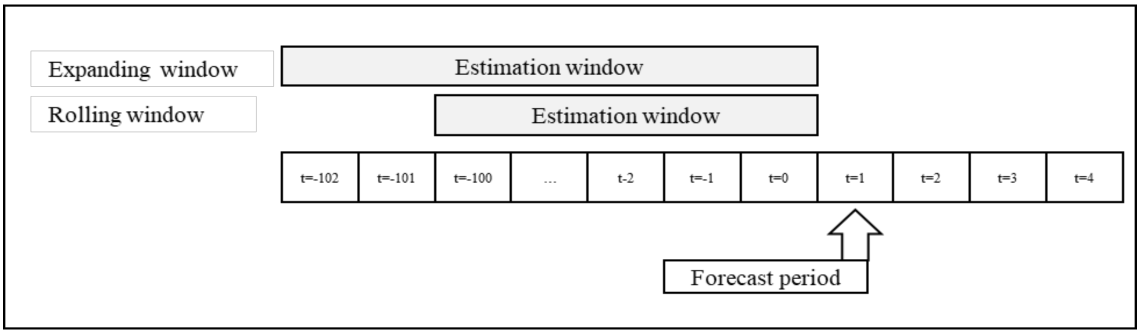
Source: Own work.

4.3.1 Estimation window and withholding period

Another important factor to consider is the estimation window for the models. In this research, the estimation window was set to 100 days solely to include the financial crisis as part of the (pseudo) out-of-sample evaluation window. However, the 'optimality' of the estimation window length was not taken into consideration, which would be sensible according to Inoue et al. (2017) if the aim would be to optimize the model performance. The process of optimizing the estimation window was omitted, because the primary aim of the research was not to perform optimal volatility forecasts, but rather to evaluate whether Google search volume contributes to the forecast.

The two different methods of estimation depicted in Figure 10 are meant to ensure that the results are representative whether there is parameter instability present or not (Inoue et al., 2017).

Figure 10: Schematic representation of the difference between 'expanding window' and 'rolling window' estimation.



Source: Own work.

4.3.2 Parameter estimation

All the models were estimated using simple Ordinary Least Squares (OLS) method. Parameters were estimated with the help of Matlab 2015b software.

5 EMPIRICAL RESULTS

5.1 Performance measures

The evaluation of whether the Google Trends data improves the prediction is based on two criteria. The first criterion is the out-of-sample performance, which is measured by:

- Mean Squared Error (MSE),
- Mean Absolute Error (MAE),
- Mean Absolute Percentage Error (MAPE).

For these performance measures, lower values indicate more precise forecasts.

The second criterion is the in-sample and out-of-sample R^2 in the Mincer-Zarnowitz (1969) regressions of the actual realized volatilities to the predicted integrated volatility. This approach has been selected following the examples of Andersen et al. (2003).

5.1.1 Mean Squared Error

MSE is a common metric used to measure the difference between a model's predicted values and the actual values. MSE performance measure is defined as follows:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(RVol_t - F_t)^2 \quad (36)$$

where:
- **RVoln** stands for Realized Volatility
- **Fn** stands for Forecast of Realized Volatility
- **t** stands for indicator of time

Properties of MSE:

- MSE is a non-negative value, which means it is always greater than or equal to zero.
- MSE is sensitive to outliers in the data, which means that extreme values can significantly affect the overall measure.

5.1.2 Mean Absolute Error

MAE is a statistical metric used to measure the average magnitude of the errors between the actual and predicted values in a set of data. It is calculated by taking the average of the absolute differences between the actual values and the predicted values. MAE performance measure is defined as follows:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|RVol_t - F_t| \tag{37}$$

Properties of MAE:

- MAE is always a non-negative value. This is because the absolute value of the difference between actual and predicted values is always non-negative, which makes relative comparison of different models easier.
- It is a simpler and more intuitive measure of error than other statistical metrics, such as MSE, because it is based on the absolute error rather than the squared error.
- MAE is robust to outliers, as it does not heavily penalize large errors like MSE does.

Overall, MAE is a useful statistic for evaluating the accuracy of predictions, particularly in situations where outliers may be present in the data.

5.1.3 Mean Absolute Percentage Error

MAPE is a statistical metric used to measure the accuracy of a forecast or prediction. It is calculated by taking the average of the absolute percentage difference between the actual and predicted values. MAPE performance measure is defined as follows:

$$MAPE = \frac{T \cdot MAD}{\sum_{t=1}^{T} RVol_t} \cdot 100 \tag{38}$$

Properties of MAPE:

- MAPE is always a non-negative value. This is because the absolute value of the percentage difference between actual and predicted values is always non-negative.
- It can be a more intuitive measure of error than other statistical metrics because it is based on a relative size of an error rather than absolute size of an error.
- MAPE is sensitive to large errors or outliers, as these can have a significant impact on the percentage difference between the actual and predicted values.

It is also important to note that MAPE has some limitations, particularly when the actual values are close to zero, as small errors in the predicted values can result in very large percentage differences.

5.1.4 Mincer-Zarnowitz regression

The Mincer-Zarnowitz regression is a statistical method used to evaluate the predictive power of economic forecasting models. It was first proposed by Jacob Mincer and Victor Zarnowitz in their paper, "The Evaluation of Economic Forecasts" (1969).

The Mincer-Zarnowitz regression involves running a linear regression of the forecast errors (the difference between the actual values and the predicted values) on the forecasted values themselves. The Mincer-Zarnowitz regression is specified as follows:

$$RVol_{t+1}^{W} = c + \beta F_{t+1}^{W} + \varepsilon_{t+1} \tag{39}$$

where:

- **c** and $\boldsymbol{\beta}$ stand for the coefficients to be estimated
- $\boldsymbol{\varepsilon}$ stands for the error term

This model can be used to estimate a wider range of forecast properties, but for this study I am only using the R^2 of the regression as a comprehensive summary measure of forecasting performance.

5.1.5 Statistical significance of the results

Note that all predictions are for data that was not part of the estimation sample, and no steps were taken to specifically optimize the results for the sample. If the Google search data did not have any predictive value, one would expect the share of models where the performance improves or deteriorates to be equal. Thus, the confidence level of whether or not the inclusion of Google search data improves prediction could be approximated using the binomial distribution.

5.2 Results

The results in this section are presented in relation to the base model which, does not include the Google Trends component as a predictor. It is important to note that each language and model configuration has a distinct base model against which the Google Trends augmented configuration is benchmarked. A list of distinct base model variations is presented in Table 4.

The results are presented in this manner to facilitate comparison between languages and models. For each of the models, the threshold of 100% represents a split between worse or better performance than the base model. For MAD, MPA, and MSE, a result lower than 100% indicates better performance than the base model, while for MZ result higher than 100% indicates better performance.

Table 4: List of model variations

	Base model	Calibration data
Model 1	HAR-RV	All the historical data
Model 2	HAR-RV	100 previous time periods
Model 3	HAR-ln(RV)	All the historical data
Model 3	HAR-ln(RV)	100 previous time periods

Source: Own work.

5.3 Overview

Overall, as part of the evaluation, a total of 180 model variations were estimated, which resulted from the inclusion of 9 search words in 5 markets across 4 model variants. The performance comparison is presented in Tables 5, 6, 7, 8, 9. The first way to evaluate if the inclusion of Google search data improved the prediction accuracy is to count the models where the result improved. There are several key observations that can be made based on the results of the analysis:

Firstly, the inclusion of search data improves the accuracy across all performance measures. This is a consistent finding throughout the study. Secondly, the improvement can be observed in each model variant. Regardless of the specific model configuration used, the inclusion of search data leads to improved accuracy. Finally, the share of models where improved performance is observed is clearly above 50%. In other words, in more than half of the models examined, the inclusion of search data leads to improved accuracy.

Table 5: Share of models split to improved and deteriorated prediction by performance measure

Measure	Deteriorated	Improved
MAD	28%	72%
MAPE	34%	66%
MSE	12%	88%
MZ	9%	91%
Average	**21%**	**79%**

Source: Own work.

Table 6: Count of models split into improved and deteriorated prediction performance measured by MAD

MAD	**>=100%**	**<100%**
HaR-RV - Moving window	12	33
HaR-RV - Expanding window	13	32
HAR-ln(RV) - Moving window	13	32
HAR-ln(RV) - Expanding window	13	32
Total	**51**	**129**

Source: Own work.

Table 7: Count of models split into improved and deteriorated prediction performance measured by MAPE

MAPE	**>=100%**	**<100%**
HaR-RV - Moving window	19	26
HaR-RV - Expanding window	12	33
HAR-ln(RV) - Moving window	15	30
HAR-ln(RV) - Expanding window	16	29
Total	**62**	**118**

Source: Own work.

Table 8: Count of models split into improved and deteriorated prediction performance measured by MSE

MSE	**>=100%**	**<100%**
HaR-RV - Moving window	2	43
HaR-RV - Expanding window	1	44
HAR-ln(RV) - Moving window	7	38
HAR-ln(RV) - Expanding window	11	34
Total	**21**	**159**

Source: Own work.

Table 9: Count of models split into improved and deteriorated prediction performance measured by MZ

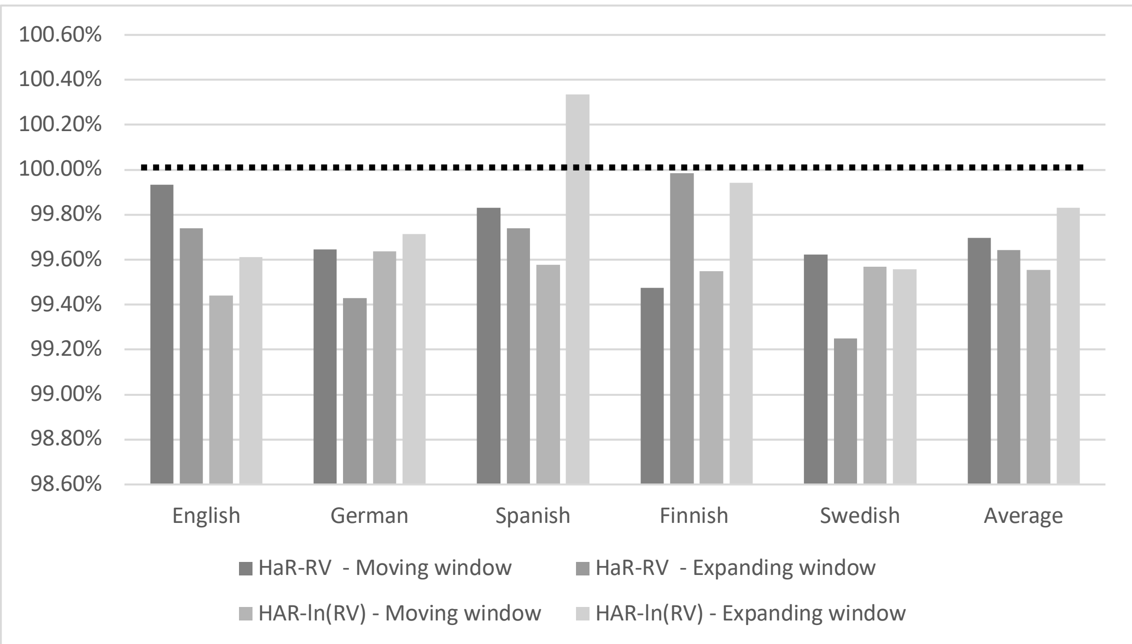
MSE	**>=100%**	**<100%**
HaR-RV - Moving window	43	2
HaR-RV - Expanding window	44	1
HAR-ln(RV) - Moving window	37	8
HAR-ln(RV) - Expanding window	39	6
Total	**163**	**17**

Source: Own work.

