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

Basistha, Arabinda and Kurov, Alexander and Wolfe, Marketa Halova

Skidmore College

2019

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

Arabinda Basistha, Alexander Kurov* and Marketa Halova Wolfe

May 22, 2019

Journal of Risk Model Validation, Forthcoming

Abstract

Recent empirical literature shows that Internet search activity is closely associated with volatility prediction in financial and commodity markets. In this study, we search for a benchmark model with available market-based predictors to evaluate the net contribution of the Internet search activity data in forecasting volatility. We conduct in-sample analysis and window-size robust out-of-sample forecasting analysis in multiple markets for robust model validation. The predictive power of the Internet search activity data disappears in the financial markets and substantially diminishes in the commodity markets once the model includes implied volatility. A further common component analysis shows that most of the predictive information contained in the Internet search activity is also present in implied volatility while implied volatility has additional predictive information that is not contained in the Internet search activity data.

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

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

---

*Corresponding author. Department of Finance, College of Business and Economics, West Virginia University, P.O. Box 6025, Morgantown, WV 26506, Tel: 304-293-7892, Fax: 304-293-3274, e-mail: alkurov@mail.wvu.edu.

We thank two anonymous referees, Theo Berger, Ana-Maria Fuertes, Gerard Kuper, Yihui Lan, Fabio Mattos, Nicolas Merener, Gabriel Power, Till Strohsal, Neda Todorova, Yaw-Huei Wang, and participants at 2015 NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management, the 14th INFINITI Conference on International Finance, 2016 Commodity Markets Conference, 2016 World Finance Conference, 2016 meetings of the Eastern Finance Association, Multinational Finance Society and European Financial Management Association, 25th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, Rensselaer Polytechnic Institute seminar and Skidmore College seminar for helpful comments and suggestions. We also thank Chen Gu for research assistance. 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, Morgantown, West Virginia.

Alexander Kurov is a professor of finance in the Department of Finance, West Virginia University, Morgantown, West Virginia.

Marketa Halova Wolfe is an associate professor of economics in the Department of Economics, Skidmore College, Saratoga Springs, New York.

Key Messages

- We evaluate usefulness of the Internet search activity data for forecasting volatility.
- The predictive power of the Internet search activity disappears in financial markets and diminishes in commodity markets once the model includes implied volatility.
- These results hold in both in-sample and out-of-sample analyses.

1. Introduction

Volatility is a key measure of financial market risk and is of keen interest to both academics and practitioners. Volatility forecasts are critical inputs in risk models and risk management, security valuation, and portfolio allocation for investors and financial institutions. While the literature on volatility forecasting is large, the search for new predictors of volatility continues. Although better structural models of volatility are of great academic interest, the crucial challenge for a practitioner is finding predictors that can be used to improve the volatility forecasts. Internet search activity is a new, theoretically interesting and promising predictor of volatility that can be viewed as a proxy for investor attention. Andrei and Hasler (2015) provide a theoretical model where the stock return variance increases with investor attention and uncertainty. This theoretical background provides motivation for analyzing the relation between the Internet search activity and volatility.

Recent empirical research shows that Internet search activity is associated with volatility in the financial and commodity markets. The literature documenting this link includes Da, Engelberg and Gao (2011), Vlastakis and Markellos (2012) for individual stocks, Chronopoulos, Papadimitriou and Vlastakis (2018), Dimpfl and Jank (2016) and Dzielinski (2012) for stock indices, Da, Engelberg and Gao (2015) for stock indices, exchange traded funds and Treasury bonds, Goddard, Kita and Wang (2015) and Smith (2012) for exchange rates, and Vozlyublennaia (2014) for stock and bond indices, gold and crude oil. Although the above studies examine this topic, the role that the Internet search activity plays in improving the accuracy of volatility predictions has not been consistently analyzed across different markets, forecast windows, and alternative predictors. These are important empirical model validation considerations that determine the overall usability of a predictor for a practitioner.

The previous literature has established that Internet search activity can have predictive information about volatility. It is important to ask, however, whether this predictor is useful in practice. If the predictive ability of the Internet search activity is subsumed by other available predictors, then the search activity data cannot be used to improve volatility forecasts. The aim of our paper is to evaluate the marginal predictive informational content of the Internet search activity in forecasting volatility in models with several market-based predictors for six financial and commodity markets. Since we do not know the most useful predictors in a given market, in a key intermediate step we search for the best forecast with the market-based predictors in each market. We use returns, trading volume, and option implied volatility as market-based predictors.[1] We conduct in-sample analysis and recursive, window-size robust out-of-sample forecasting analysis to quantify the predictive content of each predictor.

We find that adding the implied volatility delivers large gains in forecasting weekly realized volatility in all six markets. In contrast, the Internet search activity plays either no or rather limited role in forecasting realized volatility once the model includes implied volatility measures. We further show that most of the predictive information contained in the Internet search activity is also present in implied volatility while implied volatility has additional predictive information that is not contained in the Internet search activity. These results hold both in in-sample and out-of-sample analyses, providing mutual validation and strengthening our empirical evidence on a predictor's net contribution.

