Volatility Forecasting: The Role of Internet Search Activity and Implied Volatility

Arabinda Basistha, Alexander Kurov and Marketa Halova Wolfe

First Draft: April 18, 2015
This Draft: January 15, 2016

Abstract

Growing literature has shown usefulness of the Internet search activity data in numerous fields. We reexamine the role of the Internet search activity in a broader context of forecasting volatility in the financial and commodity markets both in-sample and out-of-sample. We find that usefulness of the Internet search activity data disappears in the stock index and foreign exchange markets and substantially declines in the commodity markets once implied volatility is included in the benchmark model. This finding contributes to our understanding of what informational content is captured by the Internet search activity data. It appears to capture similar information as implied volatility.

Keywords: Volatility forecasting, realized volatility, implied volatility, Internet search activity, Google search volume, Google Trends, information

JEL Classification: C32, C52, G12, G14, G17

We thank Fabio Mattos, Nicolas Merener, and participants at 2015 NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management and Washington State University seminar for helpful comments. Errors or omissions are our responsibility.

Arabinda Basistha is an associate professor of economics in the Department of Economics, College of Business and Economics, West Virginia University, P.O. Box 6025, Morgantown, WV 26506, Phone: +1-304-293-7854, E-mail: arbasistha@mail.wvu.edu.

Alexander Kurov is an associate professor of finance in the Department of Finance, College of Business and Economics, West Virginia University, P.O. Box 6025, Morgantown, WV 26506, Phone: +1-304-293-7892, E-mail: alkurov@mail.wvu.edu.

Marketa Halova Wolfe is an assistant professor of economics in the Department of Economics, Skidmore College, Saratoga Springs, NY, Phone: +1-518-580-5090, E-mail: mwolfe@skidmore.edu.

Volatility Forecasting: The Role of Internet Search Activity and Implied Volatility

First Draft: April 18, 2015
This Draft: January 15, 2016

Abstract

Growing literature has shown usefulness of the Internet search activity data in numerous fields. We reexamine the role of the Internet search activity in a broader context of forecasting volatility in the financial and commodity markets both in-sample and out-of-sample. We find that usefulness of the Internet search activity data disappears in the stock index and foreign exchange markets and substantially declines in the commodity markets once implied volatility is included in the benchmark model. This finding contributes to our understanding of what informational content is captured by the Internet search activity data. It appears to capture similar information as implied volatility.

Keywords: Volatility forecasting, realized volatility, implied volatility, Internet search activity, Google search volume, Google Trends, information

JEL Classification: C32, C52**,** G12, G14, G17

1. Introduction

Growing literature has shown usefulness of the Internet search activity data in numerous fields including Ginsberg et al (2009) for detecting influenza epidemics, Choi and Varian (2009) for predicting automobile sales, unemployment claims and consumer confidence, D'Amuri and Marcucci (2009) for forecasting unemployment rate, and Wu and Brynjolfsson (2013) for predicting housing market trends. Recent studies have shown that online search activity is also associated with volatility and returns in the financial and commodity markets such as Da, Engelberg and Gao (2011) and Vlastakis and Markellos (2012) for individual stocks, Dimpfl and Jank (2015) for the Dow Jones Industrial Average (DJIA) index, Da, Engelberg and Gao (2015) for stock indices, exchange traded funds and Treasury bonds, Goddard, Kita and Wang (2012) and Smith (2012) for exchange rates, Vozlyublennaia (2014) for stock and bond indices, gold and crude oil, and Guo and Ji (2013) for crude oil.

We reexamine the role of the Internet search activity data in a broader context of forecasting volatility in the financial and commodity markets both in-sample and out-of-sample. We use Google search volume data available at weekly frequency since January 2004.[1] We include data for stock indices (S&P 500 and DJIA), foreign exchange (Euro and Canadian dollar), and commodities (gold, silver, copper, crude oil, natural gas and corn). We build on seminal work by Andersen, Bollerslev, Diebold and Labys (2001) who propose measuring volatility as realized volatility and subsequent research by Andersen, Bollerslev, Diebold and Labys (2003) and Andersen, Bollerslev and Meddahi (2004) who propose forecasting volatility

[1] Google is the most popular Internet search engine with the U.S. market share of 63.8% as of August 2015 (http://www.comscore.com/Insights/Market-Rankings/comScore-Releases-August-2015-U.S.-Desktop-Search-Engine-Rankings).

by reduced-form models of realized volatility as they outperform models such as the generalized autoregressive conditional heteroskedasticity (GARCH) model.

In the in-sample analysis, we employ a vector autoregressive (VAR) model, Granger causality tests, and forecast error variance decomposition. In-sample results are often prone to pitfalls involving overfitting and spurious correlations. Therefore, we follow with out-of-sample evaluations that have been quite effective in reducing these problems. Here, the key out-of-sample evaluation concept is encompassing. It argues that if model 1 contains all relevant information for forecasting a target variable over model 2, forecast errors of model 1 should be identical to forecasts from model 2. Otherwise, model 2 provides additional information in the forecasts and is not encompassed by model 1. This is especially useful in our context because we want to examine the marginal contribution of different predictors. We begin with a simple four-lag autoregressive model of realized volatility (AR4) measured as the realized standard deviation (computed using 5-minute continuously compounded returns). Against this benchmark we evaluate the marginal contribution of four predictors proposed in the previous literature: trading volume, returns, Google search volume, and implied volatility. We conclude that the AR4 model with implied volatility substantially outperforms the other models, which agrees with extensive literature (for example, Busch, Christensen and Nielsen, 2011) employing implied volatility to forecast realized volatility. It is against this benchmark that we evaluate the Google search volume. We find that the usefulness of Google search volume for forecasting realized volatility disappears in the financial markets and substantially declines in the commodity markets once implied volatility is included in the model.

This result contributes to our understanding of what informational content is captured by the Internet search activity data. Previous papers, for example, Da, Engelberg and Gao (2011), Goddard, Kita and Wang (2012), Vlastakis and Markellos (2012), and Vozlyublennaia (2014) discuss that the Internet search activity captures investor attention or information demand. Our results suggest that it captures some of the same information as implied volatility: market's expectation of future volatility over the life of the options. Neely (2005) analyzes news events around the largest changes in implied volatility of options on Eurodollar futures from 1985 to 2001. The stock market crash of 1987, a decrease in the U.S. trade deficit, President George H. W. Bush asking Congress for authority to oust Iraq from Kuwait, the Russian debt crisis, and another sharp decline in the U.S. trade deficit rank among the top five events. Developments in the financial markets and the U.S. monetary policy feature among other influential events. In this sense, the previous studies about usefulness of the Internet search activity for forecasting realized volatility are not misguided; the Internet search activity does likely reflect interest in acquiring more information about an assortment of events. However, perhaps because most of these internet searches come from the general public and do not translate into *trading* in the financial and commodity markets, the effect of implied volatility subsumes the noisier effect of the Internet search activity.

Our finding is novel because previous papers do not include implied volatility in their analysis of usefulness of Internet search activity for realized volatility forecasting. There are two exceptions. Dimpfl and Jank (2015) use keywords such as "dow" to study the DJIA volatility from 2006 to 2011. They show in a VAR framework that the effect Google search volume decreases but does not disappear when implied volatility is added in their in-sample analysis of realized volatility.[2] Dzielinski (2012) extracts search volume for the keyword "economy" and in a single-regression ordinary least squares (OLS) regression finds that this measure remains

[2] Dimpfl and Jank (2015) present both an in-sample analysis and out-of-sample forecasting, but implied volatility features only in the in-sample analysis.

significant in-sample even after controlling for existing measures of uncertainty such as implied volatility in the S&P 500 from 2005 to 2009. Our results extend findings from these studies because we use the benchmark model with implied volatility not only for in-sample analysis but also for out-of-sample forecasting.[3]

The remainder of this paper is structured as follows. Section 2 describes the data, Section 3 presents the empirical results, and Section 4 concludes with a brief discussion.