5.4 Results split by language

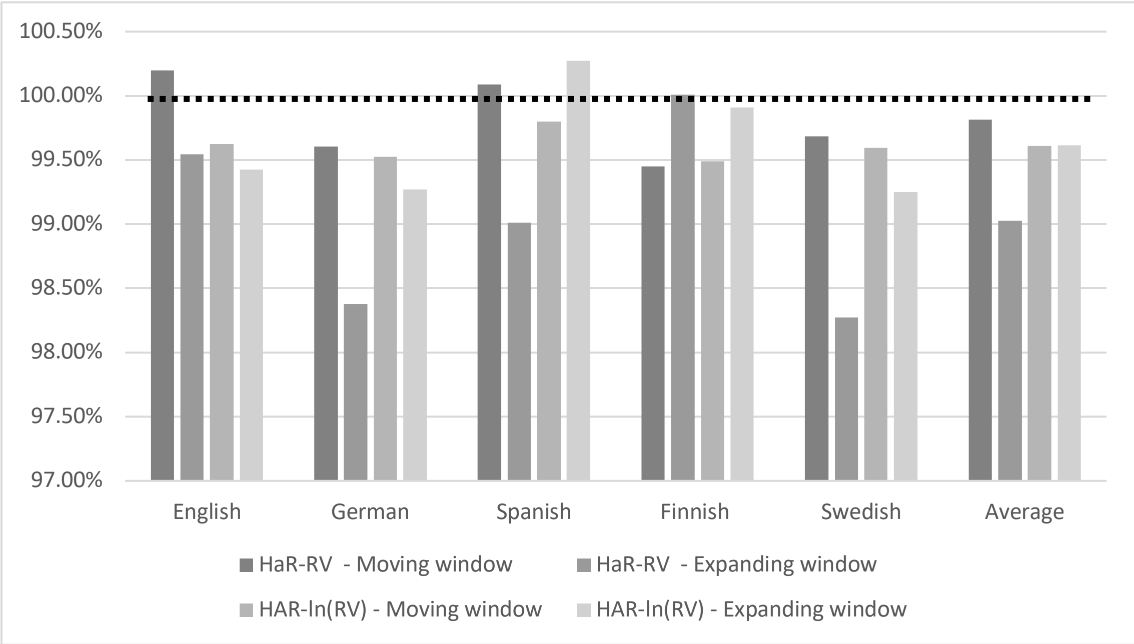
For each of the languages, the average performance compared to the benchmark is calculated for each of the model variants. This is done separately for each of the performance measures. The performance results are depicted in Figures 11, 12, 13, 14. Overall, the results indicate consistent improvement in prediction accuracy across all languages and model configurations and performance measures. No language, measure or model configuration stands out in any way even when the results are viewed based on the language split. Still, one must note the scale of the improvement is very minor in each of the measures.

Figure 11: Average MAD per language for various model configurations



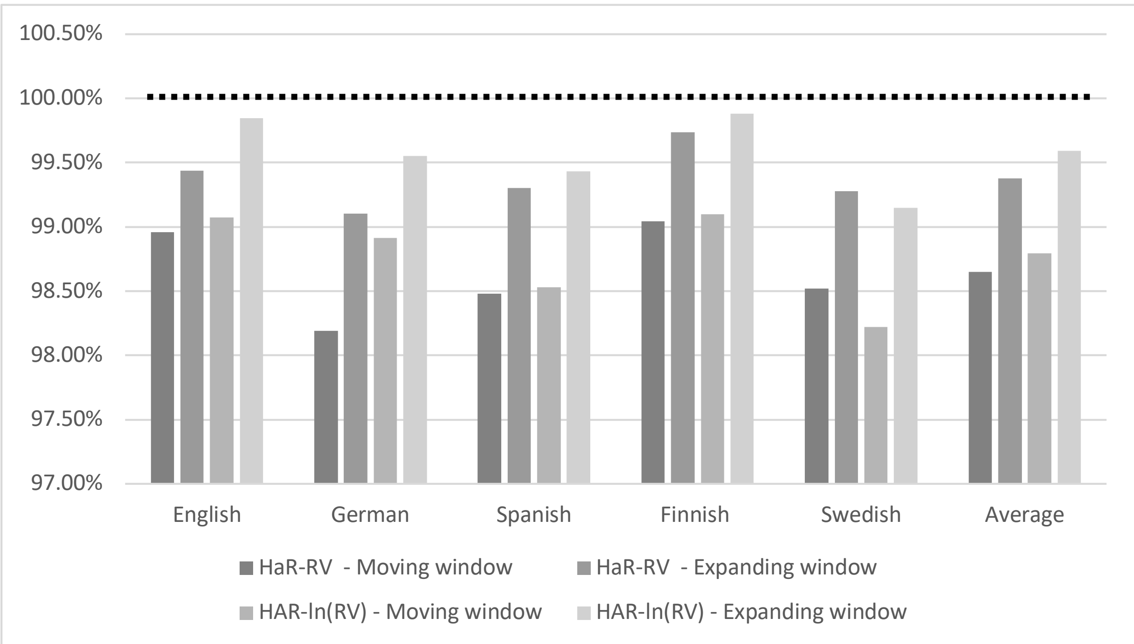
Source: Own work.

Figure 12: Average MAPE per language for various model configurations



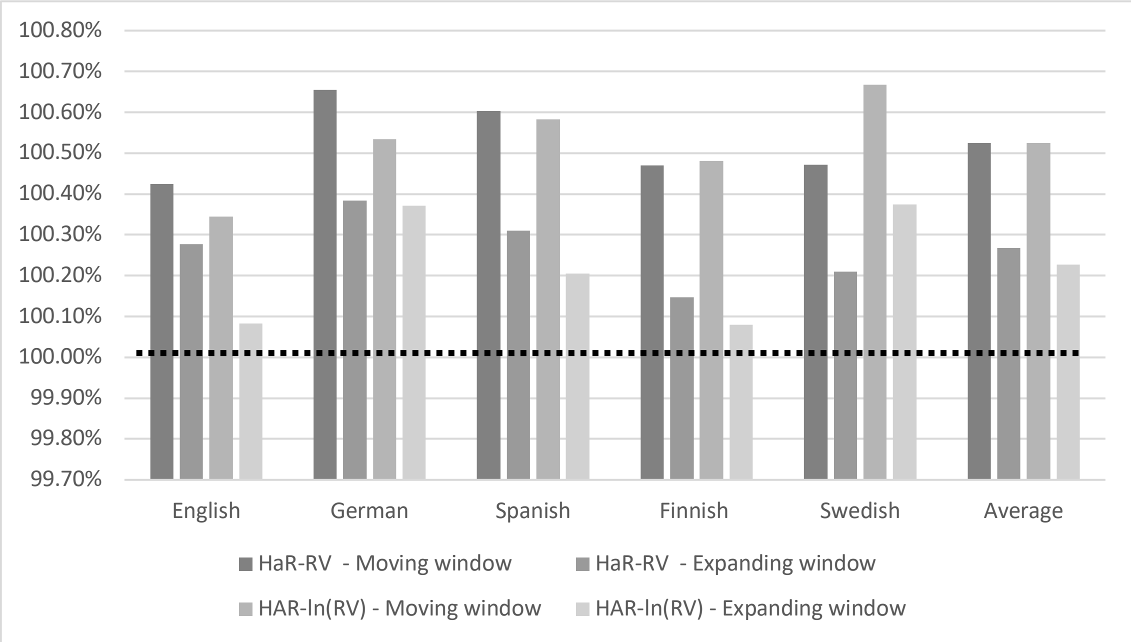
Source: Own work.

Figure 13: Average MSE per language for various model configurations



Source: Own work.

Figure 14: Average MZ per language for various model configurations

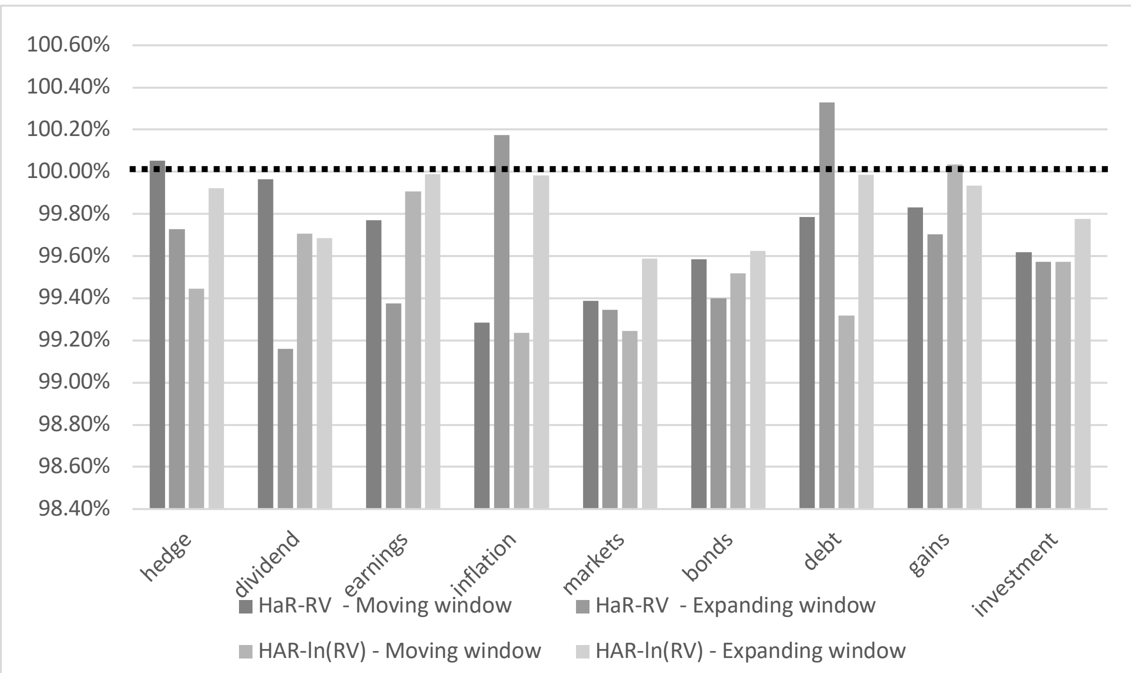


Source: Own work.