These results advance our understanding of the information captured by Internet search activity because its relation to implied volatility and other market-based predictors has not been

[1] Lamoureux and Lastrapes (1990) and Darrat, Zhong and Cheng (2007) argue in favor of using trading volume as a predictor of volatility. Christensen and Prabhala (1998), Szakmary, Ors, Kim and Davidson (2003) and Busch, Christensen, and Nielsen (2011) stress the predictive information present in implied volatility although recent studies by Han and Park (2013) and Kambouroudis, McMillan, and Tsakou (2016) point out limitations in using implied volatility.

comprehensively analyzed. Most studies on forecasting volatility with Internet search activity do not simultaneously consider the role of implied volatility in multiple markets.[2] Our paper, therefore, connects two strands of volatility forecasting literature: the theoretical and empirical literature on the role of implied volatility in forecasting volatility (for example, Christensen and Prabhala, (1998), Szakmary, Ors, Kim and Davidson (2003) and Busch, Christensen, and Nielsen (2011)), and the empirical literature on the role of Internet search activity in forecasting volatility mentioned above. While previous empirical studies on the role of Internet search activity surmise that the Internet search activity captures investor attention or information demand, we show that the Internet search activity is a relatively noisy predictor of realized volatility.[3] In contrast, implied volatility represents risk-neutral expectation of volatility reflected in option prices. This expectation of future volatility includes influences from the given market as well as broader forces such as economic and geopolitical developments and subsumes the information in the noisier Internet search activity.

2. Data

We describe the financial and commodity markets data in Section 2.1, the Internet search activity data in Section 2.2 and correlations between the two data sets in Section 2.3.

2.1. Financial and Commodity Market Data

We use data from the largest financial markets (stock and foreign exchange markets) and commodity markets. For each market type, we include two markets. For stock indices, we include the S&P 500 Index and the DJIA Index because they are among the world's most

[2] Dzielinski (2012) finds that Internet search activity remains a significant in-sample predictor of realized volatility even after controlling for implied volatility in the S&P 500.
[3] Cole (1969) elaborates how measurement errors in predictors affect forecast accuracy.

important stock indices and they were used in previous papers on the relation between Internet search activity and volatility (Dzielinski (2012), Vozlyublennaia (2014), and Dimpfl and Jank (2016)). 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. For commodities, we include gold from metal commodities and crude oil from energy commodities, the largest commodity markets in these sub-categories.

To measure implied volatility, we use the CBOE VIX and VXD that measure implied volatility with a horizon of 30 days for the S&P 500 Index and DJIA Index options, respectively. We use spot data for the S&P 500 and DJIA indices to obtain prices and trading volume stated as the number of shares of stocks in the indices. For commodities, we have implied volatility of options on the gold and crude oil nearby futures contracts and use gold and crude oil futures data to obtain prices and trading volume stated as the number of futures contracts. For foreign exchange, we use implied volatility of spot options with one week to maturity.[4] Since we do not have spot foreign exchange trading volume data, we use the Euro and Canadian dollar futures prices and trading volume stated as the number of futures contracts. Trading volumes in all markets are total weekly trading volumes. [5]

To measure volatility, we build on seminal work by Andersen, Bollerslev, Diebold and Labys (2001) who propose measuring volatility as the realized standard deviation:

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

[4] The results using implied volatility of foreign exchange options with one month to maturity are similar.
[5] The futures and spot market 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. Because futures contracts become increasingly illiquid close to their expiration, we use the next-to-maturity contracts when their daily trading volumes exceed the nearby contract volumes; the total weekly trading volume is then computed using this series. Futures returns are appropriately adjusted for contract rollovers.

where RV_t is the realized standard deviation and $r_{t,i}^2$ is the squared continuously compounded return in intraday interval i during week t. We follow existing literature (for example, Bollerslev, Tauchen and Zhou, 2009) and use 5-minute intraday intervals and include returns over non-trading periods.[6] Summary statistics for all variables are available in the Internet Appendix.

2.2. Internet Search Activity

We obtain Internet search activity data from the Google search engine using the Google Trends website (http://www.google.com/trends). Google has been the most popular Internet search engine throughout our sample period with 64% of the U.S. market share in February 2016 (Lella, 2016), and its search volume data has been used by the previous studies. This data shows how frequently search terms were used in the Google search engine. Google Trends computes a search volume index scaled by the maximum value over the selected time period. The index ranges from zero to 100 with the value of 100 representing the peak of search activity.

Google Trends organizes the search activity data by topic categories and regions. We use the U.S. region[7] and Finance category. Within the Finance category, we use the most appropriate subcategory for each market: the Investing subcategory for the 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. For search terms, we follow previous studies; for example, Dimpfl and Jank (2016) use search terms such as 'dow' and show that the results for the DJIA are robust to the choice of search terms such as *dow*, *dow jones*, and

[6] Fleming, Kirby and Ostdiek (2003) provide evidence that including overnight returns when computing the realized variance improves volatility estimates for stock, bond, and commodity futures. Overnight returns include returns over weekends and other non-trading days. In their analysis of stock return variances, French and Roll (1986) show that the variance of returns from Friday close to Monday close is only 10.7% higher on average than the variance of one-day returns.

