Forecasting Volatility with Empirical Similarity and Google Trends

Alain Hamid
Department of Statistics, University of Augsburg, Universitaetstr. 16, D-86159 Augsburg, Germany
alain.hamid@wiwi.uni-augsburg.de

Moritz Heiden*
Department of Statistics, University of Augsburg, Universitaetstr. 16, D-86159 Augsburg, Germany
moritz.heiden@wiwi.uni-augsburg.de

This paper proposes an empirical similarity approach to forecast weekly volatility by using search engine data as a measure of investors attention to the stock market index. Our model is assumption free with respect to the underlying process of investors attention and significantly outperforms conventional time-series models in an out-of-sample forecasting framework. We find that especially in high-volatility market phases prediction accuracy increases together with investor attention. The practical implications for risk management are highlighted in a Value-at-Risk forecasting exercise, where our model produces significantly more accurate forecasts while requiring less capital due to fewer overpredictions.

Key words: empirical similarity; google trends; investor attention; volatility; forecasting

1. Introduction

When examining how volatility in stock markets arises, one of the most interesting questions includes the role played by investors behavior, especially the interest they take in the market as well as their uncertainty about its future outcome. Recent research focusses on retail investors attention on the stock market, as studies show that retail investors trades do indeed move markets on short (weekly) horizons (see e.g. Barber et al. (2009)) and investors disagreement causes higher trading volume (e.g. Li and Li (2014)). Traditionally, interest in the market is measured by indirect proxies like volume, turnover and news. While volume might be the natural candidate to link investor attention and volatility, several studies, e.g. Brooks (1998) and Donaldson and Kamstra (2005) demonstrate that it does not improve the accuracy of volatility forecasts. News as an alternative measure are mostly irregular and may underly a considerable publication lag. Recent publications use internet message postings (Kim and Kim 2014), Facebook users sentiment data (Siganos et al.

* The authors gratefully acknowledge the support of the Research Center Global Business Management at Augsburg University, as well as financial support from the German Research Foundation (DFG)-Project "Wishart Processes in Statistics and Econometrics: Theory and Applications".

2014) or search frequencies (Vozlyublennaia 2014) to assess the influence of retail investors attention on the stock market. Among these studies, Da et al. (2011), Vlastakis and Markellos (2012) and Andrei and Hasler (2013), suggest that Google search volume is a driver of future volatility.

In this paper our scope of work is twofold. Since previous studies, prominently Vlastakis and Markellos (2012) and Vozlyublennaia (2014), have focused on analyzing the in-sample properties of the relationship between investor attention and volatility by using Google search volume, we take the discussion further and concentrate on the predictability in an out-of-sample forecasting framework. We argue, that if investor attention is not only correlated with, but indeed a cause of volatility, this should enable us to form superior predictions from a model that makes use of the dynamics of both time-series. As theoretical investigations and in-sample comparisons in previous research (e.g. Vlastakis and Markellos (2012), Andrei and Hasler (2013) and Vozlyublennaia (2014)) have found that volatility and investor attention show strong correlation on short horizons, we focus on the weekly horizon, which is the shortest one possible, as Google has restricted the access to (nonstandardized) daily data. We statistically test the ability of Google data to forecast volatility compared to relevant benchmark models using the Model Confidence Set approach of Hansen et al. (2011). We do not find that simply adding a Google component in a regression context results in significant better out-of-sample forecasts compared to conventional models of volatility, even if a cross-correlation and in-sample analysis implies a strong dependence between investor attention and volatility.

This leads us to the question, whether the dynamics of investor attention can be correctly depicted by an additive component in conventional time-series models for volatility. As an alternative, we suggest including Google data in the framework of empirical similarity introduced by Gilboa et al. (2006), which has already been applied in studying behavioral phenomena in portfolio theory by Golosnoy and Okhrin (2008). An adaption of the model, making it suitable for forecasting volatility was proposed in Golosnoy et al. (2014). Lieberman (2012) suggests a simple extension to an AR(1) model, where the autoregressive parameter is determined by empirical similarity. We follow his approach and augment an AR(1) model by a time-varying coefficient determined by the empirical similarity between last periods Google data and volatility. The unique assumption behind the model is, that volatility increases and decreases with investor attention, depending on the previous level of volatility. This approach allows us to study the relationship between investor attention and volatility while being more flexible in the dynamics of the process, as it allows for stationary, non stationary and explosive behavior. Thereby, the model provides a simple, yet flexible framework for forecasting volatility. Comparing and testing predictive accuracy, we find that particularly in crisis phases, our model significantly improves upon standard time-series models with and without additional Google components. This is consistent with the theory of Andrei and Hasler (2013), who state that in "'panic states" where volatility is high, investors pay more attention to the market. In an economic application forecasting weekly Value-at-Risk (VaR), we show that more accurate volatility forecasts also lead to improved VaR forecasts. Since VaR exceedances tend to cluster in crisis periods (see e.g. Candelon et al. (2011)), our model is beneficial for risk management as forecasting accuracy translates in more precise VaR forecasts as well as less overall capital requirements.

2. Theory and prior literature

Retail investors behavior and their impact on the stock market are well documented in the agent-based literature, e.g. Lux and Marchesi (1999) and Alfarano and Lux (2007), where uninformed investors (noise traders) act as an additional source of volatility. Barber et al. (2009) study the trading behavior of individual investors and confirm that their buying and selling leads to over-, respectively underpricing of the assets on a weekly horizon. A recent study by Li and Li (2014) on household investors suggests, that not all of their trading behavior is unsophisticated or random. In their sample of 30 years of survey data, they find that dispersion of believes about the economic outlook among investors are positively related to stock market trading. However, as the authors point out, they do not test or necessarily assume rationality behind the trades.

Search engine data as a measure of investor attention follows a similar path. The use of e.g. Google to find information on a certain stock does not imply nor deny a rationale behavior, but seems to be strongly linked to stock market participation (see e.g. Preis et al. (2010)). Being the most commonly used search engine for collecting information on the internet, accounting for 77.46% [1] of all desktop user search queries worldwide in 2013, Google search frequency data is a regular (daily) and contemporary data source. As Da et al. (2011) point out, Google is likely to be representative of the general internet search behavior, but searching a term is rather a measure for retail investors than professional investors attention.

Google data has already been applied in forecasting flu (Ginsberg et al. 2009), economic indicators (Choi and Varian 2012) and private consumption (Vosen and Schmidt 2011). Koop and Onorante (2013) use Google data in a dynamic model selection approach for macroeconomic nowcasting, stressing the fact that including the Google variables in a regression framework might not always be optimal because of the nonlinear dynamics of the attention process. Da et al. (2011) construct a Google search volume Index, showing that it captures the attention of retail investors, while being different from existing proxies for investor attention. They find that the search volume is highly time-varying, rises in periods of high volatility and retail investors are likely to create additional noise in the market. In a study of 30 NYSE and NASDAQ stocks,

[1] as reported by http://www.netmarketshare.com

Vlastakis and Markellos (2012) show, that information demand is positively related to volatility in a GARCH framework. Motivated by these findings Andrei and Hasler (2013) set up a dynamic equilibrium model where variance increases quadratically with investors attention and uncertainty. Especially when they control for lagged volatility, attention is an even more powerful driver of future volatility. This suggests that volatility increases as attention increases, not the other way round. The working mechanism of attention in their model is twofold. Higher attention in bad times accelerates the incorporation of news into prices and increases volatility. Due to higher attention and elevated information demand, investors are more informed, resulting in lower uncertainty, which finally decreases the volatility. Despite these findings on the relationship between volatility and investor attention, the before mentioned articles do not test out-of-sample predictability. Partly, this may be explained by the complexity and number of parameters of the models used for in-sample studies, which makes forecasting rather cumbersome. Dimpfl and Jank (2011) are among the first ones to use daily Google data to forecast daily and weekly realized variances by including it as an additive component in Autoregressive (AR) and Heterogeneous Autoregressive (HAR) models. In their study, Google searches lead to an improvement of in- and out-of-sample performance. However, neither do they show statistically significant forecasting improvements, nor include relevant benchmark models usually applied in forecasting realized variances, such as the ARFIMA (Andersen et al. 2003). In a recent study Vozlyublennaia (2014) studies the interaction between index returns, volatility and Google searches. Using Granger Causality tests in a Vector Autoregressive framework, she shows that investor attention does in many cases have short term and in some cases even long term effects on returns. Yet, the results on volatility are less pronounced, due to the low frequency of her data and as a result, she does not directly investigate the predictability of volatility by investor attention.

Our paper tries to contribute to this recent path of literature by providing a simple and flexible model to study the predictability of volatility by means of investor attention, which takes into account the specific nature of both, volatility and Google search volume time-series data. We argue that, if investor attention causes volatility, a well specified model should allow us to obtain predictions that are significantly more accurate than the ones of conventional volatility models, thus providing harder evidence for the relationship between the two variables. Furthermore, our models dynamic parameters can easily be interpreted, while the number of parameters to estimate remain reasonable small, making the model attractive for economic reasoning as well as practical application.