2. Data

This section describes the financial and commodity markets data followed by a description of the Internet search activity data. We then present correlations between the two data sets.

2.1 Financial and Commodity Market Data

To investigate the effect of Internet search activity on volatility in the financial and commodity markets, we use a variety of assets including stock indices, foreign exchange and commodities. For each asset class, we include multiple markets as listed in Table 1. From stock indices, we include the S&P 500 Index and DJIA Index that are among the world's most important stock indices. For foreign exchange, we include the Euro and Canadian dollar with exchange rates denominated in U.S. dollars per unit of the foreign currency. The Euro is the second largest currency following the U.S. dollar; the Canadian dollar is the largest commodity currency. From commodities, we include all commodities for which uninterrupted Internet search activity data is available: gold, silver and copper among metal commodities, crude oil and natural gas among

[3] Three other papers (Da, Engelberg and Gao, 2015, Goddard, Kita and Wang, 2015, and Vlastakis and Markellos, 2012) include implied volatility as a *measure* of volatility and analyze the effect of Internet search activity on implied volatility forecasting. Our focus is different: we forecast realized volatility and examine the effect of Internet search activity after *controlling* for implied volatility.

energy commodities, and corn from agricultural commodities. Since the results for all commodities are similar, we report results only for gold and crude oil.

In our analysis we need a measure of implied volatility for each of the six markets. For the stock indices, we use the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) and DJIA Volatility Index (VXD) that measure implied volatility of the S&P 500 Index and DJIA Index options, respectively. Correspondingly, we use spot data for the S&P 500 and DJIA indices to obtain prices and trading volume. For commodities, we have available implied volatility of options on the nearby gold and crude oil futures contracts. Correspondingly, we use futures data to obtain prices and trading volume of gold and crude oil. For foreign exchange, we have available implied volatility of spot options. Since we do not have foreign exchange spot price and trading volume data, we use futures prices and trading volume for the Euro and Canadian dollar.[4]

[Insert Table 1 about here]

Following Andersen, Bollerslev, Diebold and Labys (2001) who propose measuring volatility as realized volatility, we compute the weekly realized volatility as follows:

$$RV_t = \sqrt{\sum_{i=1}^{n} r_{t,i}^2}, \tag{1}$$

where RV_t is the realized standard deviation during week t and $r_{t,i}^2$ is the squared continuously compounded return in intraday interval i during week t.[5] Following the existing literature (for

[4] The futures and spot markets data is obtained from Genesis Financial Technologies. The implied volatility data is obtained from Bloomberg except for the VIX and VXD indices that are publicly available on the internet.
[5] For the futures data, the returns are calculated using prices from a continuous series of the most liquid futures contract. The nearby contract becomes relatively illiquid in the last few days of its trading. Therefore, we switch to the next month contract when its daily contract volume exceeds the nearby contract volume.

example, Bollerslev, Tauchen and Zhou, 2009), we use 5-minute intraday intervals in the calculation.[6]

2.2 Internet Search Activity

To measure trader attention to the financial and commodity markets, we obtain Internet search activity data from Google Trends (http://www.google.com/trends), a Google service that provides data showing how frequently search terms have been used in the Google Search engine. Market participants looking for information about financial and commodity markets use many possible search terms. Instead of displaying the *number* of searches for each search term, Google Trends calculates a search volume *index* scaled by the maximum value over the time period selected for each search term. The search volume index ranges from zero to 100, with a value of 100 representing the peak of search activity for the given search term during the sample period. This normalization makes it difficult to aggregate search volume indices for multiple search terms because the number of searches differs across search terms. Fortunately, Google Trends aggregates search activity data for related searches by topic categories and regions.

We extract search activity within the Finance category in the U.S. region using the most appropriate subcategory for each market: the Investing subcategory for S&P 500 and DJIA, the Currencies & Foreign Exchange subcategory for the Euro and Canadian dollar, and the Commodities & Futures Trading subcategory for gold and crude oil. Specifically, we download search volume indices for the following search terms: ‘s&p,’ ‘dow,’ ‘euro,’ ‘canadian,’ ‘gold,’

---

[6] As a robustness check, we use range-based volatility estimators proposed by Garman and Klass (1980) and Rogers and Satchell (1991). The results are similar in all volatility estimators. Therefore, we report only the realized volatility results.

and ‘oil’.[7] These search volume indices represent search activity data aggregated across many search queries that contain these commodity names. For example, the top five search queries containing the term ‘s&p’ were *s&p*, *s&p 500*, *s&p index*, *s&p 500 index* and *s&p futures*, and the top five search queries containing the term ‘oil’ were *oil prices*, *oil price*, *crude oil*, *price of oil* and *crude oil prices*. Section 3.3 discusses robustness checks with these individual search queries.

Google Trends data is available since January 2004 at weekly frequency. We examine the sample period from January 4, 2004 to August 28, 2015. When search activity for a given search term is too low, Google Trends reports missing (zero) values of the search volume index. We have uninterrupted non-missing values for our search terms during our entire sample period.

Figure 1 shows the Google search volume indices and realized volatility over our sample period. We present only one market from each asset class (DJIA for stock indices, Euro for foreign exchange, and crude oil for commodities) to save space. Increases in search activity coincide with periods of high realized volatility. The other markets (S&P 500, Canadian dollar and gold) exhibit similar patterns and the corresponding figures are available upon request. Periods of high realized volatility appear to coincide with periods exhibiting high Google search volume.

[Insert Figure 1 about here]

2.3 Correlations

Table 2 shows correlations of log-differences of Google search volume with realized volatility, implied volatility, trading volume and returns.[8] Changes in the search activity are positively and

---

[7] Using these subcategories ensures that our data is not polluted by searches containing our search terms but not related to these financial and commodity markets. For example, the crude oil data is not polluted with searches for ‘olive oil’ or ‘baby oil.’

[8] We use log-differences to avoid potential spurious correlations.

significantly correlated with contemporaneous changes in realized volatility and trading volume in all six markets. For example, the correlation between changes in search activity and changes in realized volatility ranges from 0.30 for S&P 500 to 0.55 for crude oil. This result is consistent with trader attention reflected in the Google search queries that translates into trading activity. In several markets, there is a negative and statistically significant correlation of changes in the Google search volume with contemporaneous returns.

[Insert Table 2 about here]

3. Results

We employ two complementary approaches to examine the role of the Internet search activity data in a broader context of forecasting volatility in the financial and commodity markets: in-sample analysis and out-of-sample forecasting. We follow with robustness checks.

3.1. VAR Estimation Results

We proceed in two steps. First, we estimate a vector autoregression (VAR) for each market without implied volatility. Second, we add implied volatility to the VAR to examine whether implied volatility affects usefulness of the Google Trends search volume index for predicting realized volatility.