[7] We conduct a robustness check with "Worldwide" region. The results (available upon request) are similar to the ones with the U.S. region.

djia. We download search volume indices for ‘s&p,’ ‘dow,’ ‘euro,’ ‘canadian,’ ‘gold,’ and ‘oil.’ Searching for these terms within the Finance subcategories ensures that our search data is not polluted by searches containing our search terms but unrelated to the financial and commodity markets; for example, the Canadian dollar data is not polluted with searches for ‘canadian hockey league’ because these searches do not show in the Currencies & Foreign Exchange subcategory of the Finance category.

Google Trends data is available since January 2004. We examine the sample period from January 4, 2004 for S&P 500, DJIA, Euro, and Canadian dollar and October 23, 2005 for gold and crude oil due to availability of the implied volatility data. The end of the sample period is August 28, 2015 for all markets.[8] We use weekly data because large portions of daily Google Trends data are missing.[9]

Figure 1 shows the logs of Google Trends search volume indices, realized volatility, and implied volatility during our sample period. We present only one market from each market type (DJIA for stock indices, Euro for foreign exchange, and crude oil for commodities) to save space and show the other markets (S&P 500, Canadian dollar, and gold) in the Internet Appendix. Periods with increased search activity appear to coincide with periods of high volatility as noted by the previous studies.

[Insert Figure 1 about here]

[8] As a robustness check, we split our sample period into three sub-periods: before, during and after the financial crisis using September 2007 and March 2009 as the crisis beginning and end, respectively. We then perform our analysis separately in each of these sub-periods. The results indicate that our findings are not driven by the financial crisis. These results are available upon request.

[9] Some of the previous studies, for example, Dimpfl and Jank (2016) and Chronopoulos, Papadimitriou and Vlastakis (2018), use daily data as their focus is on a stock index; however, the daily data has missing observations for the broader cross-section of markets that we examine, which renders the daily data analysis infeasible.

2.3. Correlations

We test all variables for stationarity using the Phillips and Perron (1988) test. Panel A of Table 1 shows that the null hypothesis of a unit root is strongly rejected for all variables in all markets. Panel B of Table 1 shows correlations of Google search volume with realized volatility, trading volume, returns, and implied volatility. Google search volume is positively and significantly correlated with realized volatility in all six markets (with correlations ranging from 0.33 in S&P 500 to 0.79 for crude oil) which is consistent with previous literature showing that Google search volume is associated with realized volatility. The correlation between Google search volume and implied volatility is also positive and significant in all six markets with correlation coefficients similar to those showing correlation between Google search volume and realized volatility. These numbers are also mostly higher than the other correlations of Google search volume with trading volume and returns, making the prediction evaluation between Google search volume and implied volatility an interesting empirical exercise.

[Insert Table 1 about here]

3. Methodology and Results

This section begins with two complementary approaches for examining the role of the Internet search activity in forecasting realized volatility: in-sample analysis in Section 3.1 and recursive, window-size robust out-of-sample forecasting in Section 3.2. We then expand our predictive analysis using an unobserved components model in Section 3.3. Robustness checks are discussed throughout the paper (robustness to using implied volatility of foreign exchange options with one month to maturity rather than one week to maturity in Section 2.1, the choice of Google Trends key words in Section 2.2, Google Trends region in Section 2.2, sample sub-periods to account for the 2007-2009 financial crisis in Section 2.2, GARCH model as the out-of-sample forecasting benchmark in Section 3.2.1, and out-of-sample forecasting window size in Section 3.2.1) and in the Internet Appendix (choice of lags in Section A.3 and an alternative measure of investor attention/information demand in Section A.5).

3.1. In-Sample Analysis Methodology and Results

The starting point of full-sample multivariate predictive analysis is usually a reduced form vector autoregressive model (VAR). VAR models are particularly useful in conducting Granger causality tests, a key indicator of potential predictability. Andersen, Bollerslev, Diebold and Labys (2003) and Andersen, Bollerslev and Meddahi (2004) also propose forecasting volatility by reduced-form models of realized volatility as they outperform models such as the generalized autoregressive conditional heteroskedasticity (GARCH) model. We employ a VAR model and Granger causality tests on the relation between Internet search activity and volatility and also provide an analysis of mean square prediction error ratios to characterize the incremental predictive content of the variables in the VAR.

Our VAR analysis proceeds in two steps. In the first step, we estimate the VAR with four variables (realized volatility, trading volume, return, and Google search volume) for each market to examine whether Google search volume predicts realized volatility in the full sample:

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

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: realized volatility, trading volume, return, and Google search volume. $\boldsymbol{\varepsilon_t}$ is a vector of random disturbances. This allows us to see whether the Google search volume helps in predicting realized volatility. We take natural logs of the realized volatility, trading volume, and Google search volume; this transformation reduces skewness and excess kurtosis of these variables. We include two lags of variables in the VAR based on the Schwarz information criterion and show in the Internet Appendix that our results are not affected by the choice of lags.