3. The concept of empirical similarity

Consider an investor being asked to predict a variable of interest Y_n with the help of explanatory variables $X_n = (X_n^1, \dots, X_n^m)$ and a database consisting of historical cases $(X_i^1, \dots, X_i^m, Y_i)$

for $i = 1, \dots, n-1$. Assuming some kind of memory of the investor, it is natural to include the similarity of past situations (states of nature, cases) into the forecasting process. The Empirical Similarity concept (ES) as developed in Gilboa et al. (2006) allows to predict Y_n as a weighted average of all previously observed values Y_i. The weights are determined by the similarity of historical cases $(X_i^1, \dots, X_i^m)$ and the current situation with characteristics $X_n = (X_n^1, \dots, X_n^m)$. This *similarity-weighted averaging* approach is highly related to Case-Based Decision Theory (CBDT), introduced in Gilboa and Schmeidler (1995), which models the human decision making process via the similarity of past situations and the current decision problem. Within the ES approach the data generating process (DGP) for a variable Y_t is modeled as

$$Y_t = Y_t^s = \frac{\sum_{i<t} s\left(X_i, X_t\right) Y_i}{\sum_{i<t} s\left(X_i, X_t\right)} + \varepsilon_t, \tag{1}$$

with $\varepsilon_t \sim N\left(0, \sigma^2\right)$, Y_1 being an arbitrary random variable and $s\left(\cdot,\cdot\right)$ is the similarity function we define next.

$$s\left(x_i, x_j\right) = \exp\left(-\sqrt{\sum_{m \leq M} \omega_m \left(x_i^{(m)} - x_j^{(m)}\right)^2}\right), \quad \text{with} \quad \omega_m \in \mathbb{R}. \tag{2}$$

The similarity function includes weighted Euclidean distances and transforms large distances into small similarities and vice versa. The weighting parameters ω_m can be estimated from the data using maximum likelihood methodology.

Note that the ES model includes all past observations into the DGP and the memory does not decay without additional structure. Though volatility series usually show long memory behavior, it is unlikely to assume that very old volatility observations (e.g. older than one year) influence the value of today's volatility. Consequently, the original ES approach must be adapted for the purposes of volatility modeling. A recently developed ES approach in Golosnoy et al. (2014) shows a possibility to augment HAR-forecasts when the HAR-components are combined exploiting empirical similarity. Another idea described in Lieberman (2012) is to improve the simple AR(1) model. Thereby, the ar_1 coefficient is determined via ES and changes from a static parameter to a time-varying coefficient. Within this framework it is possible to model time-series that show stationary, unit-root and explosive behavior.

In our paper we suggest an ES model which is applicable for our Google Trends data, as it simply compares today's level of investor attention and volatility to perform forecasts for the next period. Hence, it allows us to draw inference for different states of investor attention and volatility. If investor attention is high and volatility is low, we expect future volatility to rise due to increased participation of retail investors in the market. Otherwise, if attention and volatility are both high, we expect the dynamics of volatility to remain unchanged compared to its recent past. Similar, two possibilities exist for the case of low investor attention. If the previous level of volatility was also low, future volatility should remain constant, as no additional participation is to be expected. However, if the previous level of volatility was high, low attention indicates a change point of volatility dynamics, meaning investors are losing interest in the market. In this case, due to decreased participation, future volatility should decrease, too.

We follow the idea of Lieberman (2012) with respect to using the similarity between two different variables rather than the similarity between the same variable across time. Note that the latter one is the conventional framework of empirical similarity, but as shown by Lieberman (2012) the first methodology is statistically valid as well.

Our ES model for the volatility process v_t is of the following form:

$$v_t = \omega_0 + \beta_t\left(g_{t-1}, v_{t-1}; \omega_1\right) v_{t-1} + \varepsilon_t, \quad t = 1, \ldots, n \tag{3}$$

where g_t is the value of the Google Trends data as a measure of investor attention at time t and $\boldsymbol{\omega} = (\omega_0, \omega_1) \in \mathbb{R}$ is an unknown parameter vector, $\beta_t\left(g_{t-1}, v_{t-1}; \omega_1\right)$ is a real-valued, non-negative and non-stochastic function, parametrized by ω_1, and ε_t is a sequence of i.i.d. random variables with zero mean and variance σ^2. The function $\beta_t\left(g_t, v_{t-1}; \omega_1\right)$ measures the similarity between g_{t-1} and rv_{t-1}. When $\beta_t\left(g_{t-1}, v_{t-1}; \omega_1\right)$ is constant for all t, namely if the level of attention is very similar to the level of volatility, the model collapses to the standard AR(1) model

$$v_t = \omega_0 + \omega_1 v_{t-1} + \varepsilon_t, \quad t = 1, \ldots, n. \tag{4}$$

Otherwise for a given t the coefficient on v_{t-1} can be smaller, equal or greater than one and describe a stationary, unit-root or explosive manner respectively. The function β_t is defined as:

$$\beta_t\left(g_{t-1}, v_{t-1}; \omega_1\right) = \exp\left(\omega_1\left(g_{t-1} - v_{t-1}\right)\right) \tag{5}$$

In contrast to the definition in equation 2 the difference between g_t and v_t is not squared, hence our model considers which argument of the similarity function has the greater value (however the ordering of the arguments is unimportant, since the parameter ω_1 is real-valued and a multiplication of ω_1 with -1 is equivalent to a permutation of g_t and v_t). Note that our previous theoretical argumentation on the comparison of attention and volatility for forecasting holds if $\omega_1 > 0$, which is the case as shown in section 7.1.

A corresponding definition of the similarity function as well as asymptotic theory and score based hypothesis tests for this model class can be found in Lieberman (2012).

The ES model can be estimated with Maximum Likelihood methodology by replacing the true unobservable v_t with the realized volatility rv_t from section 4.2, which we use as a proxy for v_t.

4. Data

4.1. Google Trends data

The data on Google search queries are obtained through Google Trends[2], where we use weekly data on search volume from 16.01.2004 – 18.10.2013 (510 obs.) for the keyword "dow". In contrast to weekly financial data which usually covers the period from Monday to Friday, Google Trends search volume ranges from Saturday to next Sunday and is made available on Monday morning. Hence, weekly predictions for volatility in our model can be performed on Monday for the current week, see figure 1.

[Figure 1 is about here.]

We restrict our analysis to search data within the US to limit the bias resulting from different meanings of the keyword or parts of it in other languages. Google Trends measures weekly searches based on a query share, which is the total query volume for the search term within a particular geographic region divided by the total number of queries in that region during the examined period of time. After this normalization, Google divides each observation by the highest value and multiplies by 100, so that all observations lie between 0 and 100. Hence, different time-series of Google Trends data are only comparable if their highest values before scaling are the same. This is especially problematic for the use of daily data, since at the moment, Google Trends restricts the access for daily data to windows of 90 days. Being a rather new specification, the problem does not apply to the previous literature that uses daily data. Due to the normalization and standardization, the windows cannot be merged into one meaningful time-series, which makes many of the results of e.g. Dimpfl and Jank (2011) not applicable in practice. Figure 2 illustrates this problem for the daily case, where the Google Trends series rather resembles a white noise process than an informative predictor. Nevertheless, similar to Dimpfl and Jank (2011), we investigate the possibility to use daily Google Trends data for forecasting volatility in section 7.3.

[Figure 2 is about here.]

To assess the correct specification of our rather general keyword, we compare its correlation to the level of search volume of other terms. Using Google Correlate[3], we find a near perfect correlation with the search terms "dow jones", "dow stock" and "djia". Amongst other highly correlated terms are "current dow", "dow close" and "current dow jones". In general, abbreviations seem to be more likely to be googled and our keyword "dow" has the highest search volume amongst all alternatives. Hence we restrict ourselves to this selection. Considering alternatives for the Dow Jones, we share the findings of Dimpfl and Jank (2011), which point out that, e.g. using the S&P 500 is not suitable due to the lower search volume as well as the use of the acronym "S&P" for rating downgrades.

[2] Source: http://www.google.com/trends

[3] Source: http://www.google.com/trends/correlate

4.2. Realized Variance for the DJIA

We use data provided by the Oxford-Man Institute, in particular the daily realized variance series for the Dow Jones based on 5 minutes returns, the standard measure for daily volatility[4]. Further information on data preparation and cleaning can be found in the documentation of the Oxford-Man Institute[5]. We aggregate the data to receive weekly volatility estimators. The chosen data covers the period from 16.01.2004 - 18.10.2013 (510 weekly observations). This period is characterized by a calm volatility period until the end of 2007 and high volatility phase in the second part from the beginning of 2008. For this reason, we pay attention to a subperiod from 04.01.2008 - 18.10.2013 (303 weekly observations) which covers the recent subprime crisis, the Fukushima nuclear disaster and the Eurozone crisis. The basic statistic properties of the RV series and the Google data are summarized in table 1 for the full sample and subsample.

[Table 1 is about here.]

Additionally, the p-values of the Augmented Dickey-Fuller test for stationarity are provided. All time-series appear to be stationary. The statistics support the common stylized facts about realized volatility. The DJ series shows strong and persistent memory, a high kurtosis and is heavily right skewed. The average level of volatility is higher in the crisis period. Surprisingly the Google series shows a very similar behavior, according to memory, skewness and kurtosis. Indeed figure 3, showing the plots and autocorrelation functions for the RV and Google series, makes the similarity between these two time-series apparent, especially for the crisis period. Both series exhibit a collateral behavior and the autocorrelation during the crisis phase decays faster than for the full sample.

[Figure 3 is about here.]

4.3. Relationship between investor attention and realized volatility

The high extent of similarity between the series support the idea of an information content within the Google data. However this can only be the case if Google search queries are not subsequent. Consider the case that a high degree of volatility attracts the investors attention and leads to a high search volume. In this case the information of Google data cannot be exploited for volatility forecasting. If instead, the majority of investors seeks to gain information about the market situation prior to possible transactions, then search volume would contain a valuable information content for volatility forecasting. To asses this issue figure 4 shows the cross correlation between RV and Google data for the full sample (left) and the crisis subperiod (right).