We begin by estimating the following VAR:

$$\boldsymbol{x_t} = \boldsymbol{\alpha} + \sum_{j=1}^{4} \boldsymbol{\beta_j}\, \boldsymbol{x_{t-j}} + \boldsymbol{\varepsilon_t}, \tag{2}$$

where $\boldsymbol{\alpha}$ is a vector of constant terms, $\boldsymbol{\beta_j}$ is the vector of coefficients for lag j, and $\boldsymbol{x_t}$ is a vector of four variables: weekly Google Trends search volume index, realized volatility measured by the realized standard deviation, trading volume and return.[9] $\boldsymbol{\varepsilon_t}$ is a vector of random disturbances. Following Dimpfl and Jank (2015), we take the natural logs of the realized standard deviation, search volume index and trading volume. This transformation reduces skewness and excess kurtosis of these variables. We also test for stationarity using the Phillips and Perron (1988) test. The null hypothesis of a unit root is strongly rejected for all variables in all markets. We include four lags of variables in the VAR.

For all markets except DJIA, the coefficient estimate of the first lag of the Google search volume in the realized volatility equation is positive and statistically significant at the 5% level.[10] To examine whether Google search volume has predictive power for realized volatility, trading volume and returns, we use the VAR estimation results to perform Granger causality tests. Table 3 Panel A shows the results. In all six markets, there is strong evidence that Google search volume Granger-causes realized volatility after controlling for the other variables, i.e., lags of realized volatility, trading volume and return. Google search volume is also a useful predictor of trading volume in four of the six markets and returns in three of the six markets. In most markets, the relation between realized volatility and trading volume is bidirectional as described in previous studies, for example, Darrat, Zhong and Cheng (2007).

[Insert Table 3 about here]

An alternative way of characterizing the relative predictive content of variables in the VAR is the decomposition of the forecast error variance. The decomposition represents the relative contribution of innovations in each variable to the other variables. Table 4 Panel A shows the variance decomposition results for the log of realized volatility. The variance

[9] We also considered including futures open interest in the model. However, the open interest of the nearby contract is driven to a large extent by periodic rollovers from the nearby to the next-to-mature contract.
[10] The VAR coefficient estimates are not tabulated for brevity but are available upon request.

decomposition results depend on the ordering of variables. Placing a variable earlier in the decomposition tends to increase its contribution to the forecast error variance. Therefore, we use two alternative orderings. When the Google search volume is placed last in the ordering, its contribution to the forecast error variance of log realized volatility ranges from about 3% for S&P 500 to almost 20% for crude oil and averages about 8% across the six markets. This can be viewed as the lower bound of the contribution of Google search volume in predicting realized volatility. When we place the Google search volume first in the ordering, its contribution in the forecast error variance of realized volatility increases and averages about 44% across the six markets. Overall, the variance decompositions suggest that Google search volume explains a significant portion of the forecast error variance of realized volatility.

[Insert Table 4 about here]

Figure 2 shows impulse responses that represent the effect of a one standard-deviation shock in a given variable on the other variables in the model. The time periods on the horizontal axis correspond to weeks. We present data only for one market from each asset class (DJIA for stocks, Euro for foreign exchange and oil for commodities) to save space. Impulse responses for the other three markets are generally similar and available upon request. The first column shows the effect of the Google search volume shock on the other variables. Unexpected increases in the Google search volume predict higher realized volatility and higher trading volume. The first line shows how Google search activity reacts to shocks in the other variables. Google search volume tends to decrease after price increases. Google search volume reacts positively to realized volatility and trading volume. As previous papers surmise, this could reflect investors searching for information on the Internet as a reaction to news that caused unexpected volatility and trading volume in the financial and commodity markets.

[Insert Figure 2 about here]

Overall, the above results suggest that the Google search volume is a useful predictor of the realized volatility. We now add implied volatility to the VAR described in equation (2) to examine whether implied volatility affects usefulness of the Google search volume for predicting realized volatility. Table 3 Panel B presents results of these Granger causality tests. The highly significant coefficients indicate that implied volatility predicts realized volatility in all six markets. This agrees with previous literature such as Busch, Christensen and Nielsen (2011). The impact of including implied volatility in the VAR is striking in another respect: the significance of the Google search volume disappears in the stock index and foreign exchange markets and declines in the commodity markets.

The forecast error variance decomposition for the log of realized volatility in Table 4 Panel B leads to the similar conclusion. Again, the variance decomposition results depend on the ordering of variables. When we place the Google search volume last in the ordering, its contribution to the forecast error variance of log realized volatility ranges from about 1% for S&P 500 to less than 16% for crude oil and averages less than 5% across the six markets. When we place the Google search volume first in the ordering in predicting realized volatility, its contribution in the forecast error variance averages about 32% across the six markets compared to 44% in Panel A. This shows that the relative contribution of innovations in the Google search volume to the other variables decreases once implied volatility is included in the VAR.

3.2. Out-of-Sample Forecasting

This section takes an out-of-sample approach to evaluating the role of Google search volume in forecasting volatility. The out-of-sample approach has been quite effective in reducing the problem of in-sample overfitting with spurious regressors. However, one potential issue is the sensitivity of results to the estimation window size. To address this shortcoming, we use the Rossi and Inoue (2012) methods robust to the window size to avoid concerns about data-snooping over window sizes. Their encompassing (ENC) tests build on Clark and McCracken (2001) study that compares forecast errors in nested models.

For each market, our benchmark model forecasts the log of realized volatility based on lags of realized volatility. We use four lags to be consistent with the VAR. This is our restricted model, model 1. Let the forecast errors from this model be denoted as u_{1t}. We then add – one at a time – four lags of returns, log of trading volumes, log of Google search volume, and log of implied volatility to the benchmark model to form our unrestricted model, model 2. Let the forecast errors from these unrestricted models be denoted as u_{2t}. Let *R* be the number of observations used to estimate the parameter estimates to form the first one-step forecast. After that, the models are recursively estimated adding one observation at a time. If *T* denotes the total number of observations, there will be *T-R* forecasts from restricted and unrestricted models. The ENC test statistic is denoted as:

$$ENC = \frac{\sum_{t=R+1}^{T} u_{1t}(u_{1t} - u_{2t})}{\sum_{t=R+1}^{T} u_{2t}^2} (T - R). \tag{3}$$

To compute the ENC tests recursively, we start at the lower end of the estimation window with R^L observations and after adding one observation at a time we go up to the upper end, R^U. We follow the Rossi and Inoue (2012) recommendation to use 15% trimming on each side of the sample for choosing R^L and R^U. For example, the R^L corresponds to observations 91 and 81 in the S&P 500 and DJIA markets, respectively.

Rossi and Inoue (2012) recommend using two versions of the ENC test. The tests are denoted as

$$Sup\text{-}ENC = Sup_{R \in (R^L, \ldots, R^U)} \{ENC(R)\} \tag{4}$$

and

$$Ave\text{-}ENC = \frac{1}{R^U - R^L + 1} \sum_{R=R^L}^{R^U} ENC(R). \quad (5)$$

Results of these average (Ave-ENC) and supremum (Sup-ENC) encompassing tests are reported in the upper panels of Tables 5, 6 and 7 for stock, foreign exchange and commodity markets, respectively. High values indicate that u_{2t} is relatively small compared to u_{1t}. Bold text indicates statistical significance at 5% level based on critical values of Rossi and Inoue (2012). In all six markets, among the four variables (return, trading volume, Google search volume and implied volatility) it is the model with implied volatility that performs the best relative to the AR4 benchmark.