We begin by estimating the four-variable (realized volatility, trading volume, return, and Google search volume) VAR in equation (2) and use the results to perform Granger causality tests.[10] Table 2 Panel A shows the results. In five of the six markets, we reject the null hypothesis that Google search volume does not Granger cause realized volatility after controlling for lags of realized volatility, trading volume, and return. The Google search volume is also a useful predictor of trading volume in four of the six markets and returns in two of the six markets.

[Insert Table 2 about here]

Figure 2 shows impulse responses after one standard-deviation reduced form shocks. We present only key variables for one market from each asset class to save space; additional impulse response functions are available in the Internet Appendix. The first row shows the effect of a Google search volume shock on realized volatility. Increases in the Google search volume predict higher realized volatility. This is consistent with previous empirical studies that show the Google search volume predicts volatility.

[Insert Figure 2 about here]

In the second step, we add implied volatility as another endogenous variable to the VAR in equation (2) to examine whether implied volatility affects usefulness of the Google search volume for predicting realized volatility. The model then becomes:

$$\boldsymbol{y_t} = \boldsymbol{\gamma} + \sum_{j=1}^{2} \boldsymbol{\delta_j}\, \boldsymbol{y_{t-j}} + \boldsymbol{\epsilon_t}, \tag{3}$$

[10] The VAR coefficient estimates are not tabulated for brevity but are available upon request.

where $\boldsymbol{y_t}$ is a vector of five variables: realized volatility, trading volume, return, Google search volume, and implied volatility. Similar to equation (2), $\boldsymbol{\gamma}$ is a vector of constant terms, $\boldsymbol{\delta_j}$ is the vector of coefficients for lag j, and $\boldsymbol{\epsilon_t}$ is a vector of random disturbances. The highly significant coefficients on implied volatility in Table 2 Panel B indicate that implied volatility predicts realized volatility in all markets. This provides in-sample support for the argument that implied volatility contains useful predictive information for realized volatility as in Christensen and Prabhala (1998) and Busch, Christensen, and Nielsen (2011).[11] The impact of including implied volatility in the VAR is more striking in another respect: the significance of the Google search volume disappears in the financial markets and diminishes in the commodity markets. Interestingly, implied volatility predicts Google search volume in all markets whereas Google search volume predicts implied volatility in only three markets.

To characterize the incremental predictive content of variables in the VAR, we conduct an analysis of mean square prediction error (MSPE) ratios. We first estimate a restricted (or benchmark) VAR that contains the log of the realized volatility, log of trading volume, and returns. Based on this estimation, we compute the MSPE of realized volatility. We add the Google search volume and implied volatility in the model one at a time and then together. We compute the MSPEs of realized volatility based on these alternative unrestricted models. The ratios of the unrestricted model MSPEs to the restricted model MSPE are presented in Table 3. MSPE ratios below one indicate that the additional variable improves on the benchmark model's prediction of realized volatility. For all six markets, Google search volume improves the prediction of realized volatility. However, this improvement is generally small with the MSPE

[11] We used the Ljung-Box test to examine serial correlation in the residuals in the realized volatility equation of the VAR. In all six markets, the residual serial correlation in the realized volatility equation is low and not statistically significant. These results are available upon request.

ratios ranging from 0.99 to 0.93 indicating improvements of 1 percent to 7 percent. Adding implied volatility generates a much larger improvement with the MSPE ratios ranging from 0.85 to 0.69 indicating improvements of 15 to 31 percent.

Relative to the benchmark model containing implied volatility (shown in the bottom panel of Table 3), the incremental improvement in the in-sample forecasts of realized volatility from adding the Google search volume is small. Adding Google search volume reduces the MSPE by only about two percent for gold and about one percent for the S&P 500 index and Canadian dollar while generating essentially no improvement for the DJIA and Euro. The only exception is the crude oil market where the improvement in the MSPE is about five percent. Overall, the in-sample results indicate that usefulness of the Google search volume is limited once implied volatility is included in the model.

[Insert Table 3 about here]

3.2. Out-of-Sample Forecasting Methodology and Results

In-sample outcomes can be susceptible to such pitfalls as spurious associations and overfitting. An out-of-sample analysis is also closer to the scenario a practitioner faces in reality. We analyze the relation between the Internet search activity and volatility by conducting recursive out-of-sample evaluations that have been quite effective in reducing the in-sample issues. We use window-size robust evidence to further reduce the sensitivity of our evidence to the choice of the forecasting window. We also quantify the forecasting gains based on MSPE ratios as well as expected loss and provide a calculation of Value-at-Risk.[12]

[12] Smith (2012) mentions that Google searches help in out-of-sample forecasts of one-week ahead volatility in exchange rates. Dimpfl and Jank (2016) come to a similar conclusion for stock indices.

3.2.1. Forecast Encompassing Tests

The key out-of-sample evaluation concept we use is encompassing, which quantifies the marginal contribution of different predictors in nested models to evaluate the independent forecasting information in each predictor. It argues that if a nested Model 1 contains all relevant information for forecasting a target variable over an unrestricted Model 2, forecast errors of Model 1 (i.e., the difference between the actual values and Model 1 forecast values), should be close to forecast errors from Model 2. Otherwise, Model 2 provides additional information and is not encompassed by Model 1.