[Figure 4 is about here.]

[4] We use the term realized volatility for the series of realized variances, as these terms are mostly used synonymously

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

The line in the middle of each picture shows the correlation of the RV and Google series for lag=0, which is unsurprisingly the highest value (higher than 0.8). The lines on the left hand side of each picture show the correlation between RV and lagged values of the Google series, the opposite holds for the right hand side. A closer look at figure 4 shows that the correlation between RV and lagged values of Google data is explicitly higher than the correlation between the Google series and lagged RV values, if the lag is smaller or equal to 6. This finding holds for both sample periods. Hence, for small lags (our ES model uses lag 1 in which case the correlation is still about 0.8) the Google data appears to show predictive abilities, supporting the second theory about the interrelation of Google searches and market activity.

In order to control for the effects of previous levels of attention and volatility on the contemporaneous relationship, we use a VAR model similar to the one of Vozlyublennaia (2014). Additionally, this allows us to assess the sign and timing of the effects between investor attention and volatility. The model can be written as

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \cdots + \alpha_k X_{t-k} + \varepsilon_t, \tag{6}$$

where X_t contains the realized volatility rv_t and the Google data g_t. Regarding the lag length k, an analysis up to 10 lags revealed no more significant parameters and a decrease in adjusted R^2 of the regressions after a number of $k = 4$ lags. Results on the estimated parameters are reported in table 2, where the first column gives the regression coefficients with realized volatility as dependent variable, while the second column gives the results for search volume as dependent variable.

[Table 2 is about here.]

For volatility, search volume has an significant impact for lags 1 and 4. While the short term (weekly) influence of lag 1 is positive, the long term (monthly) effect of lag 4 points in the opposite direction, which emphasizes the rapid dynamics of both series, where signs are likely to change on different horizons. Vozlyublennaia (2014) discovered similar significances for the same lags in the case of index returns and investor attention. Altogether, this is consistent with the theory of Barber et al. (2009), where investors create short term price pressure, leading to higher volatility. In the long-run a trend reversal is expected. For example, in periods of high volatility, future prices are expected to return to fundamental values, which coincides with low volatility. Investor attention on the other hand, seems to be mainly driven by short term to mid-term volatility, as lags 1 and 3 in the regression of volatility on the time-series of Google search volume are highly significant. Here, the signs on the coefficients are reversed in comparison to the previous regression equation. We interpret these results as follows. Once investors information demand increases, volatility is likely to rise due to their participation in the market. However, once they participated, less information is required and future search volume is expected to decline, while volatility still being high. Despite supporting our assumptions on the dynamics between investor attention and volatility, the VAR analysis remains in-sample based and models from this class suffer from being not suitable for forecasting due to the very large number of parameters. Similar to the correlation analysis, even if the model controls for lagged volatility, a lag number of 1 for the Google search volume seems to be a reasonable choice to study the short term impact in an out-of-sample forecasting framework. Based on these insights, we suggest several benchmark models more suitable for predicting volatility in the following section, which will finally be augmented by a Google component in section 5.3 and used as benchmark models in the forecasting analysis.

5. Benchmark volatility models

5.1. HAR model

The recently proposed heterogeneous autoregressive (HAR) model of Corsi (2009) is an additive cascade model of volatility components defined over different time horizons and appears to be a very successful approach in capturing the dynamic features of volatility time-series. It can be seen as an parsimonious approximation of more sophisticated MIDAS (Ghysels et al. 2006) and ARFIMA volatility models (Andersen et al. 2003). The different volatility measures (components, aggregates) are combined in a simple linear regression framework. The HAR approach allows to mimic many stylized facts of volatility time-series and hence provides reasonable out-of-sample forecasting results in applications.

The HAR model for the volatility process v_t is given as

$$v_t = \alpha_0 + \omega_1 v_{t-1}^{(s)} + \omega_2 v_{t-1}^{(m)} + \omega_3 v_{t-1}^{(l)} + \varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} (0, \sigma^2), \tag{7}$$

where $v_{t-1}^{(s)} = v_{t-1}$ is short term, $v_{t-1}^{(m)}$ and $v_{t-1}^{(l)}$ are the average mid- and long term volatility measures, respectively. They are defined as $v_t^{(m)} = 5^{-1} \sum_{i=1}^{5} v_{t-i+1}$ and $v_t^{(l)} = 22^{-1} \sum_{i=1}^{22} v_{t-i+1}$. The HAR model can be estimated within a standard OLS regression framework by replacing the true unobservable v_t with the realized volatility rv_t. The standard HAR model uses daily-, weekly- and monthly volatility aggregates to model daily volatility. In our paper we are interested in forecasting weekly volatility and therefore use a HAR-type model with weekly-, 5-weekly and 22-weekly volatility aggregates. (In other words we replace the daily component in the standard HAR model by the weekly one and the aggregation horizons remain constant). See Andersen et al. (2007) and Byun and Kim (2013) for further approaches to predict weekly volatilities within a modified HAR approach. There are two possible economic interpretations of the volatility components. On the one hand, the long term component $v^{(l)}$ can be related to fundamental macroeconomic uncertainty factors. The medium-term component $v^{(m)}$ is assumed to reflect the current market uncertainty concerning the evaluation of news, whereas the short term component $v^{(s)}$ accounts for the speculative momentum uncertainty. On the other hand, the components can be mapped to the different investment horizons of market participants. Financial markets consist of long term investors that leave their portfolios constant for several months, mid-term investors acting monthly and short term investors like hedge funds that tend to rebalance their portfolios on a weekly basis or even more frequently.

5.2. ARFIMA

Opposing to the HAR model the autoregressive fractional integrated moving average (ARFIMA) model as applied in (Andersen et al. 2003) lacks an econometric interpretation. However, it is suitable for volatility modeling since the autocorrelation function of ARFIMA processes has a slow hyperbolical decay matching the autocorrelation behavior in volatility time-series. The slow decay of the autocorrelation function in ARFIMA models is the reason why ARFIMA is called the long memory model.

The standard ARFIMA model for the volatility process v_t with parameters (p, d, q) is denoted by:

$$\Phi(L)(1-L)^d(v_t - \mu) = \Theta(L) + \varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} (0, \sigma^2), \tag{8}$$

where L is the backward-shift operator, $\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$, $\Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$, and $(1-L)^d$ is the fractioning differencing operator defined by

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d) L^k}{\Gamma(-d)\Gamma(k+1)} \tag{9}$$

with $\Gamma(\bullet)$ denoting the gamma function. Generally the parameter d is allowed to be any real value, however in volatility modelling d is restricted to the interval $(0, 0.5)$, since $d > 0.5$ results in non-stationary time-series. The parameters can be estimated via maximum likelihood methodology. In our study we restrict ourselves to the conventionally used ARFIMA$(1, d, 0)$ model. Both, the ARFIMA as well as the HAR model are the most commonly used models for modeling and forecasting volatility, being able to model its stylized facts (clustering, long memory) while retaining parsimony (see e.g. Corsi (2009)).

5.3. Additional models including Google Trends

Based on the strong linear relationship of rv_t and g_t (see section 4.3), a natural extension of the previous models in section 5.1 and 5.2 is the inclusion of a linear Google Trends component in the regression (see e.g. Dimpfl and Jank (2011)). We include four additional models into our analysis. First, an AR(1) model (referred to as simply AR in the figures and tables). Second, an AR(1) which we extend with an additive Google Trends component of lag $t-1$ and call it AR G. Third, both ARFIMA and HAR models are extended by the same lagged regressor and we call these models ARFIMA G respectively HAR G.

Despite being a common extension, we do not advocate the use of a Google Trends component in a simple regression framework as, in our opinion, it does only partly reflect the nature of the data. While the inclusion of a Google Trends component makes sense from our cross-correlation analysis and leads to highly significant in-sample estimates (see our analysis in section 7.1), a specific problem of conventional volatility models lies in the mixture of long run and short run dynamics. In general, both components of fractional processes are driven by the same innovations, see e.g. Sun and Phillips (2003). Hence, rapid adaption when changes from calm to volatile phases and vice versa occur, can be problematic for forecasting with long-memory models with additional regressors. In our data, especially in highly volatile phases we observe an explosive behavior of both RV and Google data (see figure 3). The HAR partly overcomes this problem by its additive structure of volatilities over different time horizons, but we expect the inclusion of a short term Google Trends component for the ARFIMA to be more perturbing in forecasting due to the inevitability of mixing dynamics of different horizons.

In opposition to the aforementioned time-series models, the ES model is able to model unit-root and non stationary behavior by changing the static ar_1 coefficient w_1 in equation 4 to a time-varying coefficient β_t, see equation 5. Both parameters and the time-series of realized volatility and Google Trends are depicted in figure 5 for the the crisis period.

[Figure 5 is about here.]

While being stable with little fluctuation during calm periods, β_t is highly dynamic in both subsample periods associated with the largest shocks in the market (August to October 2008 and June 2011), where we observe an increase in the coefficient. When volatility drops after the first shock around January 2009, the parameter declines considerably, which is a sign of adaption of the model from very high levels of volatility to lower ones. Remarkably, even during the periods of high volatility, β_t shows signs of very fast adaption and variation, which resembles the dynamics we observe in the original time-series of volatilities. This behavior is in line with the basic economic idea of the model (see section 3). If volatility and search volume is high, β_t increases explosively, thus giving more weight to past volatilities and resulting in a rise in predicted future volatility. Once a certain level of volatility is reached and investor attention decreases, β_t falls to levels smaller than 1 and future volatility is expected to fall. In cases of low volatility and low attention or high volatility and high attention, the process resembles an AR(1). Hence, we expect the ES model to particularly differ from the conventional models in its forecasting performance during the periods where the parameter shows large deviations in either direction, namely the periods around August to October 2008 and June 2011.