5.5 Results split by search term

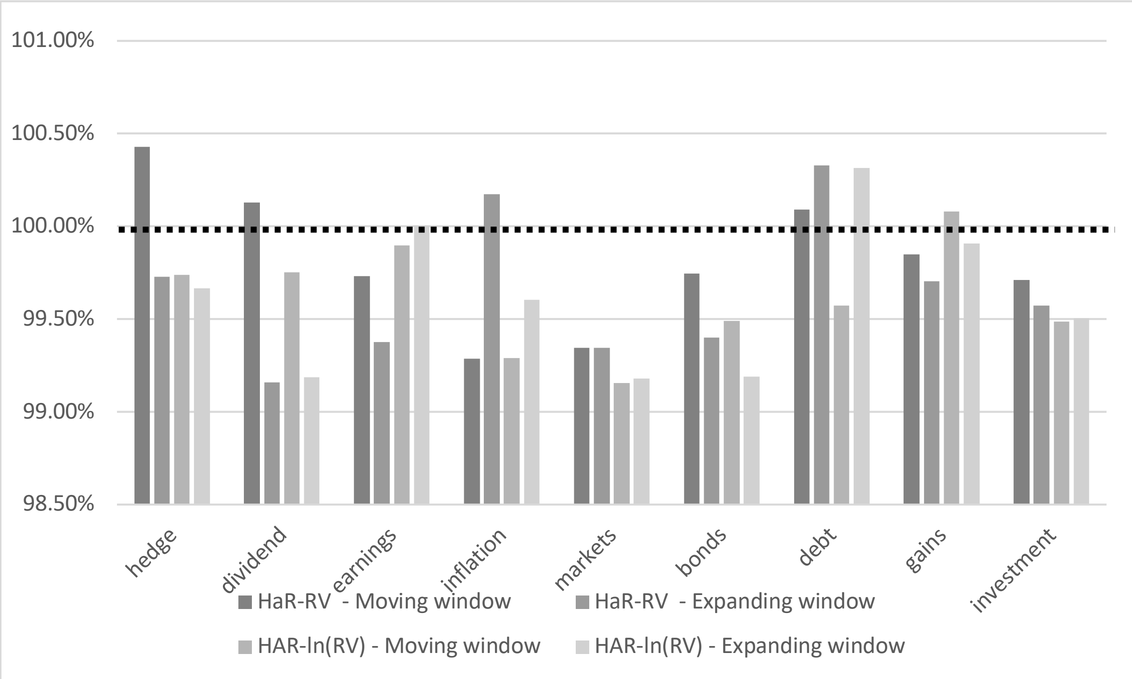
When split by search term and averaged across languages, the results still indicate a consistent improvement in predictive accuracy. Again, no search word or model specification stands out as an outlier. The results split by language are depicted in Figures 15, 16, 17, 18.

Figure 15: Average MAD per model specification



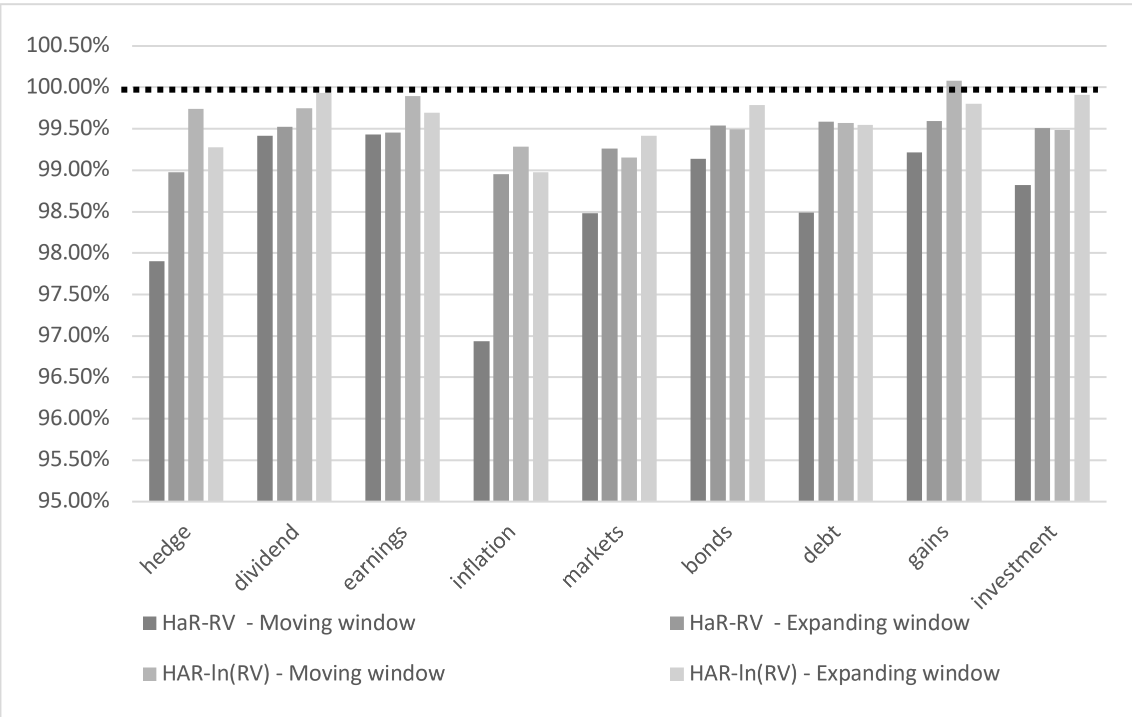
Source: Own work.

Figure 16: Average MAPE per model specification



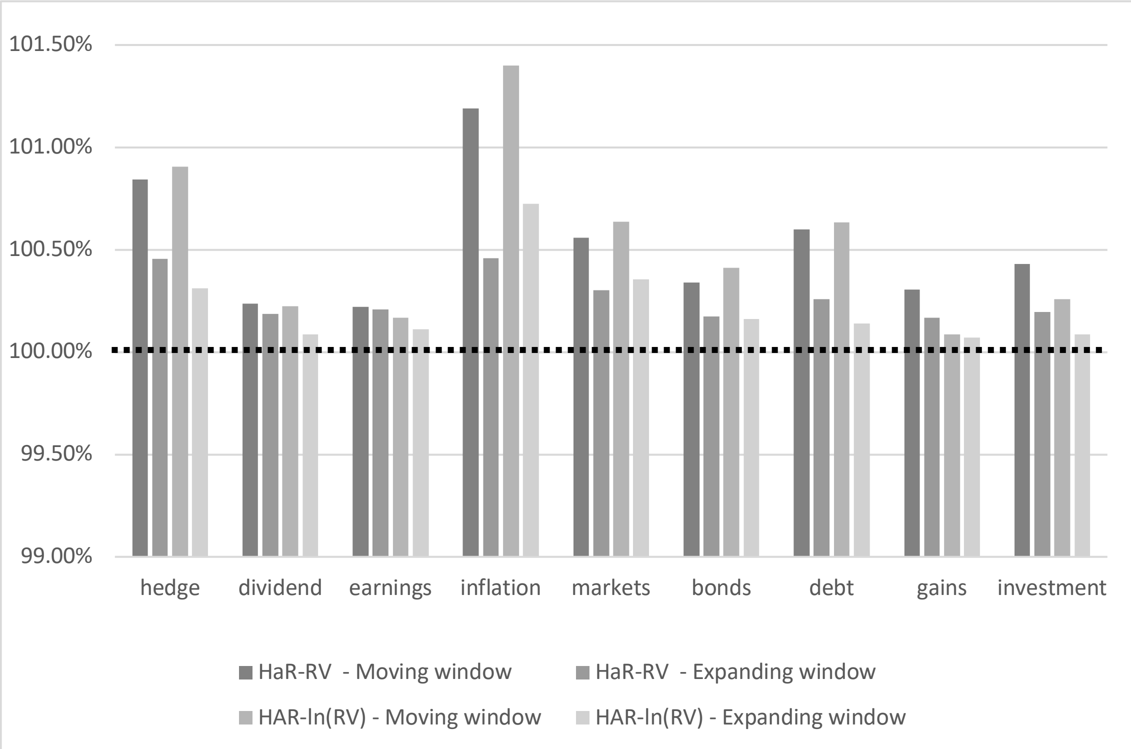
Source: Own work.

Figure 17: Average MSE per model specification



Source: Own work.

Figure 18: Average MZ per model specification



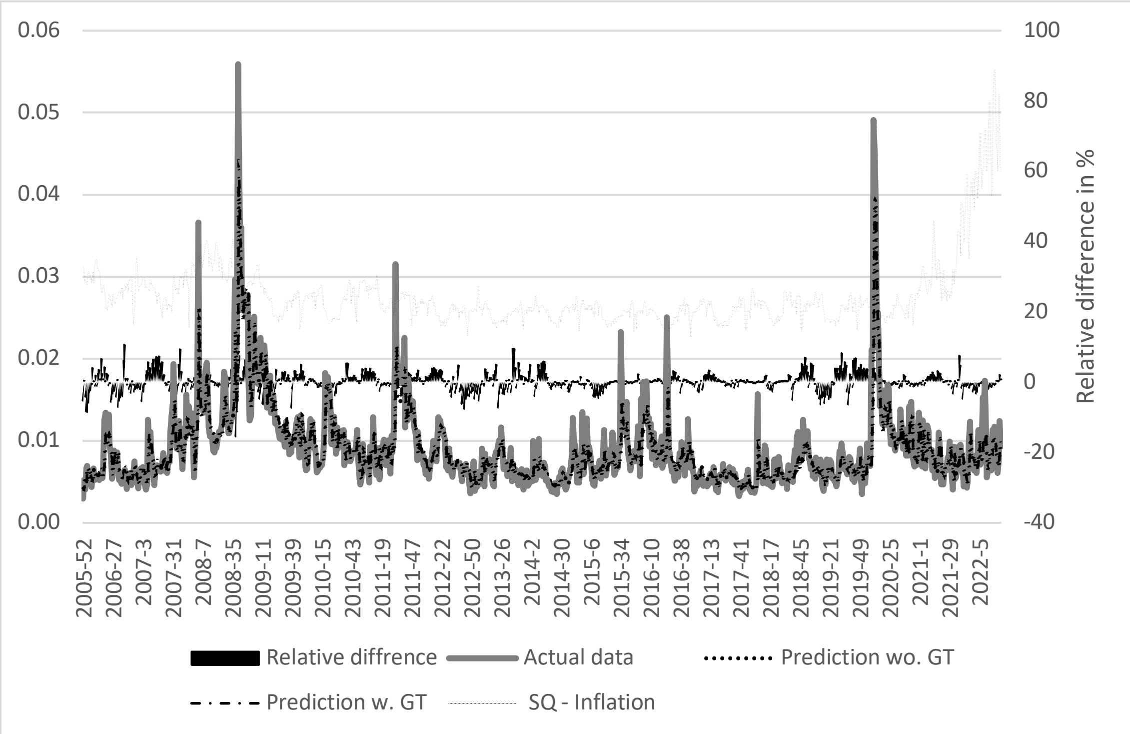
Source: Own work.

5.6 Discussion

Overall, the results indicate that Google search data of financially significant words has the potential to enhance volatility predictions. Still, this conclusion comes with major caveats. The main caveat is whether or not the improvement in prediction accuracy is sufficient to be of any practical significance. In all performance measures, the prediction accuracy improvements were in the range of 1%, and no language, model specification, or keyword stood out significantly from this pattern.

Figure 19 shows the difference between the prediction from a model without Google Trends keyword, and prediction from a model augmented with the Google Trends keyword (Moving window HAR-RV). Additionally, the relative difference of the Google Trends augmented model prediction from the model without it is shown as bars. One can observe there are mostly small differences between the model results compared to overall variability of the RV over time. Still, in relative terms, the predictions are frequently different by up to 5%, which may be significant enough difference in some applications to warrant the inclusion of the Google Trends component.