[Insert Tables 5, 6 and 7 about here]

To quantify *how much* the forecasting improves over the unrestricted model, we compute the ratio of the restricted model mean square prediction errors (MSPE) to the unrestricted model MSPE. The MSPEs are averaged across all windows in computing the ratio. For example, the ratio of 0.79 for the AR4+IV model in the S&P 500 market indicates that adding implied volatility to the AR4 model results in 21% decrease in the MSPE. Among the four unrestricted models the MSPE ratio is the lowest in the AR4+IV model in all six markets.

We, therefore, use the AR4+IV as the benchmark for further evaluation of potential predictors. Here, it is important to note that our purpose is not building a state-of-the-art volatility forecasting model[11]; the aim is finding a simple benchmark against which we can evaluate potential predictors. We add returns, trading volume and Google search volume one at a time to see if any of these variables improve on the AR4+IV forecast. Bottom panels of Tables 5,

[11] For example, recent research such as Andersen, Bollerslev and Diebold (2007) and Lee and Mykland (2008) suggests modelling jumps when forecasting volatility.

6 and 7 present the results. Google search volume does not improve the AR4+IV forecast in the stock index or foreign exchange markets. It does improve the AR4+IV forecast in the commodity markets, but the improvement is not substantial as indicated by the MSPE ratios of 0.99 and 0.94 for the gold and crude oil markets, respectively.

Overall, the out-of-sample results show that although the Google search volume appears to improve the AR4 forecast of realized volatility, its contribution disappears in the financial markets and declines in the commodity markets once implied volatility is included in the model.

3.3 Robustness Checks

In this subsection, we test whether our results are robust to the choice of Google Trends search terms. As explained in Section 2.1, the search volume indices for our search terms (‘sp500’, ‘dow’, ‘euro’, ‘canadian dollar’, ‘gold’ and ‘oil’) represent search activity data aggregated across many search queries containing these terms. For example, the top five search queries containing the term ‘oil’ were *oil prices*, *oil price*, *crude oil*, *price of oil* and *crude oil prices*, etc. To eliminate concerns about possible data mining over the search terms, we test each of the top five search queries separately. The results do not differ. This agrees with, for example, Dimpf and Jank (2015) who also conclude that their results are robust to the choice of keywords such as *dow*, *dow jones* and *djia*.

Finally, since our main analysis uses search activity data in the U.S. region, we conduct a robustness check using the same search terms with “World” as the region. The results are generally similar but the usefulness of Google search volume is somewhat weaker for stock index and commodity markets even in the model without implied volatility, perhaps reflecting the fact that Internet searches in some countries may not be closely linked to trading. For example, India ranks third in searches for ‘gold’ that Google classifies into “Commodities &

Futures Trading" category. This interest in gold is, however, most likely related to utilization of gold for jewelry and storage of value in general rather than gold futures trading.

4. Conclusion

We reexamine the role of Internet search activity in a broader context of forecasting volatility in the financial and commodity markets. We first find a benchmark to evaluate the contribution of Google search volume against. Here, our purpose is not building a state-of-the-art volatility forecasting model; the aim is finding a simple benchmark against which we can evaluate potential predictors. We find that usefulness of the Google search volume disappears in the financial markets and substantially declines in the commodity markets once implied volatility is included in the benchmark model. This suggests that the Google search volume data captures some of the same informational content as implied volatility.

It is interesting that Google search volume maintains some, although small, predictive power in the commodity markets. Forecasting commodity price volatility is critical not only for traditional hedgers who use commodities in production but also for investors such as commodity index funds who increasingly include commodity futures in their portfolios as documented in the literature on financialization of commodity markets (for example, Büyükşahin, and Robe, 2014 and Tang and Xiong, 2012). In these markets it may be worth exploring how to include Google search volume in the volatility forecasting models.

References

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

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, 42–55.

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

Andersen, T. G., Bollerslev, T., & Meddahi, N., 2004. Analytical evaluation of volatility forecasts. *International Economic Review* 45, 1079–1110.

Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, 22, 4463–4492.

Busch, T., Christensen, B. J. & Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 160, 48–57.

Büyükşahin, B., & Robe, M. A. (2014). Speculators, commodities and cross-market linkages. *Journal of International Money and Finance*, 42, 38–70.

Choi, H. & Varian, H. (2012). Predicting the present with Google Trends. *The Economic Record*, (88), 2–9.

Clark, T. E. & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105, 85–110.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, 66, 1461–1499.

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS: Investor sentiment and asset prices. *The Review of Financial Studies*, 28, 1–32.

D'Amuri, F. & Marcucci, J. (2013). "Google It!" Forecasting the US unemployment rate with a Google job search index. *Fondazione Eni Enrico Mattei Working Paper No. 31.2010.*

Darrat, A. F., Zhong, M. & Cheng, L. T. W. (2007). Intraday volume and volatility relations with and without public news. *Journal of Banking & Finance*, 31, 2711–2729.

Dimpfl, T., & Jank, S. (2015). Can Internet search queries help to predict stock market volatility. *European Financial Management*, forthcoming.

Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters*, 9, 167–175.

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

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, 1012–1015.

Goddard, J., Kita, A., & Wang, Q. (2015). Investor attention and FX market volatility. *Journal of International Financial Markets, Institutions and Money*, 38, 79–96.

Guo, J. & Ji, Q. (2013). How does market concern derived from the Internet affect oil prices? *Applied Energy*, 112, 1536–1543.

Lee, S. S. & Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Statistics*, 21, 2535–2563.

Neely, C. (2005). Using implied volatility to measure uncertainty about interest rates. *Federal Reserve Bank of St. Louis Review*, May/June, 87, 407–425.

Phillips, P. C. B. & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75, 335–346.

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

Rossi, B. & Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, 30, 432–453.

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

Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, 68, 54–74.

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

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

Wu, L. & Brynjolfsson, E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. *Economic Analysis of the Digital Economy*. Eds. Avi Goldfarb, Shane Greenstein and Catherine Tucker. Chicago: University of Chicago Press. 89–118.

Table 1
Summary Information for Financial and Commodity Markets Data

	Spot/ Futures[a]	Exchange[b]	Contract Symbol	Implied Volatility[c]
S&P 500	Spot	N/A	N/A	CBOE Volatility Index (VIX)
DJIA	Spot	N/A	N/A	CBOE DJIA Volatility Index (VXD)
Euro	Futures	CME	6E	Implied volatility of Euro spot options
Canadian Dollar	Futures	CME	6C	Implied volatility of Canadian dollar spot options
Gold	Futures	COMEX	GC	Implied volatility of options on gold nearby futures contract
Crude Oil	Futures	NYMEX	CL	Implied volatility of options on crude oil nearby futures contract

[a] All futures contracts are traded on the Chicago Mercantile Exchange Globex electronic trading platform.

[b] CME, COMEX and NYMEX stand for the Chicago Mercantile Exchange, Commodity Exchange and New York Mercantile Exchange, respectively.

[c] CBOE stands for the Chicago Board Options Exchange.

Table 2
Correlations of Google Search Activity, Realized Volatility, Implied Volatility, Trading Volume and Returns

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
Correlation with:						
Realized volatility	**0.30 (0.00)**	**0.53 (0.00)**	**0.49 (0.00)**	**0.32 (0.00)**	**0.55 (0.00)**	**0.55 (0.00)**
Implied volatility	0.05 (0.26)	**0.28 (0.00)**	**0.22 (0.00)**	**0.19 (0.00)**	**0.34 (0.00)**	**0.26 (0.00)**
Trading volume	**0.33 (0.00)**	**0.40 (0.00)**	**0.41 (0.00)**	**0.28 (0.00)**	**0.40 (0.00)**	**0.39 (0.00)**
Return	-0.02 (0.64)	**-0.18 (0.00)**	**-0.17 (0.00)**	-0.08 (0.05)	-0.01 (0.84)	**-0.14 (0.00)**
N	607	536	607	607	513	513

Log-differences are used for the Google Trends search volume index, realized standard deviation, trading volume and implied volatility. *p*-values are shown in parentheses. Bold text indicates statistical significance at 5% level. The beginning of the sample period is January 4, 2004 for S&P 500, Euro and Canadian Dollar, May 15, 2005 for DJIA, and October 23, 2005 for Gold and Crude Oil. The end of the sample period is August 28, 2015 for all markets. N indicates the number of observations measured in weeks.