6. Loss functions and the MCS

For comparing forecasts from various models without using a benchmark, we rely on the Model Confidence Set (MCS) approach of Hansen et al. (2011). The MCS allows us to obtain a selection of models, that will contain the best model at a given level of confidence without using the particular assumption of a “true” model. In contrast to pairwise comparisons or selecting a single model, the MCS acknowledges the informational content of the data by the possibility of containing several models with equivalent forecasting ability. Uninformative data produces a MCS with many models, while informative data results in only a few models included in the MCS. Further, a model is only removed from the MCS if it is significantly inferior to another model, hence the MCS comparison is more robust and less prone to data mining biases.

The MCS evaluates the forecasting ability of our models solely based upon the forecast error under a certain loss function. Regarding the choice of this loss function, we rely on the work of Laurent et al. (2013) and Patton (2011). They show that it is crucial for forecasting volatility to select a loss function from the family of consistent loss functions, where consistency means that the true ranking of the models is preserved, regardless if the true conditional volatility or a noisy volatility proxy is used. Among the class of consistent loss functions are the mean-squared-error (MSE) and the asymmetric quasi-likelihood (QLIKE) loss function, which penalizes underpredictions more heavily.

Regarding the MCS, we limit ourselves to the QLIKE loss, as Patton and Sheppard (2009a) find that likelihood based loss functions possess greater power in distinguishing between volatility forecasts compared to MSE loss functions. Beside the statistical properties, using an asymmetric loss function is relevant from an investor’s point of view, as West et al. (1993) show that an underestimation of variances in the asset allocation process leads to lower expected utility than the equivalent overestimation, Hence, a stronger punishment of underpredictions of volatility is consistent with rationale behavior under risk aversion. Evaluating the economic value of forecast accuracy, Taylor (2014) shows that costs associated with forecast imperfections are higher for models that tend to underpredict. As volatility is higher in market downturns than upturns (see e.g. Bekaert and Wu (2000)) underpredictions are especially costly in an already critical market environment.

The QLIKE is defined as

$$L\left(rv_t,\widehat{rv}_t\right) = \frac{rv_t}{\widehat{rv}_t} - \ln\frac{rv_t}{\widehat{rv}_t} - 1, \tag{10}$$

where rv_t is the actual realized volatility at time t and $\widehat{rv}_t$ the forecasted realized volatility for this point of time.

For the general procedure of the MCS, we start with the full set of candidate models $\mathcal{M}_0 = \{1,\ldots,m_0\}$, where in our case $m_0 = 7$. For all models the loss differential between each model is computed based upon the loss function L, so that for model i and j, $i{,}j = 1,\ldots,m_0$ and every time point $t = 1,\ldots,T$ we get:

$$d_{ij,t} = L\left(rv_{it}, \widehat{rv}_{it}\right) - L\left(rv_{jt}, \widehat{rv}_{jt}\right). \tag{11}$$

At each step of the evaluation, the hypothesis

$$H_0 \ : \ E\left[d_{ij,t}\right] = 0, \quad \forall i,j \in \mathcal{M}, i > j, \tag{12}$$

is tested for a subset of models $\mathcal{M} \in \mathcal{M}_0$, where $\mathcal{M} = \mathcal{M}_0$ for the initial step. If H_0 is rejected at a given significance level (e.g. 10%), the worst performing model is removed from the set. This process continues until a set of models remains that cannot be rejected. We follow the work of Hansen et al. (2011), who use the range statistics to evaluate H_0, which can be written as:

$$T_{R,k} = \max_{i,j \in \mathcal{M}} |t_{ij}| = \max_{i,j \in \mathcal{M}} \frac{|\bar{d}_{ij}|}{\sqrt{\widehat{var}(\bar{d}_{ij})}}, \tag{13}$$

where $\bar{d}_{ij} = \frac{1}{T}\sum_{t=1}^{T} d_{ij}$ and $\widehat{var}(\bar{d}_{ij})$ is obtained from a block-bootstrap procedure, see Hansen et al. (2011). We implement the bootstrap with 10000 replications and a variable block length ranging from 12 to 50, finding that the results are robust to the choice of block length.

As a level of confidence, at which the best model is included in the MCS, we rely on the conventional choice of 90% (see Patton and Sheppard (2009b) and Liu et al. (2012)).

The worst performing model to be removed from the set $\mathcal{M}$ is selected as model i^* with

$$i^* = \arg\max_{i \in \mathcal{M}} \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}}, \tag{14}$$

where $\bar{d}_i = \frac{1}{m-1}\sum_{j \in \mathcal{M}} \bar{d}_{ij}$ and m being the number of models in the actual set $\mathcal{M}$.

7. Empirical Illustration

7.1. In-sample fit

The intention behind the empirical investigation is to illustrate the benefit of applying Google Trends data in an ES model for weekly volatility forecasting of the DJ index. At first, we examine the in-sample fit for the whole data length and subsequent for the crisis period[6].

The parameter estimates, the corresponding standard errors in parenthesis and the residual standard deviation are given in table 3 for the full sample (top) and the crisis subsample (bottom).

[Table 3 is about here.]

[6] Our choice of the crisis phase is motivated by the increase in volatility in the year 2008 similar to Bekaert et al. (2011).

Surprisingly the parameter values change only slightly from the full sample to the subsample for all models, while the addition of the Google regressor leads to many parameter adaptions, especially in the HAR model. There the long term volatility component ω_3 changes from an insignificant to a highly significant variable. Beside the intercepts all other parameters are significant at a 5%−level, only ω_1 in the ARFIMA models has significances around a 10%−level. However, the estimation results for the ARFIMA models without Google components are unsteady with the differencing parameter being close to 0.5. Table 4 shows the losses for the respective models and loss functions.

[Table 4 is about here.]

The ES approach augments the standard AR(1) model for both samples. Of course according to the MSE measure all models benefit from the Google regressor, since the OLS estimation optimizes under a MSE loss function. Regarding the QLIKE distance including Google data as additional regressor heavily worsens the fit. The ES approach performs better in the crisis phase, where all other volatility models without Google regressors are heavily outperformed. The goodness of fit enhancement ranges from 5.3% (QLIKE) for the HAR model to 8.3% (MSE) for the AR(1) and ARFIMA model.

7.2. Out-of-sample forecasts

Forecasting is performed using a moving window of 250 obs., where the first forecast is made for the week from 27.10.2008 − 31.10.2008. We deliberately include the beginning of the subprime crisis (August 2008) in our first estimation window for two reasons. First of all, the ARFIMA based models need a reasonable number of observations for the estimation to be reliable. Second, since we do not make any assumption on the (unobservable) fundamental shock that leads to rising interest in the markt all models should be equally biased when such a shock occurs unexpectedly.

All models are re-estimated on a weekly basis, which results in overall 260 forecasts. These forecasts are then compared using the MSE and QLIKE loss functions from section 6. In addition, we perform Mincer-Zarnowitz (MZ) (Mincer and Zarnowitz 1969) regressions of the actual realized volatilities on their predicted values:

$$rv_t = \alpha_{mz} + \beta_{mz}\widehat{rv}_t + \varepsilon_t. \tag{15}$$

In table 5, we report the average loss in relation to the ES model as well as the parameters and R^2 of the MZ regression for each model.

[Table 5 is about here.]

For both loss functions and the R^2, the ES model is leading the ranking, increasing the average forecasting performance in comparison to the ARFIMA model about 6% for both loss functions. Adding a google component does improve the forecasting performance of the HAR model slightly, however this is not the case for the AR and ARFIMA model. As mentioned before, this may be explained by the inability of the ARFIMA G to mix long run and short run behavior simultaneously. Regarding the parameters of the MZ regression, the desired values of $\alpha_{mz} = 0$ and $\beta_{mz} = 1$ are nearly reached for the ES model, while most other models show large deviations in at least one of the parameters. However, as Patton (2011) points out, using the MZ regression for the evaluation of volatility forecasts is not reliable if, like in our case, a volatility proxy is used instead of the true volatility process. Hence we do only report the results for comparative reasons and instead proceed by evaluating the forecasts by the consistent MCS approach from section 6.

We apply the MCS on the whole out-of-sample period, as well as two crisis subsamples. For the full sample, the corresponding levels of confidence of the MCS, at which the model can be removed from the confidence set, are found in table 6.

[Table 6 is about here.]

For example, at a 90% confidence level, all models with a reported value > 0.1 will be included in the MCS. The ES model is always the leading model in the confidence set, improving forecasts significantly upon the other models, given a 90% confidence level. For all models except the HAR, we cannot find that adding a Google component improves forecasting performance significantly.

We are especially interested in the differences between highly volatile and calm phases in our out-of-sample study. Usually, the literature assumes the presence of retail investors to increase volatility and uncertainty in the market. Hence, we expect a period of high volatility and high investor attention, as measured by the Google Trends component, to be followed by high volatility. On the other hand, periods of low volatility might not induce sufficient investor attention and the Google Trends component may not contain additional information that is useful for forecasting. Overall, the gain in forecasting performance due to the use of a Google component should be especially obvious when the general level of volatility is high.

In our out-of-sample data, we observe the highest level of volatility around October 2008 (subprime crisis) and June 2011 (Eurozone crisis). We select a window of 100 out-of-sample forecasts beginning with these dates and apply the MCS procedure. Figure 6 shows the realized volatility, the Google Trends data and the both time periods.