Figure 19: Illustration of the contribution of Google Trends component



Source: Own work.

To contrast the level of difference between the non-search data augmented models and the base model, I have also compared the difference between different base model configurations. The difference from base performance, which is defined as average performance across model specifications for single market, is calculated for each market and model specification combination. This is done for each performance statistic separately. The results presented in Table 10 indicate that larger performance improvements can be achieved by slight changes in the model specification rather than from the search word augmentation.

Table 10: Comparison of average performance metrics across base model configurations.

Model	MAD	MAPE	MSE	1/MZ	Average
HaR-RV - Moving window	101.50%	101.50%	92.67%	97.20%	98.21%
HaR-RV - Expanding window	105.19%	105.19%	100.28%	100.32%	102.75%
HAR-ln(RV) - Moving window	95.01%	95.01%	101.65%	100.33%	98.00%
HAR-ln(RV) - Expanding window	98.30%	98.30%	105.40%	102.29%	101.07%

Source: Own work.

6 CONCLUSION

Based on the results, the research question has not been conclusively resolved. On one hand, the inclusion of Google search components in the prediction has a consistent effect of improving forecast accuracy across all performance measures. The effect can be observed in context of different languages, model configurations, markets, as well as search terms. On the other hand, the improvements in prediction accuracy are relatively minor and of questionable practical significance.

Still, the analysis presented here was primarily aimed at providing an initial validation of the potential usefulness of Google Trends data in volatility estimation. In this regard, there are promising indications. This conclusion derives from the fact that the investigation presented here purposefully used naïve approach in terms of factor selection, treatment of factors, and statistical specification of models. With a more rigorous optimization of all aspects of modeling, such as only selecting the most promising keywords and model specifications, it is possible that the level of performance gains from including Google search factors in the model, could be plausibly increased.

This work has also brought attention to several important considerations for using Google search terms in modeling. Issues such as accounting for synonyms, seasonality, and declensions in different languages are all potential areas for optimization. Moreover, the appropriate treatment of Google search data is not yet well understood. Further investigation is warranted to determine whether removing seasonality, smoothing, or removing outliers would significantly improve the model quality.

The operational hypothesis was specifically directed towards short-term volatility predictions, and as such, its applicability to longer-term multi-step forecasts remains unclear. It is therefore necessary to conduct further investigations to determine whether the model augmentation with search data would prove to be more beneficial for longer-term predictions.

REFERENCES

1. Ait-Sahalia, Y., Mykland, P. A., & Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *The review of financial studies*, *18*(2), 351-416.

2. Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, *51*(1), 169-204.

3. Andersen, T. G., Bollerslev, T., (1997). Answering the critics: Yes, ARCH models do provide good volatility forecasts *Working paper 6023*, National bureau of economic research.

4. Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association, 96*(453), 42-55.

5. Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica, 71*(2), 579-625.

6. Andersen, T. G., Bollerslev, T., & Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities*.* *Econometrica, 73*(1), 279-296.

7. Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical finance, 9*(3), 203-228.

8. Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly, 55*(2), 107-120.

9. Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting, 36*(2), 334-357.

10. Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics, 74*(1), 3-30.

11. Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial markets and portfolio management, 25*(3), 239.

12. Beer, F., Hervé, F., & Zouaoui, M. (2012). Is big brother watching us? Google, investor sentiment and the stock market. *Economics Bulletin, Forthcoming*.

13. Black, F. (1976): Studies in stock price volatility changes. *1976 Proceedings of the American statistical association bisiness and economic statistics section*, 177-181.

14. Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99-113

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

16. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics, 31*(3), 307-327.

17. Bollerslev, T. (2008). Glossary to ARCH (GARCH). *CREATES Research paper*, *49*.

18. Brownlees, C. T., & Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis, 51*(4), 2232-2245.

19. Buelens, C (2023). Googling "Inflation": What does Internet Search Behaviour Reveal about Household (In)attention to Inflation & Monetary Policy? *Discussion paper 183*, European Commission.

20. Castelnuovo, E., & Tran, T. D. (2017). Google it up! a google trends-based uncertainty index for the united states and australia. *Economics Letters, 161*, 149-153.

21. Challet, D., & Ayed, A. B. H. (2013). Predicting financial markets with Google Trends and not so random keywords. *arXiv preprint arXiv:1307.4643*.

22. Chen, C., Liu, L., & Zhao, N. (2020). Fear sentiment, uncertainty, and bitcoin price dynamics: The case of COVID-19. *Emerging Markets Finance and Trade, 56*(10), 2298-2309.

23. Choi, H., & Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc, 1(2009)*, 1-5.

24. Clarke, J., Jandik, T., & Mandelker, G. (2001). The efficient markets hypothesis. *Expert financial planning: Advice from industry leaders, 7*(3/4), 126-141.

25. Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174-196.

26. Damodaran, A. (2016). *Damodaran on valuation: security analysis for investment and corporate finance*. John Wiley & Sons.

27. Dimpfl, T., & Jank, S. (2016). Can internet search queries help to predict stock market volatility?. *European financial management*, *22*(2), 171-192.

28. Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance, 1*(1), 83-106.

29. D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting 33*(4), 801-816.

30. Du, R. Y., & Kamakura, W. A. (2012). Quantitative trendspotting. *Journal of Marketing Research*, *49*(4), 514-536.

31. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, 987-1007.

32. Engle, R. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, *17*(5), 425-446.

33. Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews, 5*(1), 1-50.

34. Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The journal of finance, 48*(5), 1749-1778.

35. Engle, R. F., & Patton, A. J. (2000). *What good is a volatility model? USA: University of California* (Doctoral dissertation, Dissertation-MBA).

36. Engel, J. F., Kollat, T. D., & Blackwell, R. D. (1972). *Consumer behavior*. Hinsdale.

37. Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM, 48*(11), 87-92.

38. Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business 38*(1), 34-105.

39. Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance, 25*(2), 383-417.

40. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012-1014.

41. Glosten, L. R., Jagannathan, R., & 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(5), 1779-1801.

42. Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National academy of sciences, 107*(41), 17486-17490.

43. Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization*, *117*, 62-81.

44. Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877-906.

45. Hansen, P. R., Lunde, A. (2001). *A comparison of volatility models: Does anything beat a GARCH (1, 1).* Unpublished manuscript. Department of Economics, Brown University.

46. Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics, 24*(2), 127-161.

47. Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics, 196*(1), 55-67.

48. Irresberger, F., Mühlnickel, J., & Weiß, G. N. (2015). Explaining bank stock performance with crisis sentiment. *Journal of Banking & Finance, 59*, 311-329.

49. Jun, S. P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological forecasting and social change, 130*, 69-87.

50. Kotler, P. (2016). *Marketing management (14th ed.)*. Englewood Cliffs, NJ: Prentice Hall.

51. Lee, J., Lin, Q., (2004). Consumer information search when making investment decisions. *Financial Services Review, 13*(4), 319.

52. Lin, Q., & Lee, J. (2002). *Consumers' information search when making investment decisions* (Doctoral dissertation, University of Georgia).

53. Longerstaey, J., & Spencer, M. (1996). Riskmetricstm—technical document. *Morgan Guaranty Trust Company of New York: New York, 51,* 54.

54. Mandelbrot, B. B.(1997). *The variation of certain speculative prices* (pp. 371-418). Springer New York.

55. Mandelbrot, B. B., & Hudson, R. L. (2010). *The (mis) behaviour of markets: a fractal view of risk, ruin and reward.* Profile books.

56. Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3-46). NBER.

57. Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013). Quantifying Wikipedia usage patterns before stock market moves. *Scientific reports, 3*(1), 1801.

58. Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., & Von Weizsäcker, J. E. (1997). Volatilities of different time resolutions—analyzing the dynamics of market components. *Journal of Empirical Finance, 4*(2-3), 213-239.

59. Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica: *Journal of the Econometric Society*, 347-370.

60. Nikkinen, J., & Peltomäki, J. (2020). Crash fears and stock market effects: evidence from web searches and printed news articles. *Journal of Behavioral Finance, 21*(2), 117-127.

61. Koop, G., & Onorante, L. (2019). Macroeconomic Nowcasting Using Google Probabilities. In *Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part A* (Vol. 40, pp. 17-40). Emerald Publishing Limited.

62. Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of econometrics, 45*(1-2), 267-290.

63. Papadamou, S., Fassas, A., Kenourgios, D., & Dimitriou, D. (2020). Direct and indirect effects of COVID-19 pandemic on implied stock market volatility: Evidence from panel data analysis.

64. Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific reports, 3*(1), 1684.

65. Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 368*(1933), 5707-5719.

66. Samuelson, P. A. (1973). Proof that properly discounted present values of assets vibrate randomly. *The Bell Journal of Economics and Management Science,* 369-374.

67. Schwert, G. W. (1989) Why does Stock volatility change over time? *Journal of Finance, 44*(5), 1115–1153.

68. Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics, 69*(1), 99-118.

69. Shmueli, G. (2010). To explain or to predict?. *Statistical science*, 289-310.

70. Szczygielski, J. J., Bwanya, P. R., Charteris, A., & Brzeszczyński, J. (2021). The only certainty is uncertainty: An analysis of the impact of COVID-19 uncertainty on regional stock markets. *Finance research letters, 43*, 101945.

71. Taylor, S. J. (1986): *Modelling Financial Time Series*. John Wiley & Sons.

72. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*(3), 1139-1168.

73. Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, *31*(2), 198-215.

74. Tuhkuri J. (2014). *Big data: Google-haut ennustavat työttömyyttä Suomessa* (No. 31). ETLA Report.

75. Wooldridge, J. M., (2002). *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

76. Wu, L., & Brynjolfsson, E. (2009). The future of prediction: how Google searches foreshadow housing prices and quantities. *ICIS 2009 Proceedings,* 147.

77. Yu, L., Zhao, Y., Tang, L., & Yang, Z. (2019). Online big data-driven oil consumption forecasting with Google trends. *International Journal of Forecasting, 35*(1), 213-223.

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

APPENDICES

Appendix 1: povzetek v slovenskem jeziku

Napovedovanje dinamike delniškega trga je verjetno ena izmed najbolj raziskanih tem v financah a so večji napredki na tem področju razmeroma redki. Vendar pa je uporaba spletnih podatkov o vedenju velikih skupin ljudi odprla novo področje raziskovanja borznih trgov. Teoretično osnovo uporabe spletnih podatkov lahko izsledimo v klasični hipotezi učinkovitega trga (EMH), ki trdi, da so nove informacije hitro in učinkovito vgrajene v ceno delnice, in da cene delnic zaradi naključnega prihoda novih informacij sledijo naključni dinamiki (Fama, 1965). S tega vidika so informacije tiste, ki poganjajo spremembe cen delnic, zato je mogoče novo ustvarjene ali nove vrste informacij do določene mere uporabiti za napovedovanje dinamike trgov. Pojav širše uporabe interneta je ustvaril nov tok spletnih informacij, ki naj bi se s potrjeno informacijsko učinkovitostjo odražal tudi v borzni dinamiki.