Table 3
Granger Causality Tests

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
			Panel A: Without IV in VAR			
GT → RV	**10.1 (0.04)**	**11.4 (0.02)**	**26.9 (0.00)**	**13.0 (0.01)**	**24.4 (0.00)**	**41.4 (0.00)**
RV → GT	6.7 (0.15)	**9.9 (0.04)**	6.5 (0.17)	4.6 (0.33)	**11.2 (0.02)**	**13.0 (0.01)**
GT → Trading volume	**22.7 (0.00)**	3.3 (0.51)	**15.2 (0.00)**	2.9 (0.57)	**16.1 (0.00)**	**16.9 (0.00)**
Trading volume → GT	**13.0 (0.01)**	**19.1 (0.00)**	1.2 (0.87)	9.5 (0.05)	9.3 (0.05)	5.2 (0.27)
GT → Return	**17.2 (0.00)**	**14.9 (0.00)**	4.2 (0.38)	7.9 (0.10)	3.0 (0.56)	**10.3 (0.04)**
Return → GT	**21.2 (0.00)**	**27.4 (0.00)**	9.0 (0.06)	6.0 (0.20)	**22.6 (0.00)**	2.3 (0.68)
RV → Trading volume	6.7 (0.15)	7.6 (0.11)	**20.4 (0.00)**	**13.6 (0.01)**	**33.3 (0.00)**	**16.0 (0.00)**
Trading volume → RV	**15.3 (0.00)**	8.0 (0.09)	**12.3 (0.02)**	**14.2 (0.01)**	**22.2 (0.00)**	**35.3 (0.00)**
			Panel B: With IV in VAR			
IV → RV	**120 (0.00)**	**100 (0.00)**	**221 (0.00)**	**164 (0.00)**	**91.1 (0.00)**	**44.1 (0.00)**
RV → IV	3.5 (0.48)	3.3 (0.51)	**12.8 (0.01)**	**22.2 (0.00)**	**16.2 (0.00)**	**9.9 (0.04)**
IV → Trading volume	**78.2 (0.00)**	**62.8 (0.00)**	**81.2 (0.00)**	**65.4 (0.00)**	**43.2 (0.00)**	**28.7 (0.00)**
Trading volume → IV	**19.3 (0.00)**	**10.8 (0.03)**	6.1 (0.19)	**11.7 (0.02)**	2.6 (0.62)	6.6 (0.16)
GT → RV	7.1 (0.13)	2.7 (0.62)	6.0 (0.20)	4.3 (0.36)	**10.0 (0.04)**	**30.8 (0.00)**
RV → GT	7.6 (0.11)	9.0 (0.06)	**32.7 (0.00)**	**17.5 (0.00)**	**21.6 (0.00)**	**24.7 (0.00)**
GT → IV	**11.8 (0.02)**	8.0 (0.09)	**10.1 (0.04)**	4.1 (0.39)	**11.3 (0.02)**	**22.2 (0.00)**
IV → GT	**36.3 (0.00)**	**47.5 (0.00)**	**51.3 (0.00)**	**26.8 (0.00)**	**24.3 (0.00)**	**32.5 (0.00)**
GT → Trading volume	**19.5 (0.00)**	9.4 (0.05)	7.4 (0.12)	4.1 (0.39)	**12.0 (0.00)**	**10.7 (0.03)**
Trading volume → GT	**12.8 (0.01)**	**16.0 (0.00)**	1.6 (0.81)	**16.8 (0.00)**	**10.5 (0.03)**	2.7 (0.60)
GT → Return	**17.1 (0.00)**	**13.6 (0.01)**	2.4 (0.66)	7.1 (0.13)	2.8 (0.60)	**10.9 (0.03)**
Return → GT	0.8 (0.94)	1.0 (0.91)	5.7 (0.23)	7.1 (0.13)	**18.3 (0.00)**	1.9 (0.75)
RV → Trading volume	**34.6 (0.00)**	**32.6 (0.00)**	**59.4 (0.00)**	**43.3 (0.00)**	**51.2 (0.00)**	**36.0 (0.00)**
Trading volume → RV	2.0 (0.74)	4.2 (0.38)	**15.6 (0.00)**	6.4 (0.17)	**19.1 (0.00)**	**31.6 (0.00)**
N	603	532	603	603	510	510

The table shows Wald test statistics of VAR Granger causality tests. GT, IV and RV stand for logs of the Google Trends search volume index, implied volatility and realized standard deviation, respectively. *p*-values are shown in parentheses. Bold text indicates statistical significance at 5% level. The beginning of the sample period is February 6, 2004 for S&P 500, Euro and Canadian Dollar, June 17, 2005 for DJIA, and November 25, 2005 for Gold and Crude Oil. The end of the sample period is August 28, 2015 for all markets. N indicates the number of observations measured in weeks.

Table 4
Variance Decomposition from the VAR

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
Panel A: Without IV in VAR						
Cholesky ordering: RV, Trading Volume, Return, GT						
GT	3.4	5.4	10.0	3.9	7.9	19.8
RV	84.0	86.8	87.1	94.5	81.2	75.5
Trading Volume	3.5	0.7	0.6	0.8	1.9	1.4
Return	9.1	7.0	5.3	0.8	9.0	3.3
Cholesky ordering: GT, RV, Trading Volume, Return						
GT	23.2	49.9	53.0	25.4	54.2	60.1
RV	65.2	43.4	45.7	73.3	38.1	35.3
Trading Volume	2.6	0.3	0.7	0.6	1.3	1.6
Return	9.0	6.5	0.6	0.7	6.3	1.1
Panel B: With IV in VAR						
Cholesky ordering: RV, IV, Trading Volume, Return, GT						
GT	2.8	1.3	2.9	1.4	4.4	15.5
RV	60.8	61.9	53.8	63.9	64.6	64.3
IV	33.0	35.2	42.5	33.7	22.7	16.2
Trading Volume	3.1	1.4	0.6	0.6	1.6	1.5
Return	0.3	0.3	0.2	0.3	6.8	2.4
Cholesky ordering: GT, RV, IV, Trading Volume, Return						
GT	14.5	29.1	32.3	16.6	46.4	52.5
RV	50.9	37.3	32.4	52.7	31.1	32.4
IV	32.1	32.3	34.6	30.7	16.2	10.9
Trading Volume	2.2	1.0	0.6	0.7	1.2	1.4
Return	0.3	0.3	0.1	0.3	5.1	2.8
N	603	532	603	603	510	510

The table shows the percentage of forecast error variance of the log of the realized volatility for a forecast horizon of 12 weeks explained by the variables in the respective rows. GT, IV and RV stand for logs of the Google Trends search volume index, implied volatility and realized standard deviation, respectively. The beginning of the sample period is February 6, 2004 for S&P 500, Euro and Canadian Dollar, June 17, 2005 for DJIA, and November 25, 2005 for Gold and Crude Oil. The end of the sample period is August 28, 2015 for all markets. N indicates the number of observations measured in weeks.