[Figure 6 is about here.]

In both cases, the ES model is leading in the MCS. Corresponding levels of confidence can be found in column 2 and 3 of table 6. This further confirms our hypotheses of section 5.3, that the time-varying parameter of the ES model might be particularly important during phases of high volatility.

For the first period, the AR(1) model is the second best choice, while for the second period, several models including the ARFIMA and HAR models are closely included in the MCS at a 90%

confidence level. This might be due to the relatively fast fading of the second crisis compared to the first one, where the models performance is less influenced by the Google component. Further, the behavior is in line with the general theory of the MCS. Due to the differences in forecasts being larger in the first period, more models are rejected from the MCS. In the second period, there are less pronounced differences in the forecasts, hence more models remain in the MCS. Overall, beside the ES being the leading model, the ARFIMA model shows a somehow consistent performance, being included at a 90% confidence level in both periods, despite its weaker performance during the first period, where it is close to rejection.

7.3. The daily-data problem

As pointed out in section 4.1, the daily Google Trends data is problematic due to its standardization. Nevertheless, we analyze the predictive ability of daily data for two reasons. First, since previous research used non standardized data and found predictability on the daily level, we want to assess how serious this new restriction is to the practical use of the Google Trends data. Second, if the performance of the ES model might purely result from the time-variability in the ar_1 coefficient, the model should show similar predictive accuracy for daily and weekly horizons, regardless of the information content of the data. Therefore, we carry out the previous out-of-sample analysis for the case of daily observations from 27.10.2008 – 18.10.2013 (1206 obs.) using a moving window of 1250 observations, which corresponds roughly to the time horizon used for weekly data. Table 7 shows the average loss of each model in relation to the ES model.

[Table 7 is about here.]

In all cases, the ARFIMA based models have the smallest average losses, closely followed by the HAR models and third the AR models, while the ES model performs worst. The inclusion of a daily Google component leaves most of the additional models unchanged from their base models. We can deduce that even being inferior to the AR(1) base model, pure superiority due to the dynamic coefficients does not seem to be an issue for the ES model. Overall, the inclusion of daily data does not lead to better out-of-sample forecasts and limits the usability of daily Google Trends data in practice. Since the differences are evident even based on the forecast errors, we do not further proceed by testing them statistically using the MCS.

7.4. VaR-forecasting

As an extension to our statistical analysis of the weekly forecasts, we want to asses possible benefits of Google Trends data for Risk-Management with a Value-at-Risk (VaR) forecasting exercise. Since banks are required to report their VaR at a 99% confidence level daily over the one-day and two-week horizon for estimation of their capital requirements, one of the usual methodologies includes scaling daily VaR forecasts to the longer horizon, see e.g. Cuoco and Liu (2006). However, this methodology relies heavily on the normality of returns and is prone to the underestimation of risks, see Danielsson and Zigrand (2006). Hence, weekly VaR forecasts gives an intermediate approximation of the mid-term risk and is suitable for the economic comparison of the models forecasting performance. In practice, the method could easily be extended to the required two week horizon.

We can calculate $\mathrm{VaR}_t(\alpha)$ forecasts with respect to the empirical return distribution f_t of r_t. Following Lopez and Walter (2001), we separate the variance dynamics $\widehat{rv}_{t+1}$ and the distributional form of $\hat{f}_{t+1}$. Hence the VaR forecast for $t+1$ is given by

$$\widehat{VaR}_{t+1}(\alpha) = \hat{\mu}_{t+1} + \sqrt{\widehat{rv}_{t+1}}\hat{F}^{-1}_{t+1}(\alpha), \tag{16}$$

where $\hat{F}^{-1}_{t+1}$ is the inverse of the (historical) cumulative distribution function of standardized returns. For reasons of simplicity, we assume that the empirical return distribution until t is valid for $t+1$, $\hat{f}_{t+1} = f_t$. Further, we use the empirical mean until t as an estimator for $\hat{\mu}_{t+1}$.

Given the VaR forecasts, for each model we define an exceedance of the VaR if $\widehat{VaR}_{t+1} < r_{t+1}$. We use the test of unconditional coverage (UC) by Kupiec (1995) to asses if the number of VaR exceedances differs significantly from $\alpha \cdot 100\%$ within the sample. To evaluate the loss size, we calculate the squared loss of the VaR over the ex-post return for the whole out-of-sample period. The number of exceedances (percentage and absolute), as well as the p-value of the UC test and the squared loss of the VaR over the ex-post return (relative to the ES loss) can be found in column 1 to 4 of table 8.

[Table 8 is about here.]

For all models, the H_0 of the UC test cannot be rejected at a 5% significance level, meaning that the number of exceedances is not significantly different from the expected number of $0.01 \cdot 260 = 2.6$ violations. Nearly all models yield a reasonable number of 3 to 4 VaR exceedances, only the HAR G produces too conservative predictions and yields 1 exceedance. Regarding average squared loss over the whole sample, all models that produce 3 exceptions show very little difference. For the ARFIMA and HAR model, including a Google component leads for fewer exceptions, but higher losses, which is puzzling at first, but easily explained by our next method of comparison.

From a bank's perspective, an optimal VaR forecasting model should not only accurately forecast the point of exceedance, but also the level of the forecasted VaR is relevant. The larger the forecasted VaR in relation to the ex-post return, the more capital is required to be held back unnecessarily due to general overprediction. A symmetric loss, like the MSE used before will not be optimal in this case, as over- and underpredictions are punished equally and no clear distinction can be made. Therefore, we use two reversed versions of the QLIKE loss function in equation 10, namely

$$L\left(\widehat{VaR}_{t+1}, r_{t+1}\right) = \frac{\widehat{VaR}_{t+1}}{r_{t+1}} - \ln\frac{\widehat{VaR}_{t+1}}{r_{t+1}} - 1, \tag{17a}$$

and

$$L\left(\widehat{VaR}_{t+1},\widehat{VaR}_{t+1,oracle}\right)=\frac{\widehat{VaR}_{t+1}}{\widehat{VaR}_{t+1,oracle}}-\ln\frac{\widehat{VaR}_{t+1}}{\widehat{VaR}_{t+1,oracle}}-1, \tag{17b}$$

Equation 17a measures the performance of the VaR forecast in relation to the ex-post return, which is a measure of capital reserves to be held back. Equation 17b uses the oracle VaR forecast, based upon the true rv_{t+1} in equation 16, for measuring the general forecasting accuracy of the VaR forecast. In both cases, overpredictions of the respective target variable lead to higher losses than underpredictions. As it can be seen in the last two columns of table 8, relative to the loss of the ES model, all others models tend to overpredict the true VaR. The gains in forecasting accuracy range from 1% to 2% compared the second best model (ARFIMA). For the other models, overpredictions heavily influence the losses, e.g. the HAR G forecasts are 28%to 55% worse than the forecasts of the ES model. This seems to be a plausible explanation for the small number of VaR exceedances in case of the HAR G. Despite fewer VaR violations, more capital is unnecessarily needed to fulfill regulatory requirements, making these models less attractive in VaR applications. Including the Google component in the ARFIMA and HAR framework seems to rather worsen the results in terms of general forecast accuracy.

Due to the suitability of the MCS for more general objects of comparison (see Hansen et al. (2011)), we also apply it for an analysis of the significance of our VaR forecasting results, based upon the loss functions in equation 17a and 17b. Table 9 gives the p-values, at which the model can no longer be included in the MCS.

[Table 9 is about here.]

For both loss functions, the ES is the leading model, followed by one of the ARFIMA models at a 90% level of confidence, which coincides with our previous results. While forecasts of the ARFIMA in relation to the benchmark are not significantly worse than for the ES, the ARFIMA requires more capital reserves, as its VaR forecasts tend to exceed the ex-post-return. We conclude, that while there are only moderate differences in the number of forecasted VaR exceedances, the ES significantly differs from the other models in its forecasting accuracy in relation to the optimal (oracle) forecast, as well as the size of capital reserves to be held back. For both measures, the ES model delivers the most consistent performance in the MCS.

8. Conclusions

This paper introduced an economically motivated model for using Google search frequency data to forecast volatility, based on the concept of empirical similarity. We studied the possible benefits on a daily and weekly horizon. For the weekly horizon our ES model showed significant better performance compared to traditional models in an in-sample as well as an out-of-sample forecasting study. Our results confirm the findings of Vlastakis and Markellos (2012), Andrei and Hasler (2013) and Vozlyublennaia (2014), who state that investor attention is a driver of volatility on short horizons. As described by Andrei and Hasler (2013), this relationship is strongest in phases of high volatility, where investor attention tends to be high. This coincides with our results of superior predictive ability of the ES model in crisis periods. We found the use of daily data not to be suitable for our volatility forecasting analysis due to the method Google limits the access to this data in standardized 90 day windows. We show, that while linear models can be useful for assessing the correlation of volatility and investor attention and studying their dependence in an in-sample framework, these models are not flexible enough when it comes to forecasting. For both horizons, the inclusion of Google Trends data as simple additive term in classical RV models did not improve forecasting accuracy. These findings partly contradict the results of Dimpfl and Jank (2011), who used daily unstandardized Google Trends data and find predictive ability. We assume that these differences arise through the way Google transforms the data. Hence, future practical application is limited to weekly data, which we study exemplary by predicting the weekly Value-at-Risk (VaR). Here, the ES model produced significantly better VaR forecasts in terms of overall accuracy and required capital, while providing an adequate number of VaR violations. Based on theoretically well documented assumptions on the relation of investor's behavior and stock market reaction, our model provides an intuitive way of studying how investors attention influences volatility. Due to its simplicity and forecasting performance, the model is highly beneficial in risk management applications.