A potentially informative predictor can become less usable if its predictive power is sensitive to the window size of estimation. To address this issue, we use the Rossi and Inoue (2012) methods that are robust to the window size. Their recursive encompassing tests build on Clark and McCracken (2001) study that compares forecast errors in nested models. These tests are more appropriate for nested predictors than the popular Diebold-Mariano test (Diebold and Mariano (1995)). We denote the forecast errors from the restricted Model 1 and the unrestricted Model 2 by u_{1t} and u_{2t}, respectively. R denotes the number of observations used to estimate the parameters needed to form the first one-step-ahead forecast. Using T to denote the total number of observations, there are $T - R$ forecasts from restricted and unrestricted models. The encompassing test statistic is then:

$$ENC = \frac{\sum_{t=R+1}^{T} u_{1t}(u_{1t} - u_{2t})}{\sum_{t=R+1}^{T} u_{2t}^2}(T - R). \tag{4}$$

High values of the encompassing test statistic mean that u_{2t} is small compared to u_{1t}, indicating that the additional predictor improves on the benchmark model. To compute the encompassing tests recursively, we start at the lower end of the estimation window with R^L observations and by adding one observation at a time we go up to the upper end, R^U. We follow the Rossi and Inoue (2012) recommendation of 15% trimming on each side of the sample when choosing R^L and R^U; for example, the R^L corresponds to observations 91 and 77 in the S&P 500 and crude oil markets, respectively. Rossi and Inoue (2012) recommend using two versions of the encompassing test: $Sup\text{-}ENC = Sup_{R\in(R^L,...,R^U)}\{ENC(R)\}$ and $Ave\text{-}ENC = \frac{1}{R^U-R^L+1}\sum_{R=R^L}^{R^U} ENC(R)$ where the Sup-ENC is the maximum value among all windows, and the Ave-ENC computes the average value across all windows.

Our first benchmark model is a simple autoregressive (AR) model that forecasts the log of realized volatility using its lags.[13] Simple time series models have a long history of strong performance as benchmark models for forecast evaluation. Autoregressive models for the U.S. economy in Nelson (1972) and the random walk model for exchange rates in Meese and Rogoff (1983) perform well relative to more structural models. A similar argument for simple autoregressive benchmark models has also been put forward by Andersen, Bollerslev, Diebold, and Labys (2003) for univariate volatility predictions. Christiansen, Schmeling and Schrimpf (2012) also use a simple autoregressive benchmark to evaluate macroeconomic predictors of financial volatility in a variety of markets. Therefore, this simple AR2 model is our Model 1 with forecast error denoted by u_{1t}.[14]

However, the challenge we are facing is evaluating the usability of a predictor. This can mean that we may need to use different benchmark models in each market because the best prediction (within our set of variables) may vary across markets. Marcellino (2008) shows that it

---

[13] We use two lags to be consistent with the VAR. As in the in-sample analysis, we verify in the Internet Appendix that our results are not affected by the choice of lags. The assumption of a known fixed model lag structure makes this analysis a pseudo out-of-sample analysis.

[14] One alternative to using an autoregressive model as our benchmark is using a GARCH model. We compute the GARCH (1,1) conditional variance for each market and estimate the realized volatility regressions with the conditional variances as regressors. The fit of the GARCH variance in the regressions is 7 to 15 percent lower than that of the autoregressive model of realized volatility. This implies that the autoregressive model of realized volatility is a more difficult benchmark to overcome than the GARCH model.

is useful to compare the outcomes of the simple models against other candidate benchmark models. A crucial role of a good benchmark model is to critically evaluate the informational content of potential predictors, especially when those predictors are correlated as in our case. In our volatility forecast evaluations, there are multiple predictors potentially outperforming the simple autoregressive model, thereby weakening its case for serving as a strong benchmark model useful for a practitioner to determine the usability of a predictor.

To overcome this issue when a simple benchmark model is not a strong benchmark model, we search for the best predictive model within our set of variables in each market. We add – one at a time – two lags of return, log of trading volume, log of Google search volume, and log of implied volatility to the simple benchmark model to form our unrestricted model, Model 2. The forecast errors from these unrestricted models are denoted by u_{2t}. The first four lines of Table 4 Panels A and B report results of the encompassing tests. As explained above, high values of the encompassing test statistic mean that u_{2t} is small compared to u_{1t}, indicating that the additional predictor improves on the benchmark model. In all markets, among the four models (AR2 + two lags of return, trading volume, Google search volume, or implied volatility), it is the AR2 + implied volatility (IV) model that performs the best relative to the AR2 benchmark.

[Insert Table 4 about here]

3.2.2. Size of the Out-of-Sample Forecasting Gains

To quantify *how much* the forecast accuracy improves over the benchmark model, we compute the ratio of the unrestricted model MSPE to the benchmark model MSPE. The MSPEs are averaged across windows in computing the ratio. This ratio is reported in Panel C. For example, the ratio of 0.70 for the AR2+IV model in the S&P 500 stock index means that adding implied volatility to the AR2 model reduces the MSPE by 30 percent on average. Among the four unrestricted models, the AR2+IV model shows the lowest MSPE ratio in all markets, suggesting that it substantially outperforms the other models. This finding provides out-of-sample support for using implied volatility in short-horizon volatility forecasting.