Table 5
Encompassing Tests and MSPE Ratios for Stock Markets
Panel A: S&P 500

	Ave-ENC	Sup-ENC	MSPE Ratio
	Benchmark model: AR4		
AR4 + return	**27.20**	**53.04**	0.91
AR4 + trading volume	1.42	2.68	1.00
AR4 + GT	1.59	4.74	1.01
AR4 + IV	**101.92**	**160.12**	0.71
	Benchmark model: AR4 and 4 lags of IV		
AR4 + IV + return	-1.36	-0.25	1.02
AR4 + IV + trading volume	-0.46	0.47	1.01
AR4 + IV + GT	1.71	5.50	1.01
AR4 + IV + return + trading volume + GT	-0.16	3.97	1.04

Panel B: DJIA

	Ave-ENC	Sup-ENC	MSPE Ratio
	Benchmark model: AR4		
AR4 + return	**22.29**	**41.36**	0.90
AR4 + trading volume	-0.35	2.79	1.02
AR4 + GT	1.76	4.77	1.01
AR4 + IV	**82.19**	**127.11**	0.72
	Benchmark model: AR4 and 4 lags of IV		
AR4 + IV + return	-0.68	0.33	1.01
AR4 + IV + trading volume	0.59	1.88	1.01
AR4 + IV + GT	-0.15	0.99	1.01
AR4 + IV + return + trading volume + GT	-0.53	1.48	1.04

The first two columns show the average (Ave-ENC) and supremum (Sup-ENC) of recursive encompassing tests on logs of the realized volatilities. All predictors have four lags. 15% trimming is used on both sides of the sample. Bold text indicates statistical significance at 5% level based on critical values from Table 2b of Rossi and Inoue (2012) (10%, 5% and 1% critical values for the Ave-ENC tests with four additional variables are 1.916, 2.790 and 4.701, respectively, and the 10%, 5% and 1% critical values for the Sup-ENC tests with four additional variables are 4.508, 5.975 and 9.501, respectively.) The third column shows the ratio of the average mean square prediction errors (MSPE).

Table 6
Encompassing Tests and MSPE Ratios for Foreign Exchange Markets
Panel A: Euro

	Ave-ENC	Sup-ENC	MSPE Ratio
	Benchmark model: AR4		
AR4 + return	1.26	3.00	1.00
AR4 + trading volume	3.43	4.52	0.98
AR4 + GT	**13.39**	**19.57**	0.95
AR4 + IV	**140.38**	**205.66**	0.66
	Benchmark model: AR4 and 4 lags of IV		
AR4 + IV + return	-0.26	0.80	1.01
AR4 + IV + trading volume	**4.72**	**7.08**	0.98
AR4 + IV + GT	1.40	**2.33**	1.00
AR4 + IV + return + trading volume + GT	**5.44**	**7.86**	0.99

Panel B: Canadian Dollar

	Ave-ENC	Sup-ENC	MSPE Ratio
	Benchmark model: AR4		
AR4 + return	**3.89**	**6.06**	0.99
AR4 + trading volume	**4.45**	5.66	0.97
AR4 + GT	**5.10**	**7.59**	0.98
AR4 + IV	**112.11**	**153.84**	0.70
	Benchmark model: AR4 and 4 lags of IV		
AR4 + IV + return	1.87	4.02	1.00
AR4 + IV + trading volume	2.34	3.69	0.99
AR4 + IV + GT	2.45	3.96	1.00
AR4 + IV + return + trading volume + GT	**3.30**	5.74	1.00

The first two columns show the average (Ave-ENC) and supremum (Sup-ENC) of recursive encompassing tests on logs of the realized volatilities. All predictors have four lags. 15% trimming is used on both sides of the sample. Bold text indicates statistical significance at 5% level based on critical values from Table 2b of Rossi and Inoue (2012) (10%, 5% and 1% critical values for the Ave-ENC tests with four additional variables are 1.916, 2.790 and 4.701, respectively, and the 10%, 5% and 1% critical values for the Sup-ENC tests with four additional variables are 4.508, 5.975 and 9.501, respectively.) The third column shows the ratio of the average mean square prediction errors (MSPE).

Table 7
Encompassing Tests and MSPE Ratios for Commodity Markets
Panel A: Gold

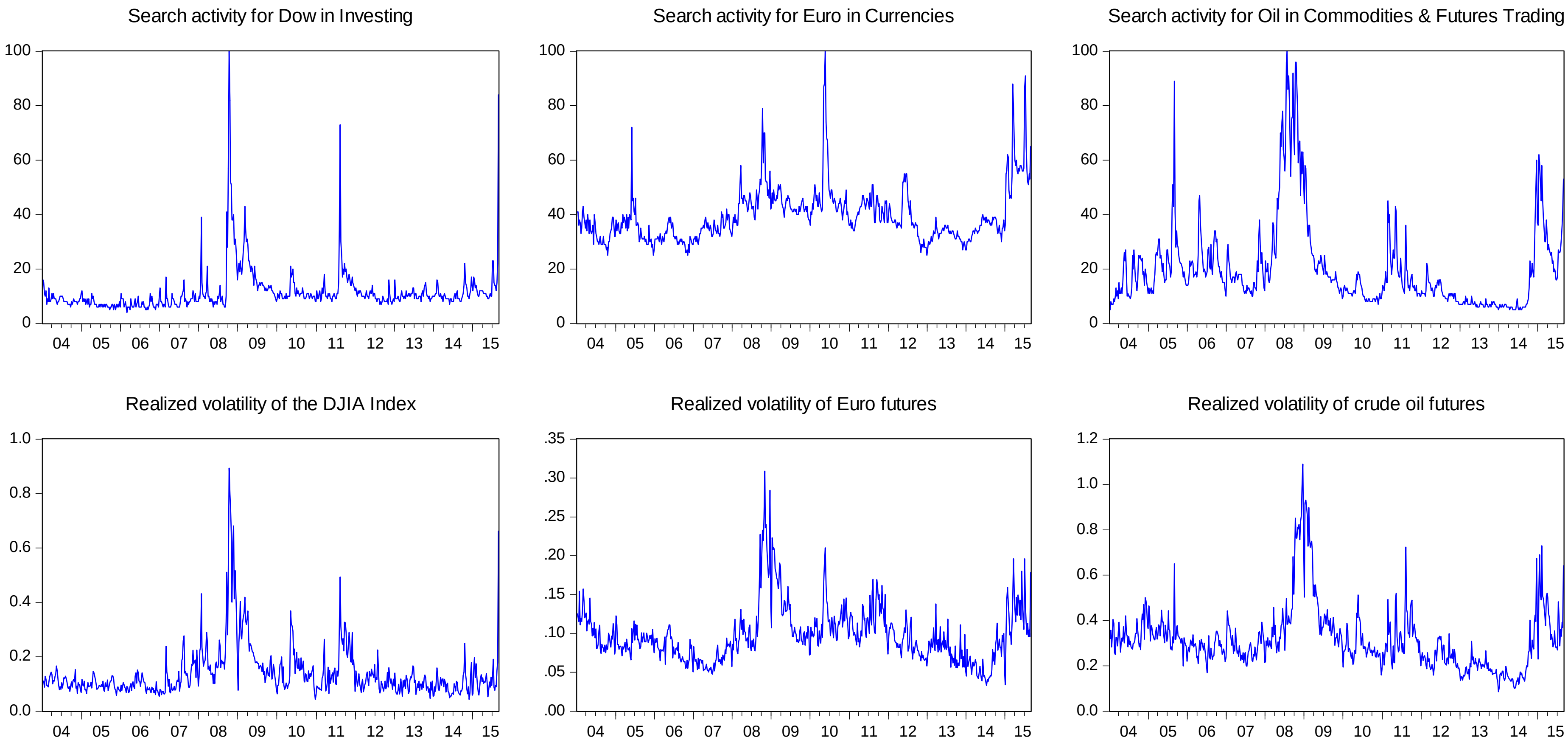
	Ave-ENC	Sup-ENC	Ave-MSPE
	Benchmark model: AR4		
AR4 + return	**5.56**	**14.19**	1.00
AR4 + trading volume	**8.05**	**17.41**	0.97
AR4 + GT	**9.95**	**29.99**	0.99
AR4 + IV	**62.56**	**100.53**	0.81
	Benchmark model: AR4 and 4 lags of IV		
AR4 + IV + return	**3.40**	**6.25**	0.99
AR4 + IV + trading volume	**7.22**	**14.81**	0.98
AR4 + IV + GT	**4.67**	**12.08**	0.99
AR4 + IV + return + trading volume + GT	**12.30**	**22.85**	0.97