References

Alfarano, Simone, Thomas Lux. 2007. A Noise Trader Model as a Generator of apparent Financial Power Laws and Long Memory. *Macroeconomic Dynamics* **11**(S1) 80–101.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold. 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics* **89**(4) 701–720.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, Paul Labys. 2003. Modeling and forecasting realized volatility. *Econometrica* **71**(2) 579–625.

Andrei, Daniel, Michael Hasler. 2013. Investor attention and stock market volatility. Working paper.

Barber, Brad M., Terrance Odean, Ning Zhu. 2009. Do Retail Trades Move Markets? *Review of Financial Studies* **22**(1) 151–186.

Bekaert, Geert, Michael Ehrmann, Marcel Fratzscher, Arnaud Mehl. 2011. Global crises and equity market contagion. Working paper.

Bekaert, Geert, Guojun Wu. 2000. Asymmetric volatility and risk in equity markets. *Review of Financial Studies* **13**(1) 1–42.

Brooks, Chris. 1998. Predicting stock index volatility: can market volume help? *Journal of Forecasting* **17**(1) 59–80.

Byun, Suk Joon, Jun Sik Kim. 2013. The information content of risk-neutral skewness for volatility forecasting. *Journal of Empirical Finance* **23**(0) 142 – 161. doi:http://dx.doi.org/10.1016/j.jempfin.2013.05.006.

Candelon, Bertrand, Gilbert Colletaz, Christophe Hurlin, Sessi Tokpavi. 2011. Backtesting Value-at-Risk: A GMM duration-based test. *Journal of Financial Econometrics* **9**(2) 314–343.

Choi, Hyunyoung, Hal Varian. 2012. Predicting the present with google trends. *The Economic Record* **88**(s1) 2–9.

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

Cuoco, Domenico, Hong Liu. 2006. An analysis of VaR-based capital requirements. *Journal of Financial Intermediation* **15**(3) 362–394.

Da, Zhi, Joseph Engelberg, Pengjie Gao. 2011. In search of attention. *Journal of Finance* **66**(5) 1461–1499.

Danielsson, Jon, Jean-Pierre Zigrand. 2006. On time-scaling of risk and the square-root-of-time rule. *Journal of Banking & Finance* **30**(10) 2701–2713.

Dimpfl, Thomas, Stephan Jank. 2011. Can internet search queries help to predict stock market volatility. Working paper.

Donaldson, R. Glen, Mark J. Kamstra. 2005. Volatility forecasts, trading volume, and the arch versus option-implied volatility trade-off. *Journal of Financial Research* **28**(4) 519–538.

Ghysels, Eric, Pedro Santa-Clara, Rossen Valkanov. 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* **131**(1-2) 59 – 95.

Gilboa, Itzhak, Offer Liebermann, David Schmeidler. 2006. Empirical similarity. *Review of Economics and Statistics* **88**(3) 433 – 444.

Gilboa, Itzhak, David Schmeidler. 1995. Case-based decision theory. *The Quarterly Journal of Economics* **110**(3) 605–639.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, Larry Brilliant. 2009. Detecting influenza epidemics using search engine query data. *Nature* **457** 1012–1014.

Golosnoy, Vasyl, Alain Hamid, Yarema Okhrin. 2014. The empirical similarity approach for volatility prediction. *Journal of Banking & Finance* **40**(0) 321 – 329.

Golosnoy, Vasyl, Yarema Okhrin. 2008. General uncertainty in portfolio selection: A case-based decision approach. *Journal of Economic Behavior & Organization* **67**(3-4) 718–734.

Hansen, Peter R., Asger Lunde, James M. Nason. 2011. The model confidence set. *Econometrica* **79**(2) 453–497.

Kim, Soon-Ho, Dongcheol Kim. 2014. Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization* (in press) –. doi:http://dx.doi.org/10.1016/j.jebo.2014.04.015. URL `http://www.sciencedirect.com/science/article/pii/S0167268114001206`.

Koop, Gary, Luca Onorante. 2013. Macroeconomic nowcasting using google probabilities. Working paper.

Kupiec, Paul H. 1995. Techniques for verifying the accuracy of risk management models. *Journal of Derivatives* **3** 73–84.

Laurent, Sebastien, Jeroen V. K. Rombouts, Francesco Violante. 2013. On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics* **173**(1) 1–10.

Li, Dan, Geng Li. 2014. Are Household Investors Noise Traders: Evidence from Belief Dispersion and Stock Trading Volume. *Finance and Economics Discussion Series 2014-35, Board of Governors of the Federal Reserve System (U.S.)* .

Lieberman, Offer. 2012. A similarity-based approach to time-varying coefficient non-stationary autoregression. *Journal of Time Series Analysis* **33**(3) 484–502.

Liu, Lily, Andrew J. Patton, Kevin Sheppard. 2012. Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes. Working paper.

Lopez, Jose A., Christian A. Walter. 2001. Evaluating covariance matrix forecasts in a value-at-risk framework. *Journal of Risk* **3**(3) 69–98.

Lux, Thomas, Michele Marchesi. 1999. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature* **397** 498–500.

Mincer, Jacob A., Victor Zarnowitz. 1969. The evaluation of economic forecasts. *Economic forecasts and expectations: Analysis of forecasting behavior and performance*. NBER Chapters, National Bureau of Economic Research, Inc, 1–46.

Patton, Andrew J. 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* **160**(1) 246 – 256.

Patton, Andrew J., Kevin Sheppard. 2009a. Evaluating volatility and correlation forecasts. Thomas Mikosch, Jens-Peter Krei, Richard A. Davis, Torben Gustav Andersen, eds., *Handbook of Financial Time Series*. Springer Berlin Heidelberg, 801–838.

Patton, Andrew J., Kevin Sheppard. 2009b. Optimal combinations of realised volatility estimators. *International Journal of Forecasting* **25**(2) 218 – 238.

Preis, Tobias, Daniel Reith, H. Eugene Stanley. 2010. Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society A* **368** 5707–5719.

Siganos, Antonios, Evangelos Vagenas-Nanos, Patrick Verwijmeren. 2014. Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization* (in press) –. doi:http://dx.doi.org/10.1016/j.jebo.2014.06.004. URL http://www.sciencedirect.com/science/article/pii/S0167268114001735.

Sun, Yixiao, Peter C. B. Phillips. 2003. Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics* **115**(2) 355–389.

Taylor, Nicholas. 2014. The economic value of volatility forecasts: A conditional approach. *Journal of Financial Econometrics* **0**(0) 1–45.

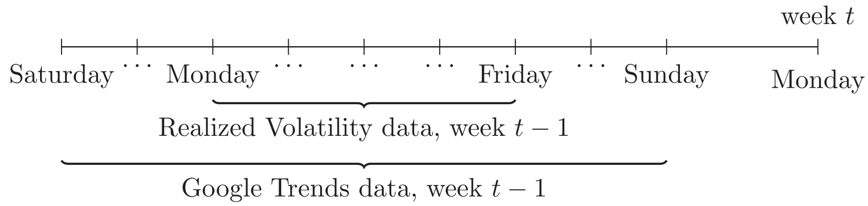
Vlastakis, Nikolaos, Raphael N. Markellos. 2012. Information demand and stock market volatility. *Journal of Banking & Finance* **36**(6) 1808–1821.

Vosen, Simeon, Torsten Schmidt. 2011. Forecasting private consumption: survey-based indicators vs. google trends. *Journal of Forecasting* **30**(6) 565–578.

Vozlyublennaia, Nadia. 2014. Investor attention, index performance, and return predictability. *Journal of Banking & Finance* **41**(C) 17–35.

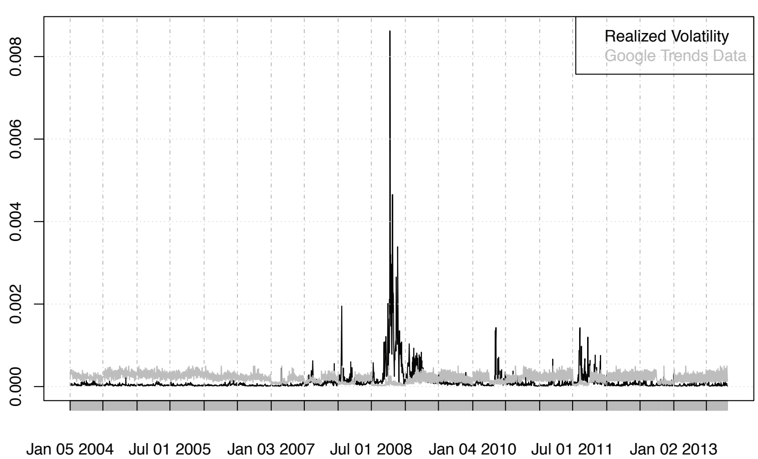
West, Kenneth D., Hali J. Edison, Dongchul Cho. 1993. A utility-based comparison of some models of exchange rate volatility. *Journal of International Economics* **35**(1-2) 23–45.