Therefore, we adopt the AR2+IV as the new benchmark based on the size of the gains in all markets, robustness to window selection, and linear simplicity. It is against this market-based, expanded benchmark model that we evaluate informational content of the non-market based Google search volume.[15] We add two lags of Google search volume to see if it improves on the AR2+IV forecast. The last rows of each panel in Table 4 present these results. We compare the encompassing statistics in the last row to those in the third row. In the DJIA stock index, for example, we compare the Ave-ENC statistically significant value of 4.72 for the AR2+GT model to the statistically insignificant value of -0.96 for the AR2+IV+GT model. The amount of additional independent information in the Google search volume relative to AR2 is higher than the amount of additional independent information in the Google search volume relative to AR2+IV, indicating that once implied volatility is included in the model, the Google search volume does not contribute with additional forecasting information. The decrease in both the Ave-ENC and Sup-ENC values from the AR2+GT model to the AR2+IV+GT model is evident in five out of the six markets. The only exception is the S&P 500 stock index where the predictive contribution of the Google search volume is low even in the simple AR2 model as shown by the statistically insignificant Ave-ENC value of -0.16.

To quantify the size of the forecasting gains over the benchmark model produced by Google search volume, we compute the ratio of the unrestricted model MSPE to the benchmark

[15] We are not claiming that the AR2+IV model generates the best forecast available to all practitioners in each market. For example, recent research such as Andersen, Bollerslev and Diebold (2007), Lee and Mykland (2008), and Busch, Christensen and Nielsen (2011) suggests that modelling jumps can potentially improve forecasts. However, availability of better forecasts in each market will further strengthen our conclusions.

model MSPE. Specifically, on the AR2+IV+GT line, we compute the ratio of the AR2+IV+GT MSPE to the AR2+IV MSPE. In the DJIA and Euro markets, the ratio is 1.01, indicating that the Google search volume worsens the AR2+IV model. In the S&P 500, Canadian dollar, and gold markets, the ratio is 0.99 indicating that adding Google search volume to the new AR2+IV benchmark does decrease the MSPE, but this decrease is rather small. The only market that shows a larger decrease in the MSPE after inclusion of Google search volume is crude oil with MSPE ratio of 0.93.

To further address any potential misspecification of our initial benchmark model (AR2), we also experiment with using ARMA(1,1) as our benchmark model. This specification corresponds to a GARCH(1,1) type of model that has been frequently used in volatility modelling. Moreover, as Basak, Chan and Palma (2001) point out, ARMA models serve as an excellent approximation of fractionally integrated processes for predictive purposes.[16] To reduce the computational burden involved with window-size robust recursive estimations, we use Durbin (1960) and Koreisha and Pukkila (1989) regression-based method to estimate our ARMA-based models. We report our empirical results in Table 5. The results are very similar to the results in Table 4. Inclusion of implied volatility is strongly preferred in all markets over the basic ARMA model and provides large forecasting gains. The predictive information gain is lower from inclusion of Google Trends search volume when the benchmark models include IV beyond the ARMA specification. The AR2 and ARMA(1,1) models have almost identical mean square prediction errors (MSPE). We, therefore, retain the computationally simpler AR2 model for the rest of our analysis.

[Insert Table 5 about here]

[16] In-sample estimates of fractionally integrated processes show that only two of the six markets (Canadian dollar and crude oil) have statistically significant long memory at 5 percent level. These results are available upon request.

Overall, these out-of-sample results show that the success of predictors depends on the selection of the benchmark model. These results also highlight the usefulness of searching for a strong benchmark model in practitioners' empirical model evaluations. Although Google search volume appears to be informative and improves on the simple benchmark model forecasts of realized volatility, its net contribution becomes limited when compared against a market-based expanded benchmark model that contains implied volatility. These outcomes are similar to the results from the in-sample analysis in Section 3.1 for all six markets and not sensitive to the choice of the window size.

An alternative way to compare out-of-sample forecasting gains from models with different predictors is based on expected loss, i.e., the difference between the actual value of volatility and the volatility forecast. Patton (2011) shows that the mean square error (MSE, same as MSPE in the previous discussion) loss function commonly used in the volatility forecasting literature is robust to the presence of measurement error in the volatility proxy. This loss function is:

$$L\left(\hat{\sigma}_t^2, h_{t|t-1}\right) = \left(\hat{\sigma}_t^2 - h_{t|t-1}\right)^2, \tag{5}$$

where $\hat{\sigma}_t^2$ is an *ex post* volatility proxy and $h_{t|t-1}$ is the volatility forecast based on information available at time $t-1$. Following Patton (2011) and Brownlees, Engle and Kelly (2012), we use realized volatility as the *ex post* proxy for volatility. To provide a simple economic interpretation of the average forecast losses, we employ the approach proposed by Brownlees et al. (2012). They argue that the typical volatility forecasts x are implied by the average forecast losses and the average value of the *ex post* volatility proxy $\hat{\sigma}_{\text{mean}}^2$ over the sample period:

$$(\hat{\sigma}_{\text{mean}}^2 - x^2)^2 = \text{average MSE} \tag{6}$$

Solving this equation provides two volatility forecasts, x_{under} and x_{over}, representing the typical underestimate and overestimate of volatility produced by a given model. One can use these typical forecasts to calculate the average volatility error reduction of Model 2 relative to Model 1 as:

$$\text{Volatility error reduction}_i = 1 - \left| \frac{x_{\text{model2},i} - \hat{\sigma}_{\text{mean}}}{x_{\text{model1},i} - \hat{\sigma}_{\text{mean}}} \right|, \quad i \in \{\text{under, over}\}. \tag{7}$$

For $\hat{\sigma}_{\text{mean}}$, we use the average realized standard deviation of returns for the second half of our sample period. The out-of-sample volatility forecasts used to compute forecast losses are also obtained for the second half of our sample period. We use our MSEs for logarithm of standard deviation of returns to solve for two volatility forecasts, x_{under} and x_{over}. Model 1 and Model 2 in our context are autoregressive out-of-sample volatility forecasting models that contain lags of the Google search volume index or lags of implied volatility or lags of both of these variables. Table 6 presents the estimates of volatility error reduction. The estimates are qualitatively similar across all six markets. The autoregressive model that includes two lags of implied volatility produces meaningful reductions in volatility forecast errors relative to the autoregressive model that includes two lags of Google search volume. For example, for the S&P 500 index this forecast error reduction ranges from 13.70% to 18.05%. However, adding lags of Google search volume to a model that includes lags of implied volatility produces little, if any, improvement in volatility forecasts in all markets except crude oil.

[Insert Table 6 about here]

Following Brownlees et al. (2012), we also compute a one-week-ahead 1% Value-at-Risk (VaR) error reduction for the S&P 500 and DJIA indices as:

$$\text{VaR error reduction}_i = 1 - \left| \frac{\Phi^{-1}(0.01; 0,\ x_{\text{model2},i}) - \Phi^{-1}(0.01; 0,\ \hat{\sigma}_{\text{mean}})}{\Phi^{-1}(0.01; 0,\ x_{\text{model1},i}) - \Phi^{-1}(0.01; 0,\ \hat{\sigma}_{\text{mean}})} \right|, \quad i \in \{\text{under, over}\}, \tag{8}$$

where $\Phi^{-1}(0.01; \mu, \sigma)$ is the inverse cumulative distribution function of a normal random variable. Consistent with Brownlees et al. (2012), the estimates of the VaR forecast error reduction are equal to the corresponding estimates of the volatility forecast error reduction shown in Table 6 to the nearest one-hundredth of a percent. Therefore, adding implied volatility to the volatility forecasting model produces an economically meaningful improvement in portfolio risk management. In contrast, the improvement in VaR forecasts from adding Google search volume to the forecasting model that includes implied volatility is low.

3.3. Unobserved Components Model of Implied Volatility and Google Search Volume Residuals

The results from the in-sample analysis in Section 3.1 and out-of-sample analysis in Section 3.2 suggest that there could be an important common component between implied volatility and Google search volume that contains predictive information about future volatility. However, we do not yet know either the size of that common component or the predictive information present in individual components. In this section, we, therefore, conduct an unobserved components analysis to better understand the sources of the predictive information.

The starting point is the reduced form VAR in equation (3) with two lags and five variables (realized volatility, trading volume, return, Google search volume, and implied volatility). The unobserved components analysis uses the reduced form residuals that are estimates of the random disturbances, $\boldsymbol{\epsilon_t}$. We denote these estimated reduced form residuals by u_t. Specifically, we are interested in the residual of implied volatility denoted by $u_{IV,t}$ and the residual of Google search volume denoted by $u_{GT,t}$. The variances of these two residuals and the correlations between these two residuals are presented in Table 7. The results confirm our conjecture of moderate positive correlation ranging from 0.161 for S&P 500 to 0.433 for gold.

[Insert Table 7 about here]

Using these residuals of implied volatility and Google search volume, we estimate a bivariate unobserved components model:

$$\begin{aligned} u_{IV,t} &= u_t^c + u_{IV,t}^i \\ u_{GT,t} &= u_t^c + u_{GT,t}^i \end{aligned} \tag{9}$$

where we allow the residuals to have a common component, u_t^c. We also allow for individual components: the implied volatility residual has an individual component denoted by $u_{IV,t}^i$, and the Google search volume residual has an individual component denoted by $u_{GT,t}^i$. The three unobserved components (i.e., u_t^c, $u_{IV,t}^i$, and $u_{GT,t}^i$) are assumed to be mutually orthogonal.