Cilj tega dela je preučiti vprašanje, ali obseg iskanja na Googlu napoveduje prihodnjo volatilnost na trgih vrednostnih papirjev. Vendar je takšno vprašanje na splošno preširoko, zato je namen tega dela le delna ocena tega področja. Bolj operativno formulacijo lahko artikuliram kot vprašanje ali lahko uporaba obsega iskanj pogosto uporabljenih finančnih besed glede na jezik lokacije borznega indeksa kot dodatnega napovednika v heterogenem avtoregresijskem (HAR) modelu izboljša napoved realizirane volatilnosti borznega indeksa.

Raziskava združuje dva vira podatkov. Prvi vir so podatki o realizirani volatilnosti delniških indeksov, ki ga je sestavil Oxford-Man Institute of Quantitative Finance. Podatki predstavljajo serijo dnevnih realiziranih volatilnosti, ki temeljijo na donosih v 5-minutnih intervalih. Ti podatki so široko uporabljeni v raziskavah volatilnosti, tudi v kombinaciji z Google Trends podatki. Borzni indeksi, uporabljeni v moji raziskavi so FTSE 100, DAX, OMXHPI, IBEX in OMXSPI.

Poleg podatkov borznih indeksov, v raziskavi uporabim tudi podatke o obsegu iskanj na Google-u, ki so na voljo preko storitev Google Trends. Ti podatki segajo od 1. januarja 2004 do datuma analize. Seznam iskalnih izrazov, za katere je bil obseg iskanja pridobljen preko Google Trends, temelji na delu Preis et al. (2013). Avtorji so ocenili stopnje finančne pomembnosti določenih besed oziroma izrazov. Seznam iskalnih izrazov uporabljen v tej raziskavi je sestavljen iz besed, ki imajo ocenjeno najvišjo finančno pomembnost.

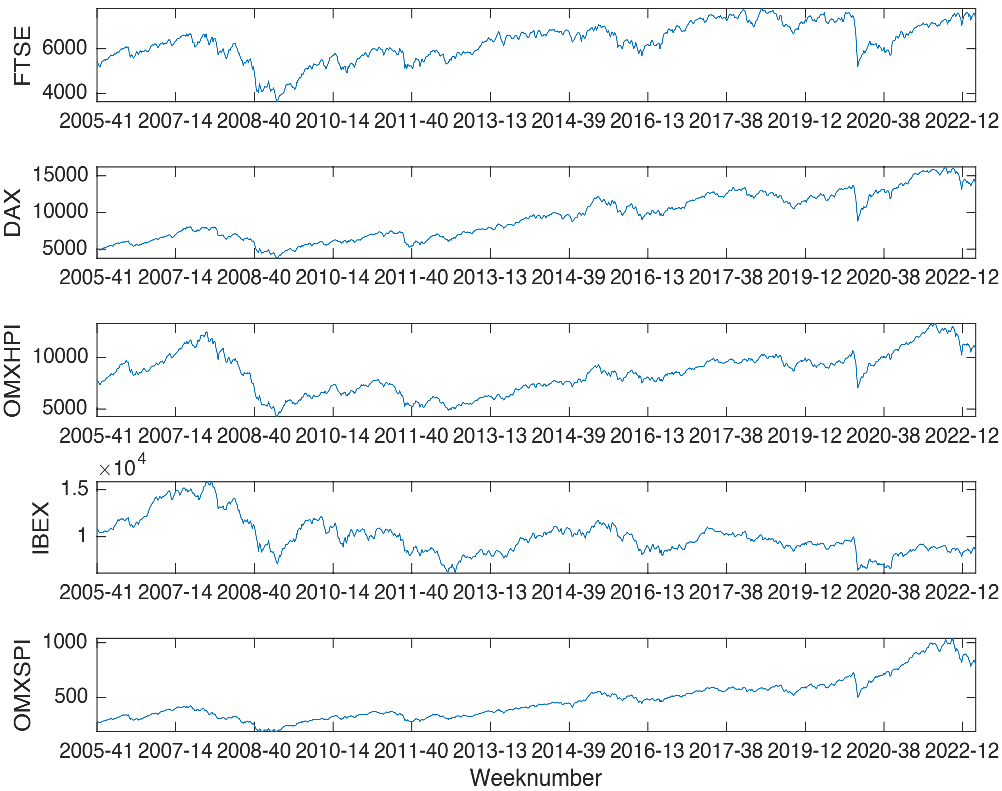
Ocena, ali uporaba podatkov Google Trends izboljša napoved, temelji na kriteriju uspešnosti napovedne natančnosti izven vzorca, pri čemer uporabim statistične mere kot so povprečna kvadratna napaka (MSE), povprečna absolutna napaka (MAE), povprečna absolutna odstotkovna napaka (MAPE) in determinacijski koeficient R^2 v Mincer-Zarnowitz regresijah (1969). Napovedno natančnost sem iterativno meril za en teden vnaprej. S to metodo so vse napovedi dejansko izven vzorca, saj čas napovednega obdobja za nobeno od napovedi ne vstopa v okno ocenjevanja modela.

V okviru raziskave sem skupno ocenil 180 različic modelov, ki izhajajo iz vključitve devet iskalnih izrazov na petih evropskih trgih v štirih konfiguracijah modelov. Vključitev podatkov Google Trends je v 79% primerov analiziranih modelov izboljšala napovedno natančnost. V povprečju vključevanje podatkov Google Trends izboljšuje natančnost napovedi, če modele ločimo glede na različne dejavnike kot so jezik, konfiguracija modela ali statistična mera. Skupni rezultati kažejo, da imajo podatki o obsegu iskanj finančno pomembnih besed na Google-u potencial za izboljšanje napovedovanja volatilnosti, vprašanje pa je ali je izboljšanje napovedne natančnosti dovolj pomembno, da bi bilo tudi praktično uporabno. Pri vseh merilih uspešnosti so se izboljšave napovedne natančnosti gibale v razponu 1%, kjer noben jezik, specifikacija modela ali ključna beseda ni odstopala iz tega vzorca.

Glede na dobljene rezultate ostaja raziskovalno vprašanje do neke mere še vedno odprto. Po eni strani je raziskava pokazala, da ima vključitev komponent obsega iskanja na Google-u v napoved dokaj zanesljiv učinek izboljšanja napovedne natančnosti v vseh raziskovanih dimenzijah. Učinek je mogoče opaziti tako v kontekstu različnih jezikov, iskanih izrazov, trgov kot tudi konfiguracijah modela. Po drugi strani pa je raziskava poakazala, da so izboljšave napovedne natančnosti razmeroma majhne in je njihov praktični pomen vprašljiv.

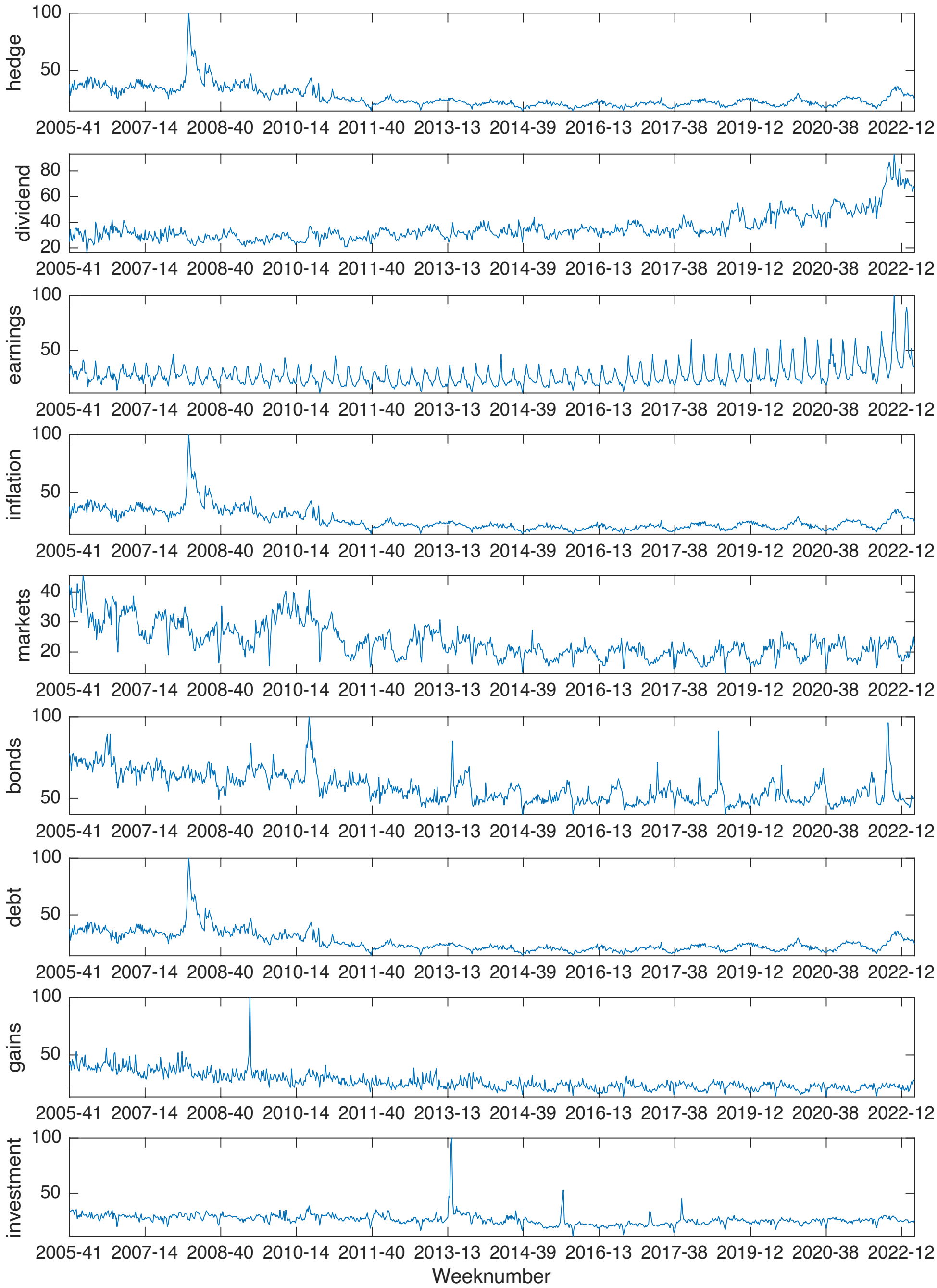
Appendix 2: Google search term time series