Figure 1 Data availability



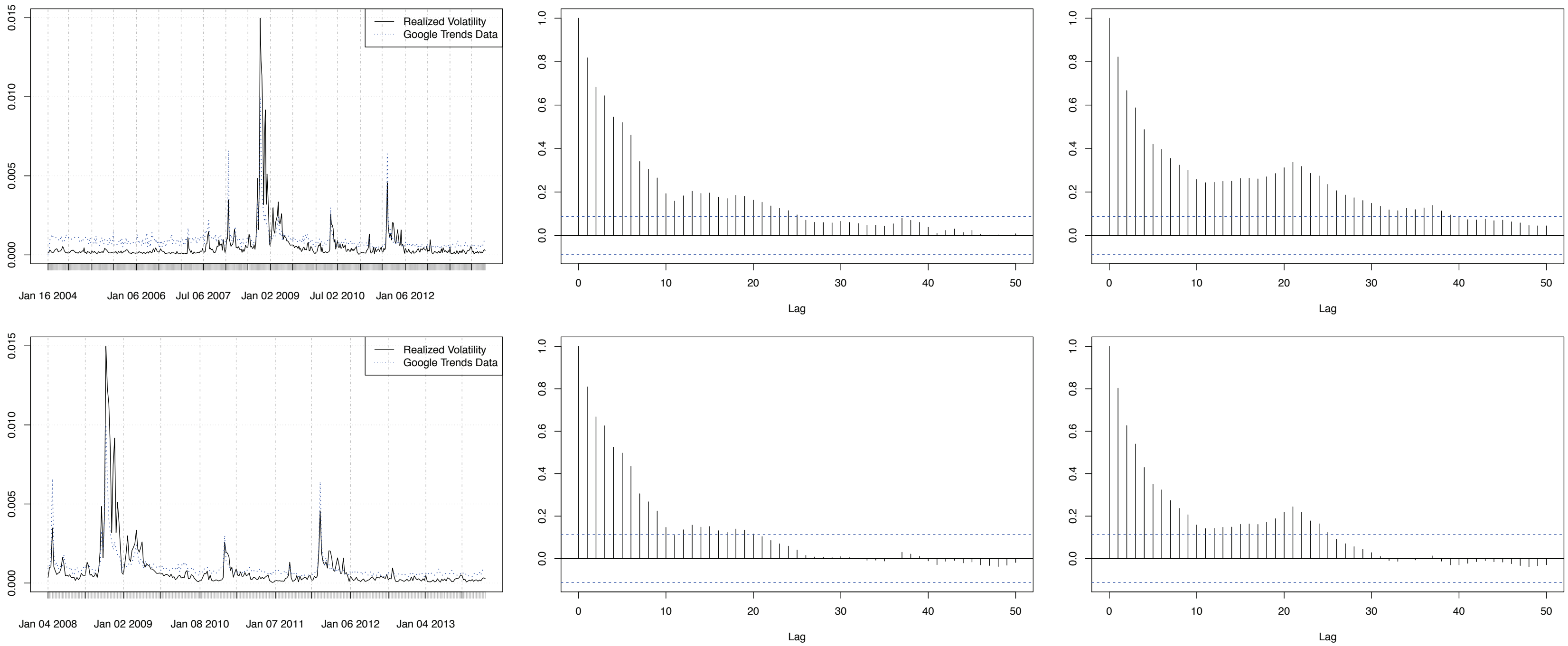
Note. Availability of weekly Realized Volatility and Google Trends data. Forecasts are performed on Monday morning.

Figure 2 Daily RV and Google Trends.



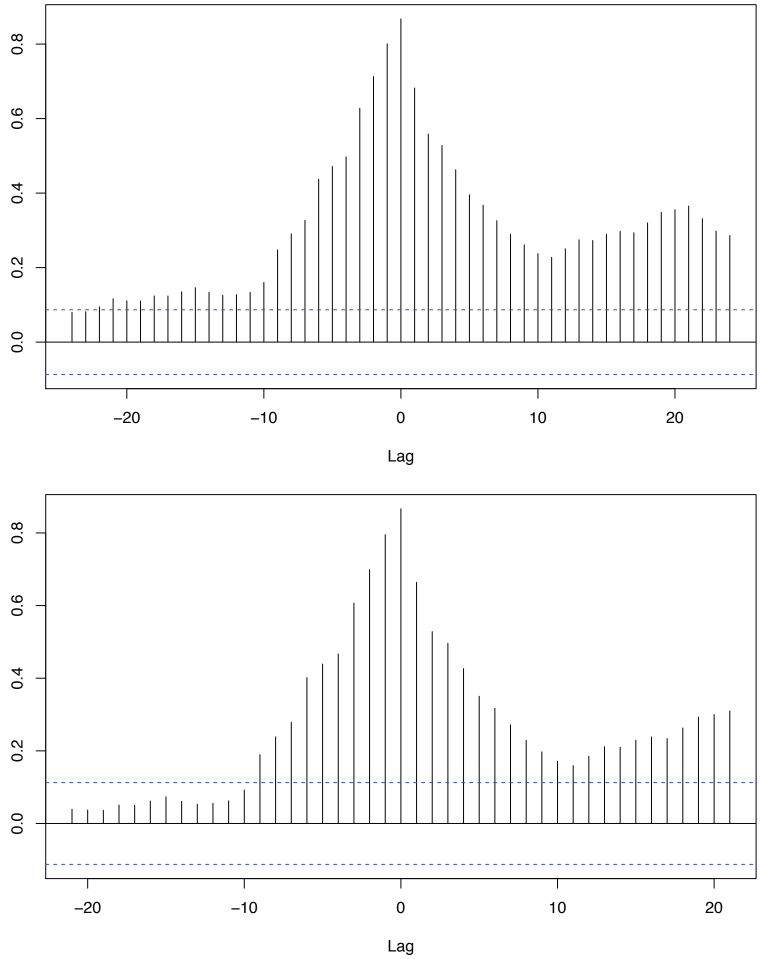
Note. The time-series of daily realized volatility (black) and lagged (lag=2) daily Google Trends data (gray) for the Dow Jones. For comparability the Google Trends data is scaled with a factor of 5e-05.

Figure 3 Autocorrelation functions for RV and Google Trends.



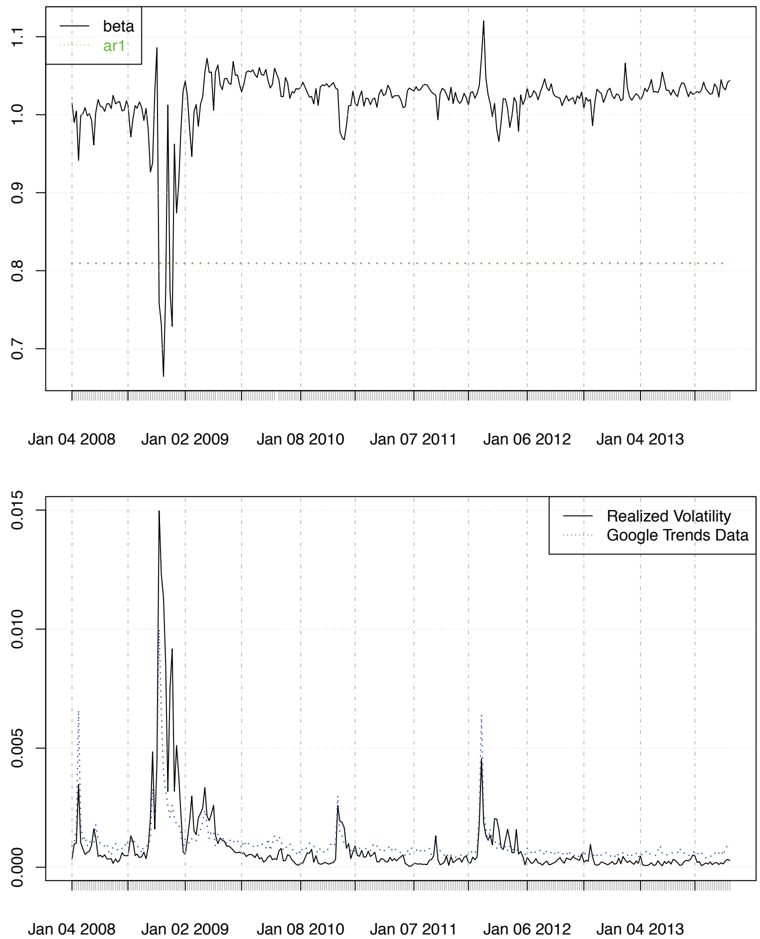
Note. The time-series of weekly realized volatilities and Google data (left) and the corresponding ACFs for the realized volatilities (middle) and the Google data (right). The plots show the full sample (top) and crisis subperiod (bottom). For comparability the Google Trends data is scaled with a factor of 1e-04.

Figure 4 Cross correlation between RV and Google Trends.



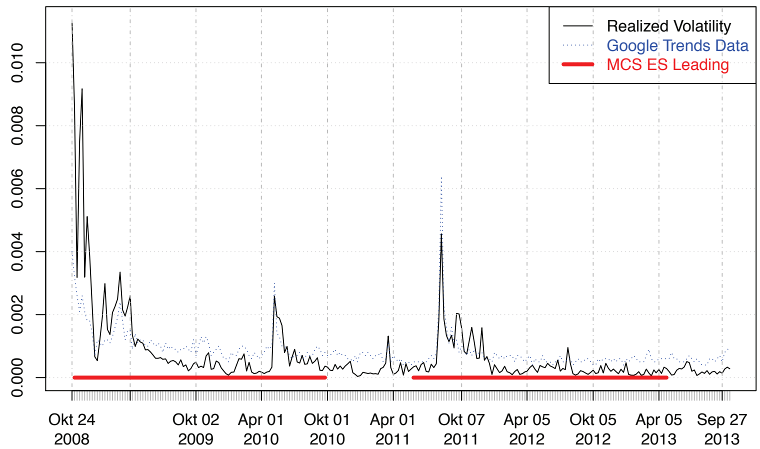
Note. Cross correlation of the RV and Google data for the full sample (top) and the crisis subsample (bottom). Negative lags indicate that the the Google data is lagged whereas positive lags show a lagged RV.