Panel B: Crude Oil

	Ave-ENC	Sup-ENC	Ave-MSPE
	Benchmark model: AR4 model		
AR4 + return	**3.38**	4.92	0.98
AR4 + trading volume	**13.10**	**28.41**	0.96
AR4 + GT	**17.65**	**31.54**	0.92
AR4 + IV	**30.61**	**53.51**	0.90
	Benchmark model: AR4 and 4 lags of IV		
AR4 + IV + return	1.77	3.44	0.99
AR4 + IV + trading volume	**10.73**	**23.07**	0.97
AR4 + IV + GT	**11.94**	**19.79**	0.94
AR4 + IV + return + trading volume + GT	**26.67**	**45.92**	0.90

The first two columns show the average (Ave-ENC) and supremum (Sup-ENC) of recursive encompassing tests on logs of the realized volatilities. All predictors have four lags. 15% trimming is used on both sides of the sample. Bold text indicates statistical significance at 5% level based on critical values from Table 2b of Rossi and Inoue (2012) (10%, 5% and 1% critical values for the Ave-ENC tests with four additional variables are 1.916, 2.790 and 4.701, respectively, and the 10%, 5% and 1% critical values for the Sup-ENC tests with four additional variables are 4.508, 5.975 and 9.501, respectively.) The third column shows the ratio of the average mean square prediction errors (MSPE).

Figure 1
Realized Volatility and Google Search Activity

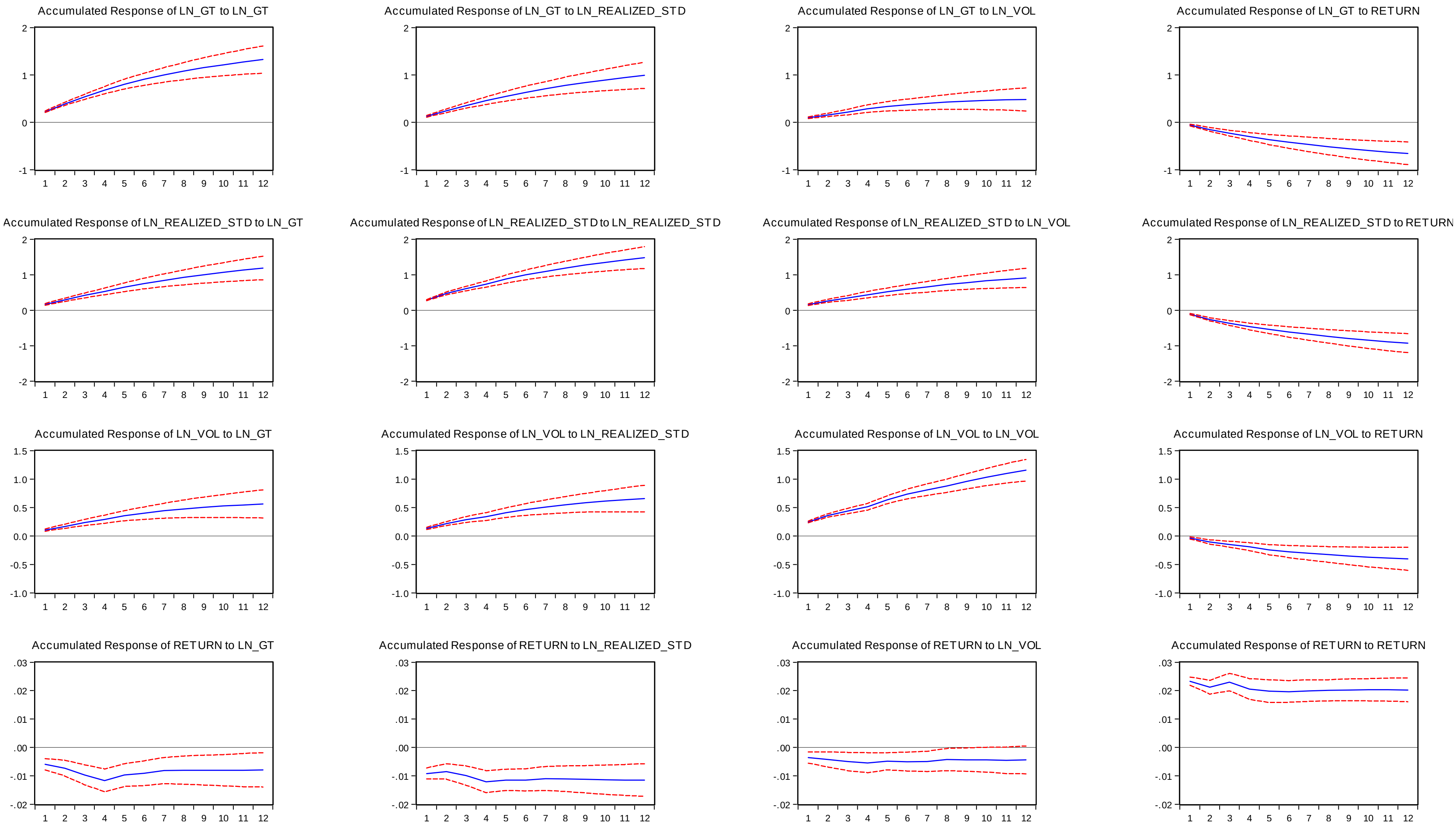


The sample period is from January 3, 2004 to August 28, 2015.