Figure 20: Index opening prices



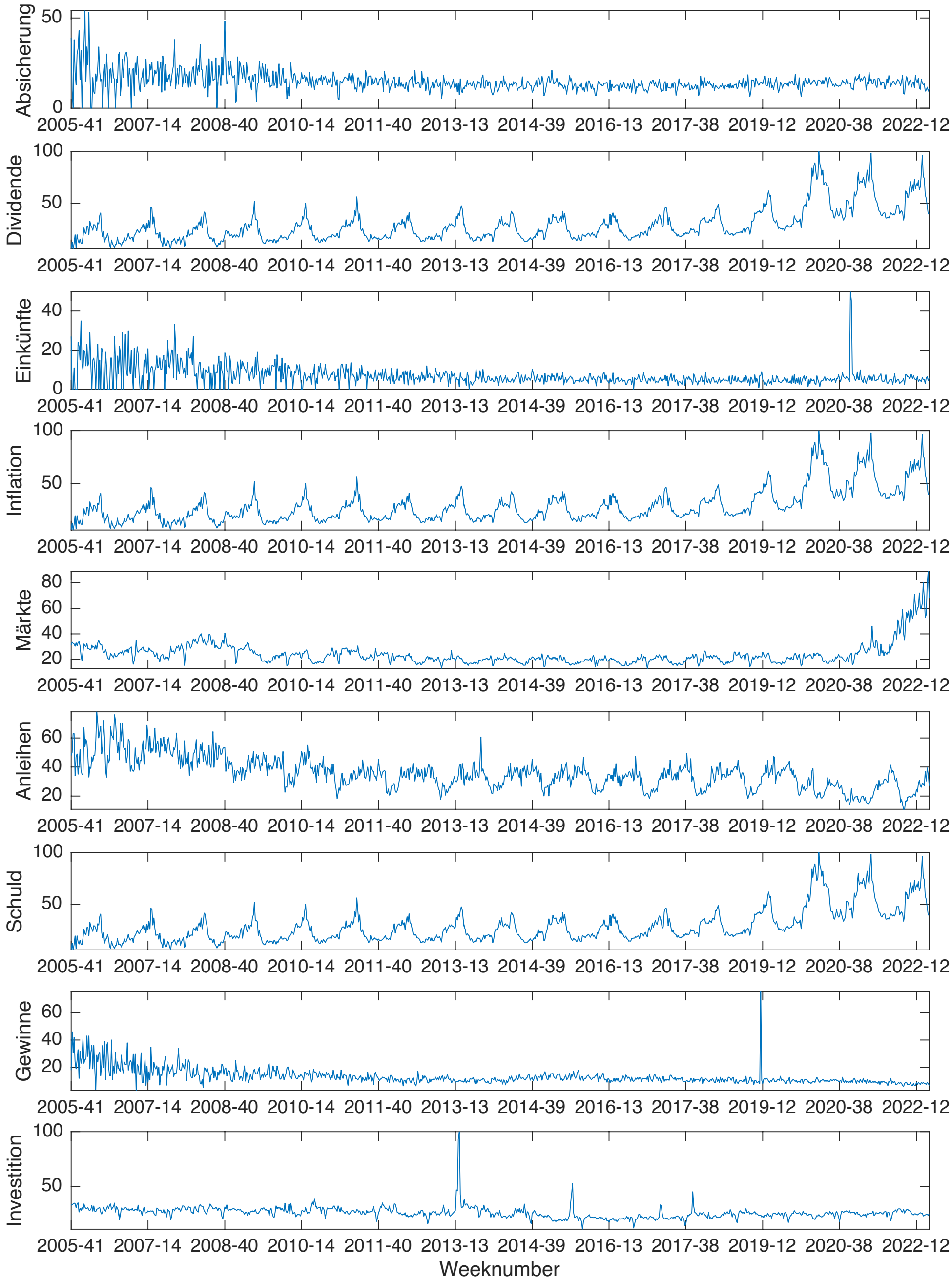
Source: Own work.

Figure 21: Google Trends search volume index for English keywords



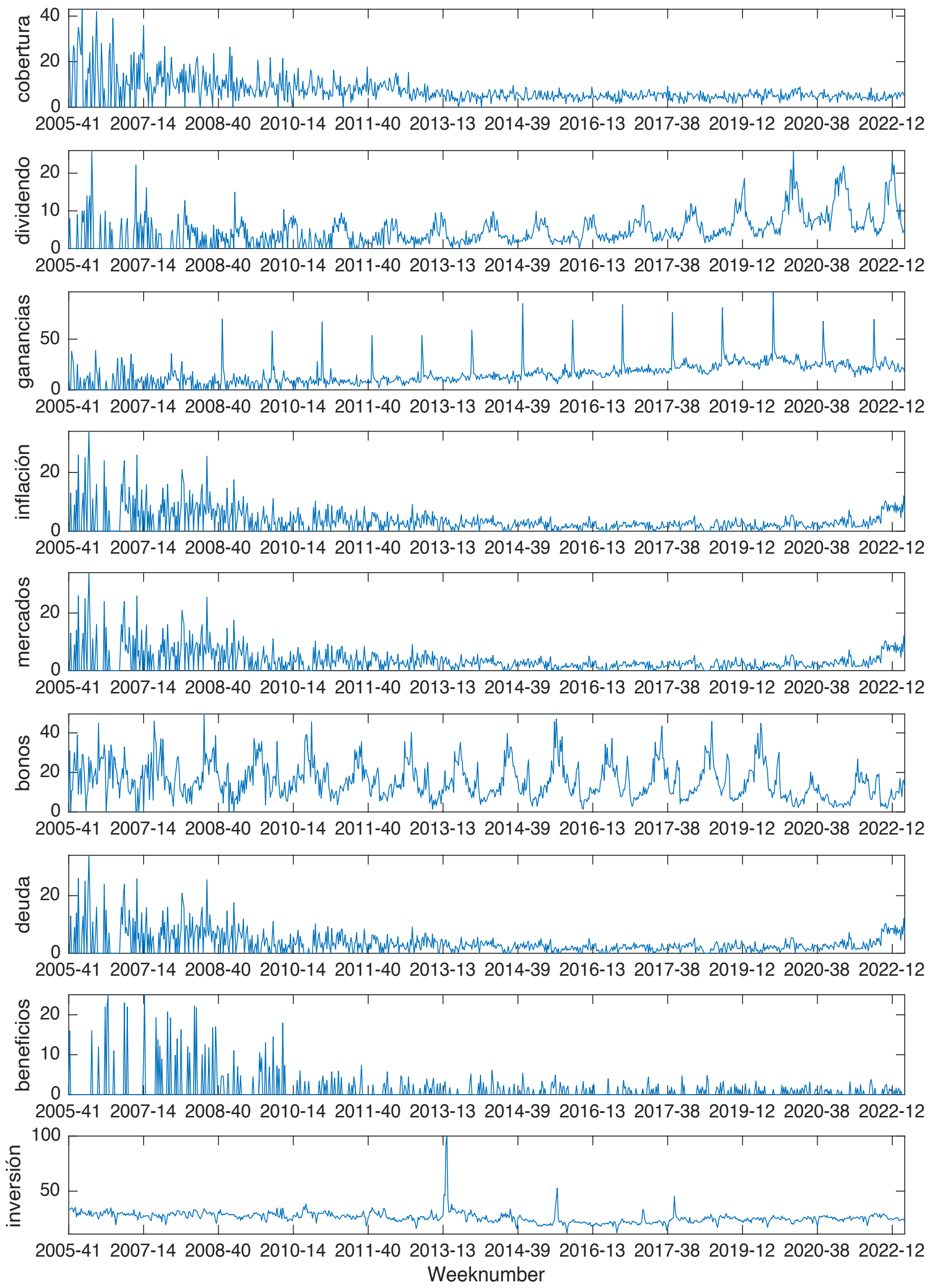
Source: Own work.

Figure 22: Google Trends search volume index for German keywords



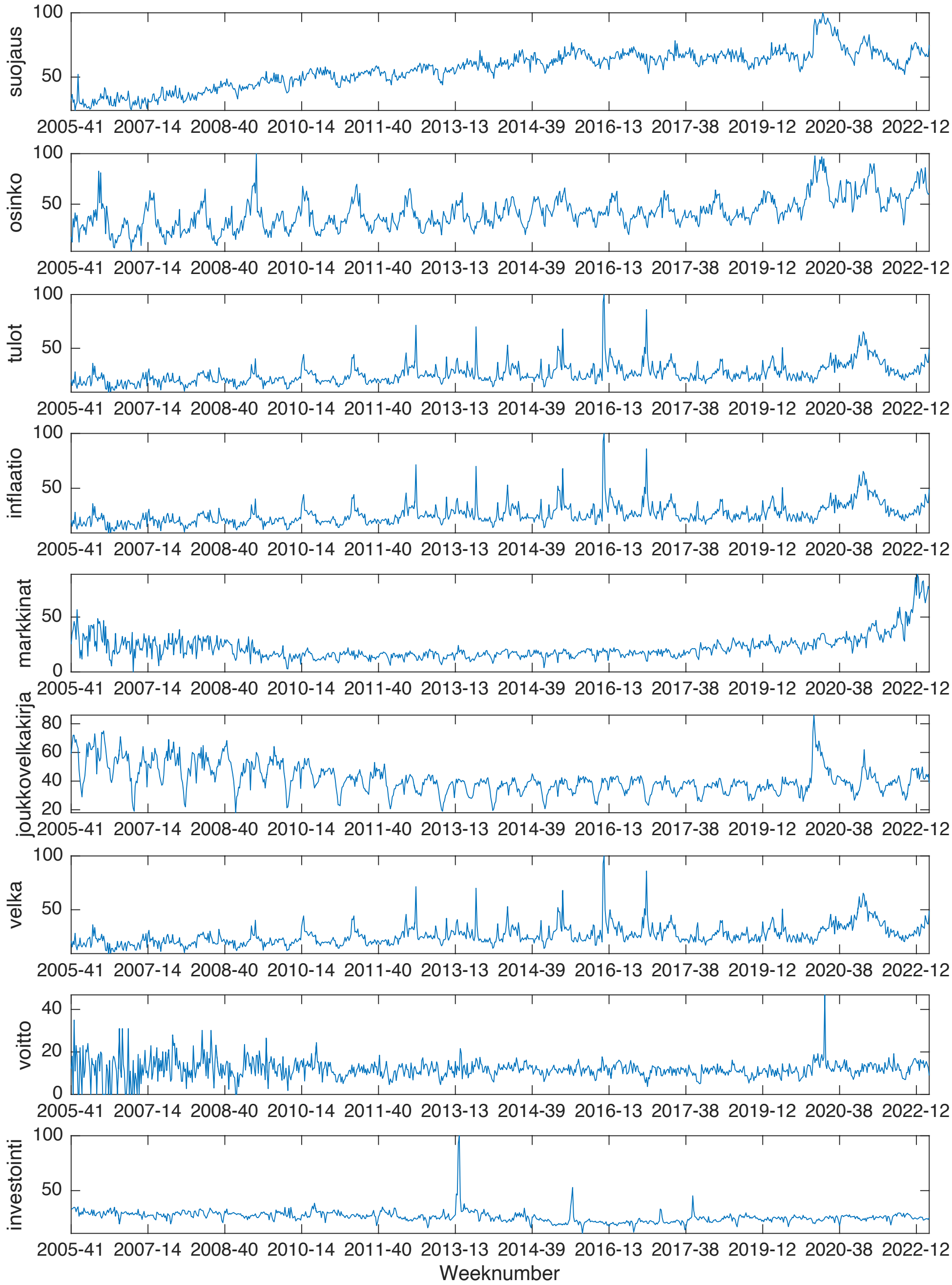
Source: Own work.