Figure 5 Coefficients for ES and AR model



Note. The evolution of the coefficient β_t for the ES-model (black) and the constant ar_1-parameter (green) for the crisis period (top). Time-series of the realized volatility (black) and Google Trends data (blue) for the crisis period (bottom).

Figure 6 Leading periods of ES model in the Model Confidence Set



Note. The time-series of the realized volatility, Google Trends data for the Dow Jones and the phases where the ES model is leading in the Model Confidence Set.

Table 1 Basis statistics for the realized volatilities and Google Trends.

	min	max	mean	sd	skewness	kurtosis	med	ρ_1	ρ_2	length	stationarity
RV	3.654e-05	1.497e-02	5.978e-04	1.310e-03	6.749e+00	5.563e+01	2.579e-04	8.178e-01	6.838e-01	5.100e+02	1.987e-11
Google	3.000e-04	1.000e-02	9.004e-04	8.134e-04	5.509e+00	4.331e+01	7.000e-04	8.215e-01	6.669e-01	5.100e+02	4.643e-11

	min	max	mean	sd	skewness	kurtosis	med	ρ_1	ρ_2	length	stationarity
RV	3.654e-05	1.497e-02	8.482e-04	1.647e-03	5.269e+00	3.316e+01	3.711e-04	8.091e-01	6.681e-01	3.030e+02	2.908e-07
Google	5.000e-04	1.000e-02	1.133e-03	9.794e-04	4.628e+00	2.933e+01	8.000e-04	8.028e-01	6.268e-01	3.030e+02	1.312e-07

Full sample statistics (upper) and crisis period statistics (lower) for RV and Google Trends.

Table 2 VAR model: Parameter estimates and standard errors

	rv_t	g_t
const.	< 0.01	< 0.01
	(< 0.01)	$(< 0.01^{***})$
rv_{t-1}	0.35	-0.18
	(0.06^{***})	(0.04^{***})
rv_{t-2}	-0.13	< 0.01
	(0.07^{*})	(0.05)
rv_{t-3}	0.46	0.12
	(0.07^{***})	(0.05^{**})
rv_{t-4}	0.10	0.06
	(0.06)	(0.04)
g_{t-1}	0.72	1.04
	(0.09^{***})	(0.06^{***})
g_{t-2}	0.19	-0.10
	(0.12)	(0.08)
g_{t-3}	-0.15	0.12
	(0.12)	(0.08)
g_{t-4}	-0.56	-0.23
	(0.10^{***})	(0.07^{***})
adj. R^2	0.75	0.70

Full sample parameters estimates and standard errors in parenthesis for the VAR model with 4 lags. The volatility equation is given in the first column, the search volume equation is given in the second column. Adjusted R^2 for both regressions are reported in the last row

Table 3 **Parameter estimates and standard deviations.**

model	α_0 $\times 10^3$	ω_1	ω_2	ω_3	g_1	σ $\times 10^3$
ES	0.029	56.034	-	-	-	0.724
	(0.032)	(6.691)	-	-	-	
AR	0.105	0.818	-	-	-	0.753
	(0.031)	(0.051)	-	-	-	
AR G	-0.187	0.638	-	-	1.234	0.512
	(0.057)	(0.089)	-	-	(0.227)	
ARFIMA	0.206	0.270	0.500	-	-	0.751
	(0.127)	(0.157)	(0.185)	-	-	
ARFIMA G	-0.391	0.269	0.321	-	1.226	0.506
	(0.092)	(0.160)	(0.124)	-	(0.220)	
HAR	0.081	0.693	0.167	0.004	-	0.748
	(0.04)	(0.047)	(0.063)	(0.053)	-	
HAR G	-0.482	0.284	0.274	-0.263	1.004	0.481
	(0.033)	(0.034)	(0.040)	(0.035)	(0.037)	

model	α_0 $\times 10^3$	ω_1	ω_2	ω_3	g_1	σ $\times 10^3$
ES	0.05	56.495	-	-	-	0.929
	(0.054)	(8.632)	-	-	-	
AR	0.154	0.810	-	-	-	0.966
	(0.049)	(0.053)	-	-	-	
AR G	-0.218	0.631	-	-	1.269	0.653
	(0.075)	(0.098)	-	-	(0.237)	
ARFIMA	0.565	0.273	0.497	-	-	0.965
	(0.372)	(0.169)	(0.188)	-	-	
ARFIMA G	-0.35	0.273	0.317	-	1.251	0.647
	(0.198)	(0.175)	(0.136)	-	(0.222)	
HAR	0.132	0.692	0.169	-0.016	-	0.963
	(0.072)	(0.061)	(0.081)	(0.070)	-	
HAR G	-0.593	0.274	0.274	-0.243	1.046	0.609
	(0.057)	(0.043)	(0.051)	(0.046)	(0.049)	

The full sample (upper table) and crisis sample (lower table) parameter estimates, the corresponding standard deviation and the standard deviation of the residuals for the considered models. All values are scaled with factor 1000.

Table 4 **Weekly in-sample average losses**

	Full sample		Crisis period	
model	MSE	QLIKE	MSE	QLIKE
ES	1.00	1.00	1.00	1.00
AR	1.08	1.03	1.08	1.08
AR G	0.50	1.66	0.50	1.15
ARFIMA	1.08	0.99	1.08	1.08
ARFIMA G	0.49	1.53	0.47	1.08
HAR	1.06	0.99	1.06	1.05
HAR G	0.44	2.70	0.42	2.67

Full sample and crisis sample weekly average loss values for the MSE and QLIKE loss functions. Results are relative to the ES losses.

Table 5 Weekly out-of-sample average losses

model	MSE	QLIKE	MZ R^2	α_{mz}	β_{mz}
ES	1.00	1.00	0.672	$4.4 \cdot 10^{-5}$	0.84
AR	1.07	1.10	0.648	$6.5 \cdot 10^{-5}$	0.82
AR G	1.59	1.45	0.528	$2.6 \cdot 10^{-5}$	0.70
ARFIMA	1.06	1.06	0.661	$8.9 \cdot 10^{-5}$	0.80
ARFIMA G	1.27	1.53	0.574	$9.5 \cdot 10^{-5}$	0.75
HAR	1.76	1.36	0.557	$2.1 \cdot 10^{-4}$	0.62
HAR G	1.50	1.36	0.597	$1.1 \cdot 10^{-4}$	0.62

Weekly out-of-sample forecast average loss values for the MSE and QLIKE loss functions. Results are relative to the ES losses. R^2, α_{mz} and β_{mz} of Mincer-Zarnowitz regression.

Table 6 Model Confidence Set p-values

MCS p-values out-of-sample forecasts

model	Full sample	Crisis 1	Crisis 2
ES	1.00	1.00	1.00
AR	0.08	0.72	0.04
AR G	0.03	0.07	0.04
ARFIMA	0.08	0.11	0.43
ARFIMA G	0.08	0.11	0.04
HAR	0.03	0.02	0.43
HAR G	0.08	0.02	0.43

Model Confidence Set p-values based on QLIKE loss for full sample and crisis subsample periods. The p-values indicate the significance level at which the H_0 that the respective model is part of the MCS can be rejected.

Table 7 Daily out-of-sample average losses

model	MSE	QLIKE
ES	1.00	1.00
AR	0.74	0.89
AR G	0.73	0.89
ARFIMA	0.59	0.61
ARFIMA G	0.59	0.62
HAR	0.61	0.64
HAR G	0.61	0.64

Daily out-of-sample forecast average loss values for the MSE and QLIKE loss functions. Results are relative to the ES losses.

Table 8 $\widehat{\mathbf{VaR}_t}(\alpha)$, summary statistics and losses for $\alpha = 0.01$

model	% exc.	no. exc.	pval UC .	sq. loss VaR	rev. QLIKE exc.	rev. QLIKE to bench
ES	1.15	3	0.81	1.00	1.00	1.00
AR1	1.15	3	0.81	1.00	1.08	1.22
AR1 G	1.15	3	0.81	1.00	1.08	1.22
ARFIMA	1.54	4	0.42	0.95	1.02	1.01
ARFIMA G	1.15	3	0.81	1.05	1.05	1.33
HAR	1.54	4	0.42	1.05	1.09	1.31
HAR G	0.39	1	0.25	1.32	1.28	1.55

VaR forecast comparison with percentage and number of exceptions, p-val. of Kupiec UC test, squared loss for VaR forecast relative to ex-post return, reverse QLIKE loss for VaR forecast relative to ex-post return, and reverse QLIKE loss relative to benchmark. All losses are relative to the ES loss.

Table 9 **Model Confidence Set p-values $\widehat{\mathbf{VaR}}_t(\alpha)$ for $\alpha = 0.01$**

model	rev. QLIKE exc.	rev. QLIKE to bench
ES	1.00	1.00
AR	0.07	0.02
AR G	0.02	0.02
ARFIMA	0.07	0.37
ARFIMA G	0.26	0.02
HAR	0.02	0.02
HAR G	0.02	0.02

Model Confidence Set p-values for VaR forecasts forreverse QLIKE loss for VaR forecast relative to ex-post return, and reverse QLIKE loss relative to benchmark, see equation 17a and 17b.The p-values indicate the significance level at which the H_0 that the respective model is part of the MCS can be rejected.