Figure 2
Impulse Responses

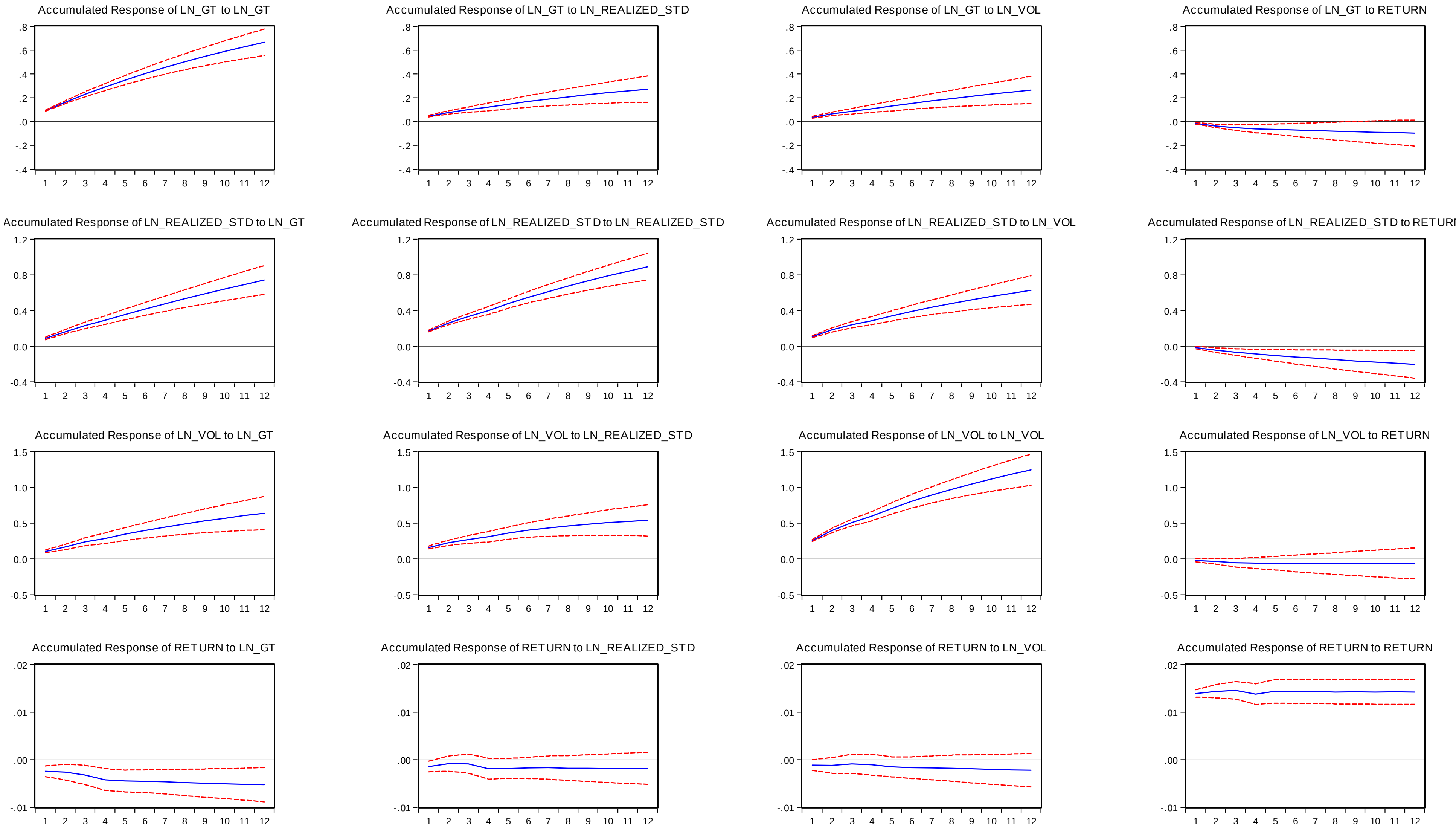
Panel A: DJIA

Accumulated Response to Generalized One S.D. Innovations ± 2 S.E.

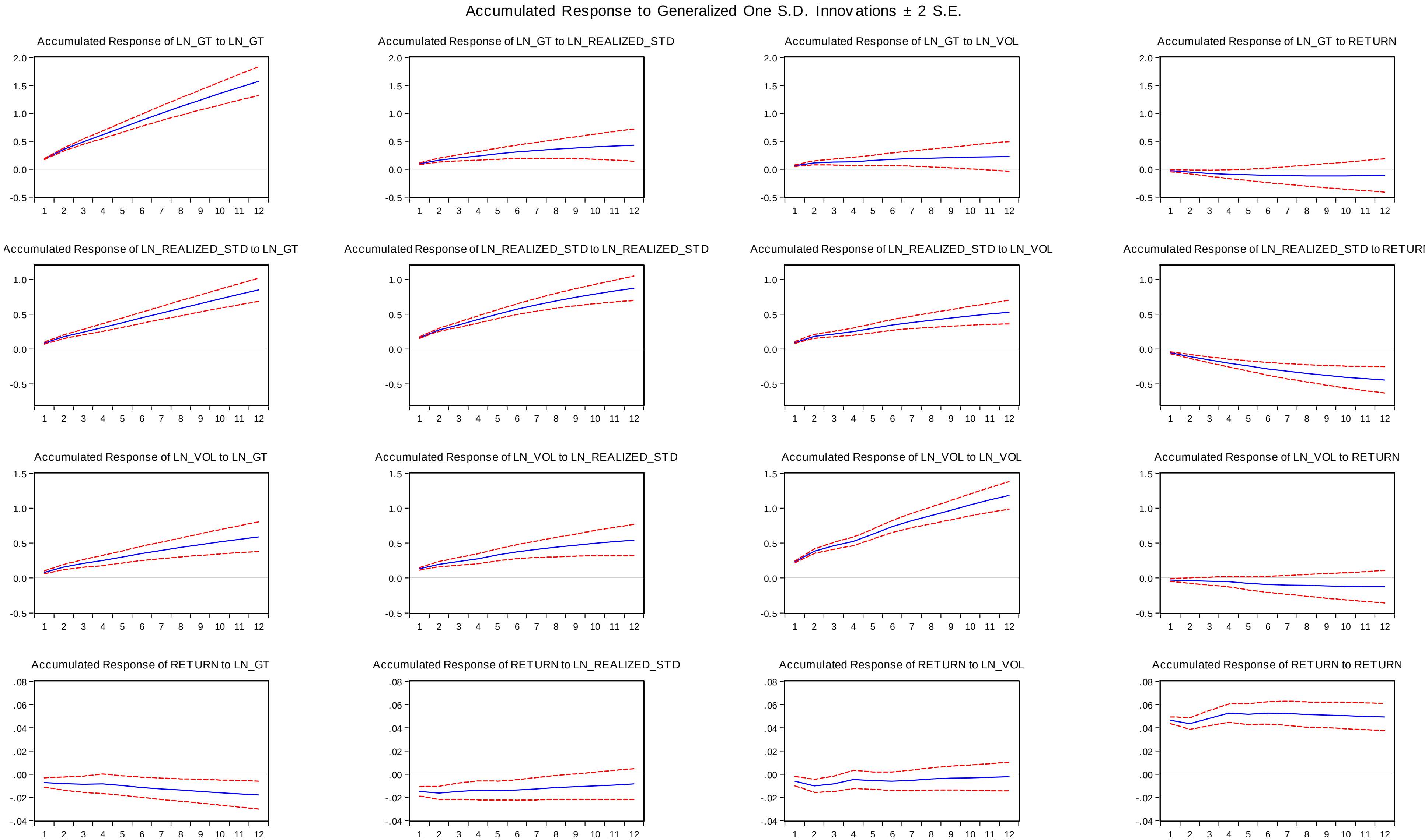


Panel B: Euro

Accumulated Response to Generalized One S.D. Innovations ± 2 S.E.



Panel C: Crude Oil



The solid (blue) lines show the accumulated responses to generalized one-standard deviation innovations. LN_GT, LN_REALIZED_STD, LN_VOL and RETURN stand for log of Google Trends search volume index, log of realized standard deviation, log of trading volume, and returns, respectively. The dashed (red) lines are two-standard-error bands. The values on the horizontal axis correspond to weeks.