Figure 23:Google Trends search volume index for Spanish keywords



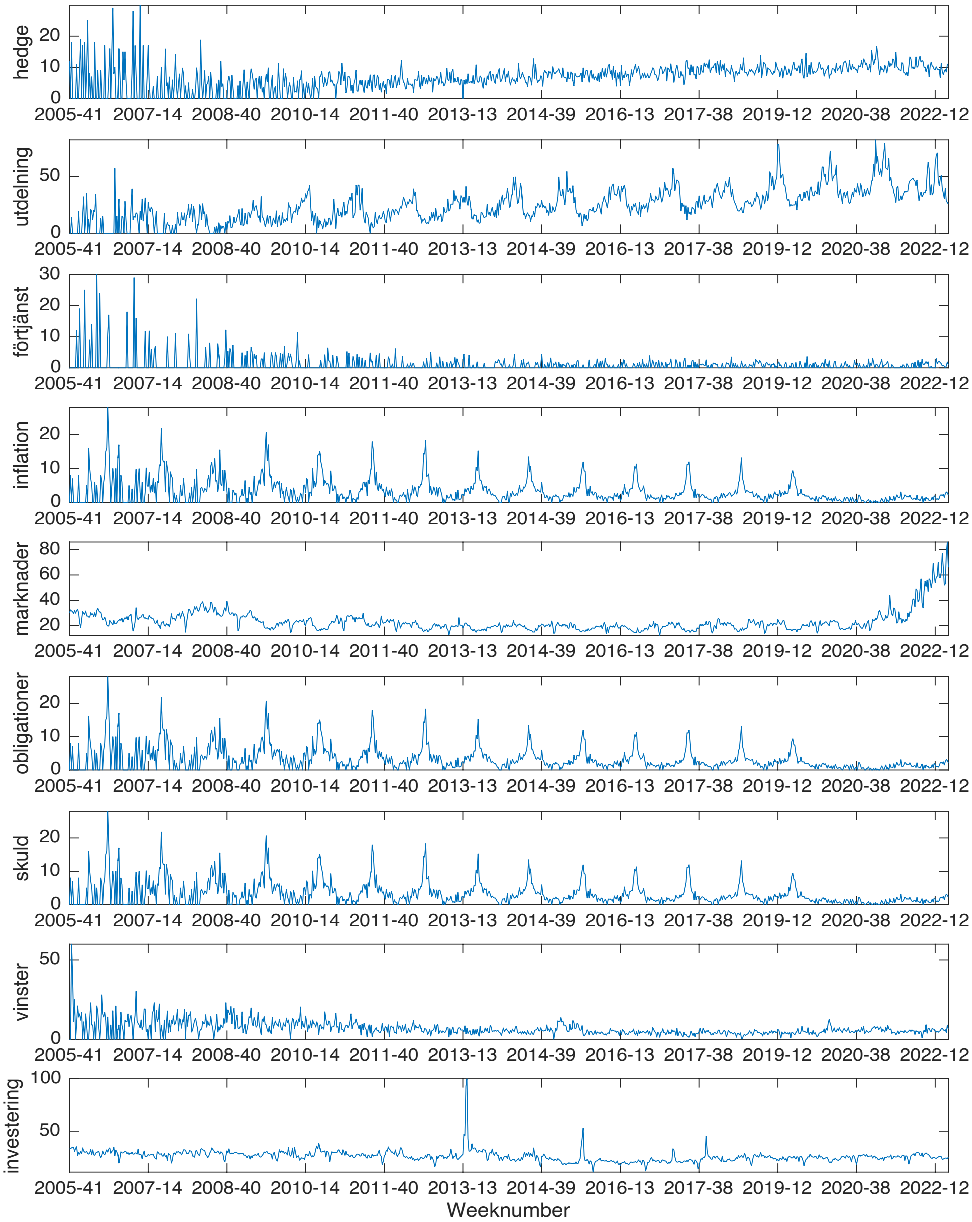
Source: Own work.

Figure 24:Google Trends search volume index for Finnish keywords



Source: Own work.

Figure 25: Google Trends search volume index for Swedish keywords



Source: Own work.

Appendix 3: Full predictive accuracy results

MAD

Table 11: HAR-RV – Moving window - MAD

	English	**German**	**Spanish**	**Finnish**	**Swedish**	**Average**
HAR-RV - MW	0.00179	0.00174	0.00163	0.00192	0.00151	0.00172
Average of GT	0.00179	0.00174	0.00163	0.00191	0.00151	0.00171
hedge	0.001825	0.001713	0.001649	0.001897	0.001511	0.00172
dividend	0.001780	0.001744	0.001638	0.001918	0.001507	0.00172
earnings	0.001775	0.001737	0.001630	0.001909	0.001519	0.00171
inflation	0.001783	0.001753	0.001599	0.001893	0.001501	0.00171
markets	0.001782	0.001734	0.001607	0.001902	0.001512	0.00171
bonds	0.001788	0.001740	0.001625	0.001895	0.001507	0.00171
debt	0.001775	0.001737	0.001635	0.001905	0.001520	0.00171
gains	0.001790	0.001738	0.001626	0.001922	0.001499	0.00172
investment	0.001787	0.001730	0.001621	0.001914	0.001505	0.00171

Source: Own work.

Table 12: HAR-RV – Expanding window - MAD

	English	**German**	**Spanish**	**Finnish**	**Swedish**	**Average**
HaR-RV - EW	0.00182	0.00180	0.00169	0.00195	0.00157	0.00177
Average of GT	0.00182	0.00179	0.00168	0.00195	0.00156	0.00176
hedge	0.00181	0.00178	0.00172	0.00195	0.00155	0.00176
dividend	0.00180	0.00181	0.00167	0.00195	0.00154	0.00175
earnings	0.00182	0.00179	0.00167	0.00194	0.00156	0.00176
inflation	0.00182	0.00183	0.00169	0.00195	0.00156	0.00177
markets	0.00184	0.00177	0.00165	0.00195	0.00157	0.00176
bonds	0.00180	0.00179	0.00169	0.00195	0.00155	0.00176
debt	0.00182	0.00179	0.00174	0.00194	0.00157	0.00177
gains	0.00183	0.00180	0.00165	0.00196	0.00157	0.00176
investment	0.00181	0.00179	0.00169	0.00195	0.00156	0.00176

Source: Own work.

Table 13: HAR-ln(RV) – Moving window - MAD

	English	German	Spanish	Finnish	Swedish	Average
HAR-ln(RV) - MW	0.00174	0.00175	0.00160	0.00189	0.00149	0.00169
Average of GT	0.00174	0.00174	0.00159	0.00188	0.00149	0.00169
hedge	0.00174	0.00173	0.00160	0.00187	0.00149	0.00168
dividend	0.00173	0.00175	0.00159	0.00188	0.00149	0.00169
earnings	0.00174	0.00174	0.00160	0.00189	0.00150	0.00169
inflation	0.00173	0.00175	0.00158	0.00187	0.00148	0.00168
markets	0.00172	0.00173	0.00158	0.00188	0.00149	0.00168
bonds	0.00174	0.00175	0.00159	0.00187	0.00147	0.00169
debt	0.00173	0.00174	0.00157	0.00187	0.00149	0.00168
gains	0.00174	0.00175	0.00160	0.00189	0.00149	0.00169
investment	0.00173	0.00174	0.00160	0.00189	0.00147	0.00169

Source: Own work.

Table 14: HAR-ln(RV) – Expanding window - MAD

	English	German	Spanish	Finnish	Swedish	Average
HAR-ln(RV) - EW	0.00180	0.00179	0.00162	0.00193	0.00152	0.00173
Average of GT	0.00179	0.00178	0.00163	0.00193	0.00152	0.00173
hedge	0.00179	0.00178	0.00165	0.00193	0.00151	0.00173
dividend	0.00178	0.00179	0.00162	0.00193	0.00151	0.00173
earnings	0.00180	0.00178	0.00162	0.00193	0.00152	0.00173
inflation	0.00180	0.00178	0.00164	0.00193	0.00151	0.00173
markets	0.00179	0.00177	0.00161	0.00193	0.00152	0.00173
bonds	0.00179	0.00178	0.00162	0.00193	0.00151	0.00173
debt	0.00180	0.00178	0.00163	0.00193	0.00152	0.00173
gains	0.00180	0.00179	0.00162	0.00193	0.00152	0.00173
investment	0.00179	0.00178	0.00163	0.00193	0.00151	0.00173

Source: Own work.

MAPE

Table 15: HAR-RV – Moving window - MAPE

	English	**German**	**Spanish**	**Finnish**	**Swedish**	**Average**
HaR-RV - MW	19.56%	18.47%	18.38%	17.21%	18.54%	18.43%
Average of GT	19.60%	18.39%	18.40%	17.11%	18.48%	18.40%
hedge	20.22%	18.22%	18.65%	17.00%	18.46%	18.51%
dividend	19.40%	18.62%	18.61%	17.24%	18.40%	18.46%
earnings	19.37%	18.41%	18.42%	17.18%	18.54%	18.38%
inflation	19.50%	18.54%	18.08%	16.97%	18.41%	18.30%
markets	19.64%	18.27%	17.98%	17.09%	18.59%	18.31%
bonds	19.59%	18.42%	18.40%	16.96%	18.55%	18.39%
debt	19.43%	18.41%	18.67%	17.12%	18.61%	18.45%
gains	19.68%	18.38%	18.39%	17.27%	18.31%	18.40%
investment	19.58%	18.26%	18.39%	17.18%	18.48%	18.38%

Source: Own work.

Table 16: HAR-RV – Expanding window - MAPE

	English	**German**	**Spanish**	**Finnish**	**Swedish**	**Average**
HaR-RV - EW	19.86%	19.13%	19.47%	17.44%	19.63%	19.10%
Average of GT	19.77%	18.82%	19.27%	17.44%	19.29%	18.92%
hedge	19.64%	18.54%	19.51%	17.55%	19.18%	18.88%
dividend	19.15%	19.15%	19.20%	17.41%	18.64%	18.71%
earnings	19.77%	18.69%	19.12%	17.37%	19.41%	18.87%
inflation	19.93%	19.10%	19.57%	17.35%	19.03%	19.00%
markets	20.41%	18.40%	18.38%	17.45%	19.51%	18.83%
bonds	19.43%	18.81%	19.47%	17.42%	19.17%	18.86%
debt	19.85%	18.89%	20.49%	17.36%	19.60%	19.24%
gains	20.06%	19.10%	18.40%	17.59%	19.69%	18.97%
investment	19.65%	18.73%	19.34%	17.47%	19.36%	18.91%

Source: Own work.

Table 17: HAR-ln(RV) – Moving window - MAPE

	English	**German**	**Spanish**	**Finnish**	**Swedish**	**Average**
HAR-ln(RV) - MW	0.18066	0.17632	0.16963	0.16361	0.17252	0.17255
Average of GT	0.17998	0.17548	0.16928	0.16278	0.17182	0.17187
hedge	0.18248	0.17415	0.17055	0.16167	0.17163	0.17210
dividend	0.17925	0.17716	0.16906	0.16315	0.17198	0.17212
earnings	0.17941	0.17576	0.16998	0.16390	0.17283	0.17237
inflation	0.17946	0.17591	0.16814	0.16201	0.17108	0.17132
markets	0.17945	0.17396	0.16643	0.16254	0.17307	0.17109
bonds	0.17964	0.17666	0.16969	0.16180	0.17056	0.17167
debt	0.17943	0.17525	0.16933	0.16241	0.17263	0.17181
gains	0.18084	0.17645	0.17023	0.16409	0.17183	0.17269
investment	0.17989	0.17404	0.17013	0.16347	0.17080	0.17166

Source: Own work.

Table 18: HAR-ln(RV) – Expanding window - MAPE

	English	**German**	**Spanish**	**Finnish**	**Swedish**	**Average**
HAR-ln(RV) - EW	0.18708	0.18151	0.17577	0.16787	0.18036	0.17852
Average of GT	0.18601	0.18019	0.17625	0.16771	0.17901	0.17783
hedge	0.18521	0.18040	0.17900	0.16757	0.17742	0.17792
dividend	0.18269	0.18214	0.17561	0.16768	0.17719	0.17706
earnings	0.18697	0.18073	0.17606	0.16831	0.18054	0.17852
inflation	0.18725	0.17972	0.17726	0.16777	0.17703	0.17781
markets	0.18717	0.17812	0.17215	0.16777	0.18006	0.17705
bonds	0.18410	0.17990	0.17576	0.16736	0.17823	0.17707
debt	0.18766	0.18021	0.17946	0.16738	0.18070	0.17908
gains	0.18752	0.18152	0.17471	0.16760	0.18041	0.17835
investment	0.18549	0.17899	0.17623	0.16798	0.17947	0.17763

Source: Own work.

MSE

Table 19: HAR-RV – Moving window - MSE

	English	German	Spanish	Finnish	Swedish	Average
HaR-RV - MW	8.7E-06	8.0E-06	8.8E-06	8.2E-06	6.6E-06	8.1E-06
Average of GT	8.6E-06	7.8E-06	8.6E-06	8.2E-06	6.5E-06	7.9E-06
hedge	8.6E-06	7.5E-06	8.6E-06	8.1E-06	6.6E-06	7.9E-06
dividend	8.6E-06	7.9E-06	8.8E-06	8.2E-06	6.6E-06	8.0E-06
earnings	8.6E-06	7.9E-06	8.8E-06	8.2E-06	6.6E-06	8.0E-06
inflation	8.6E-06	7.6E-06	8.4E-06	8.1E-06	6.3E-06	7.8E-06
markets	8.4E-06	7.9E-06	8.6E-06	8.2E-06	6.6E-06	7.9E-06
bonds	8.6E-06	7.9E-06	8.7E-06	8.2E-06	6.5E-06	8.0E-06
debt	8.5E-06	7.9E-06	8.5E-06	8.1E-06	6.6E-06	7.9E-06
gains	8.6E-06	7.9E-06	8.7E-06	8.2E-06	6.6E-06	8.0E-06
investment	8.6E-06	7.9E-06	8.6E-06	8.2E-06	6.6E-06	8.0E-06

Source: Own work.

Table 20: HAR-RV – Expanding window - MSE

	English	German	Spanish	Finnish	Swedish	Average
HaR-RV - EW	9.6E-06	8.8E-06	9.2E-06	9.0E-06	7.1E-06	8.7E-06
Average of GT	9.5E-06	8.7E-06	9.1E-06	9.0E-06	7.0E-06	8.7E-06
hedge	9.4E-06	8.6E-06	9.1E-06	9.0E-06	7.1E-06	8.6E-06
dividend	9.5E-06	8.7E-06	9.1E-06	8.9E-06	7.1E-06	8.7E-06
earnings	9.5E-06	8.7E-06	9.1E-06	9.0E-06	7.0E-06	8.7E-06
inflation	9.5E-06	8.6E-06	9.1E-06	8.9E-06	6.9E-06	8.6E-06
markets	9.5E-06	8.7E-06	9.0E-06	9.0E-06	7.1E-06	8.7E-06
bonds	9.5E-06	8.7E-06	9.1E-06	9.0E-06	7.0E-06	8.7E-06
debt	9.5E-06	8.7E-06	9.1E-06	9.0E-06	7.1E-06	8.7E-06
gains	9.5E-06	8.7E-06	9.1E-06	9.0E-06	7.1E-06	8.7E-06
investment	9.5E-06	8.7E-06	9.1E-06	9.0E-06	7.0E-06	8.7E-06

Source: Own work.

Table 21: HAR-ln(RV) – Moving window - MSE

	English	German	Spanish	Finnish	Swedish	Average
HAR-ln(RV) - MW	9.7E-06	9.0E-06	9.4E-06	8.9E-06	7.2E-06	8.8E-06
Average of GT	9.6E-06	8.9E-06	9.2E-06	8.8E-06	7.1E-06	8.7E-06
hedge	9.3E-06	8.8E-06	9.1E-06	8.8E-06	7.2E-06	8.6E-06
dividend	9.7E-06	8.9E-06	9.4E-06	8.8E-06	7.2E-06	8.8E-06
earnings	9.6E-06	8.9E-06	9.3E-06	8.9E-06	7.2E-06	8.8E-06
inflation	9.7E-06	8.7E-06	9.0E-06	8.8E-06	6.7E-06	8.6E-06
markets	9.5E-06	8.9E-06	9.2E-06	8.8E-06	7.2E-06	8.7E-06
bonds	9.7E-06	8.9E-06	9.3E-06	8.8E-06	7.0E-06	8.7E-06
debt	9.6E-06	8.9E-06	9.0E-06	8.7E-06	7.2E-06	8.7E-06
gains	9.6E-06	8.9E-06	9.4E-06	8.9E-06	7.2E-06	8.8E-06
investment	9.6E-06	9.1E-06	9.2E-06	8.9E-06	7.1E-06	8.8E-06

Source: Own work.

Table 22: HAR-ln(RV) – Expanding window - MSE

	English	German	Spanish	Finnish	Swedish	Average
HAR-ln(RV) - EW	1.0E-05	9.3E-06	9.6E-06	9.4E-06	7.4E-06	9.2E-06
Average of GT	1.0E-05	9.3E-06	9.6E-06	9.4E-06	7.3E-06	9.1E-06
hedge	1.0E-05	9.2E-06	9.5E-06	9.4E-06	7.4E-06	9.1E-06
dividend	1.0E-05	9.2E-06	9.6E-06	9.4E-06	7.4E-06	9.2E-06
earnings	1.0E-05	9.3E-06	9.6E-06	9.4E-06	7.3E-06	9.1E-06
inflation	1.0E-05	9.2E-06	9.6E-06	9.4E-06	7.1E-06	9.1E-06
markets	1.0E-05	9.3E-06	9.5E-06	9.4E-06	7.3E-06	9.1E-06
bonds	1.0E-05	9.3E-06	9.6E-06	9.4E-06	7.3E-06	9.1E-06
debt	1.0E-05	9.2E-06	9.5E-06	9.4E-06	7.3E-06	9.1E-06
gains	1.0E-05	9.3E-06	9.6E-06	9.4E-06	7.3E-06	9.1E-06
investment	1.0E-05	9.3E-06	9.6E-06	9.4E-06	7.2E-06	9.2E-06

Source: Own work.

MZ

Table 23: HAR-RV – Moving window - MZ

	English	German	Spanish	Finnish	Swedish	Average
HaR-RV - MW	0.72452	0.73809	0.71830	0.67296	0.75672	0.72212
Average of GT	0.72760	0.74293	0.72262	0.67612	0.76028	0.72591
hedge	0.72873	0.75272	0.72231	0.67922	0.75799	0.72820
dividend	0.72727	0.73972	0.71774	0.67538	0.75906	0.72383
earnings	0.72670	0.74082	0.71801	0.67393	0.75910	0.72371
inflation	0.72578	0.74998	0.72865	0.67818	0.77097	0.73071
markets	0.73092	0.74174	0.72381	0.67592	0.75838	0.72616
bonds	0.72626	0.74064	0.72011	0.67542	0.76039	0.72456
debt	0.72843	0.73985	0.72797	0.67804	0.75798	0.72645
gains	0.72704	0.74006	0.71997	0.67577	0.75880	0.72433
investment	0.72723	0.74081	0.72506	0.67320	0.75986	0.72523

Source: Own work.

Table 24: HAR-RV – Expanding window - MZ

	English	German	Spanish	Finnish	Swedish	Average
HaR-RV - EW	0.69609	0.71109	0.70628	0.64390	0.74076	0.69962
Average of GT	0.69802	0.71382	0.70847	0.64485	0.74232	0.70150
hedge	0.70156	0.71853	0.70851	0.64411	0.74129	0.70280
dividend	0.69726	0.71337	0.70650	0.64578	0.74170	0.70092
earnings	0.69714	0.71312	0.70773	0.64510	0.74233	0.70108
inflation	0.69725	0.71642	0.70836	0.64536	0.74678	0.70283
markets	0.69813	0.71388	0.71045	0.64492	0.74135	0.70175
bonds	0.69803	0.71205	0.70695	0.64443	0.74277	0.70084
debt	0.69749	0.71258	0.71060	0.64547	0.74104	0.70144
gains	0.69747	0.71229	0.70860	0.64460	0.74102	0.70080
investment	0.69789	0.71210	0.70851	0.64389	0.74263	0.70100

Source: Own work.

Table 25: HAR-ln(RV) – Moving window - MZ

	English	German	Spanish	Finnish	Swedish	Average
HAR-ln(RV) - MW	0.69393	0.70898	0.70345	0.65195	0.73964	0.69959
Average of GT	0.69632	0.71277	0.70756	0.65508	0.74458	0.70326
hedge	0.70328	0.71531	0.71041	0.65902	0.74156	0.70592
dividend	0.69326	0.71193	0.70185	0.65712	0.74166	0.70116
earnings	0.69574	0.71274	0.70388	0.65058	0.74089	0.70077
inflation	0.69258	0.72250	0.71497	0.65717	0.75970	0.70938
markets	0.70010	0.71339	0.70871	0.65542	0.74254	0.70403
bonds	0.69459	0.71109	0.70590	0.65484	0.74595	0.70248
debt	0.69580	0.71023	0.71333	0.65868	0.74204	0.70402
gains	0.69490	0.71006	0.70248	0.65187	0.74168	0.70020
investment	0.69665	0.70763	0.70646	0.65102	0.74520	0.70139

Source: Own work.

Table 26: HAR-ln(RV) – Expanding window - MZ

	English	German	Spanish	Finnish	Swedish	Average
HAR-ln(RV) - EW	0.67906	0.69680	0.69338	0.62903	0.73268	0.68619
Average of GT	0.67962	0.69939	0.69480	0.62953	0.73542	0.68775
hedge	0.68196	0.69986	0.69668	0.62958	0.73357	0.68833
dividend	0.67944	0.69816	0.69386	0.62940	0.73310	0.68679
earnings	0.67964	0.69877	0.69223	0.62945	0.73472	0.68696
inflation	0.67916	0.70682	0.69616	0.62989	0.74376	0.69116
markets	0.68275	0.69993	0.69744	0.62950	0.73352	0.68863
bonds	0.67944	0.69811	0.69311	0.63017	0.73568	0.68730
debt	0.67756	0.69876	0.69608	0.62990	0.73341	0.68714
gains	0.67906	0.69669	0.69373	0.62952	0.73438	0.68668
investment	0.67754	0.69736	0.69395	0.62838	0.73667	0.68678

Source: Own work.