The three variance parameters of model in equation (9) are estimated using maximum likelihood. We then estimate the three components of our model in equation (9) using a Kalman filter based on maximum likelihood estimates of the variances. A Kalman filter estimates unobserved dynamic state variables from observed time series data. It consists of two parts: a measurement equation part stating how the unobserved state variables are linked to the observed variables and a transition equation part that specifies the dynamic equations of the unobserved state variables. The two parts together form the state space system.[17] In our case, equation (9) is the measurement equation. In the transition equations, our three serially uncorrelated unobserved state variables are effectively following the distributions below:

$$u_t^c \sim N\big(0, V(u^c)\big), u_{IV,t}^i \sim N\big(0, V(u_{IV}^i)\big), u_{GT,t}^i \sim N\left(0, V\left(u_{GT}^i\right)\right), \tag{10}$$

where $V(u)$ corresponds to the variances of the specific state variables u.

The last two columns of Table 7 show the variance of the common component as a share of the implied volatility residual's variance ($\frac{V(u^c)}{V(u_{IV})}$ in the fourth column) and as a share of the

[17] Kim and Nelson (1999) provide a detailed discussion of state space models and Kalman filter.

Google search volume residual's variance ($\frac{V(u^c)}{V(u_{GT})}$ in the fifth column). The variance of the common component is a substantial share of both the implied volatility residual's variance and the Google search volume residual's variance. The variance of the common component as a share of the implied volatility residual's variance ranges from 0.179 (in S&P 500) to 0.622 (in DJIA). The variance of the common component as a share of the Google search volume residual's variance ranges from 0.145 (in S&P 500) to 0.663 (in Euro). However, when we compare these two shares, we do not find any pattern: $\frac{V(u^c)}{V(u_{IV})}$ is higher than $\frac{V(u^c)}{V(u_{GT})}$ in three of the markets (S&P 500, DJIA, and crude oil) and lower than $\frac{V(u^c)}{V(u_{GT})}$ in three of the markets (Euro, Canadian dollar, and gold).

We then use the two-sided estimates of each of the three components estimated in equation (9) to estimate predictive regressions of realized volatility denoted as RV_t:

$$RV_t = \beta_0 + \sum_i \sum_{j=1}^{2} \beta_j^i Z_{t-j}^i + \sum_{j=1}^{2} \gamma_{k,j} u_{k,t-j} + v_t \tag{11}$$

In the predictive equation (11), Z_t^i denotes the three other predictors, i.e., realized volatility, returns, and log of trading volume. The variables $u_{k,t}$ denote the specific unobserved component used in the regression where k denotes the common component c or the individual components IV or GT. We use two lags to be consistent with the analysis in Sections 3.1 and 3.2.

Panel A of Table 8 reports the fit measured by R^2 for these predictive regressions without the three other predictors. The results show that the common component, u_t^c, and the individual component of the implied volatility, $u_{IV,t}^i$, are important for the realized volatility prediction: the contribution of u_t^c ranges from 9 to 23 percent, and the contribution of $u_{IV,t}^i$, ranges from 5 to 20 percent. However, the contribution of the individual component of Google search volume, $u_{GT,t}^{i}$, is very low with a maximum of around 2 percent.[18]

Overall, the analysis of the unobserved components model of implied volatility and Google search volume residuals shows that most of the predictive information about realized volatility contained in Google search volume is also captured in implied volatility. In contrast, implied volatility has additional predictive information that is not captured in the Google search volume data. This suggests that the Google search volume captures a subset of information in implied volatility.

[Insert Table 8 about here]

4. Conclusion

In this study, we analyze the usability of Internet search activity data for forecasting volatility in the financial and commodity markets. We search for a benchmark model with available market-based predictors to evaluate the net contribution of the Internet search activity data in forecasting volatility. While the Internet search activity has predictive power when implied volatility is not included in the model, its usefulness for forecasting volatility disappears in the financial markets and substantially diminishes in the commodity markets once implied volatility is included in the model. We highlight this using both in-sample analysis and recursive, window-size robust out-of-sample forecasting analysis. A further unobserved component analysis shows that most of the predictive information contained in the Internet search activity is also present in implied

[18] We follow up in Panel B by including two lags of realized volatility, returns, and log of trading volume as additional predictors used in Section 3.1. The fits of the regressions show a large increase in all three columns. Although the collinearity of the variables compresses the differences in fit between the columns, the basic pattern of the individual component of Google search volume showing the lowest fit still holds in all six markets.

volatility while implied volatility has additional predictive information that is not contained in the Internet search activity data.

This is not to claim that Internet search activity data is not useful in other fields. Usefulness of this data has been shown in other contexts including Ginsberg et al. (2009) for detecting influenza epidemics, Choi and Varian (2012) for predicting automobile sales, consumer confidence, unemployment claims, and travel destinations, and Wu and Brynjolfsson (2015) for predicting housing market trends. In these scenarios where one cannot take advantage of information contained in options, the Internet search activity could be a useful predictor. However, when implied volatility is available, usefulness of the Internet search activity data is limited, which is an important finding for researchers, practitioners, and policy-makers who forecast volatility of prices in the financial and commodity markets.

References

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (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. https://doi.org/10.1162/rest.89.4.701.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* **96**(453), 42–55. https://doi.org/10.1198/016214501750332965.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modelling and forecasting realized volatility. *Econometrica* **71**(2), 579–625. https://doi.org/10.1111/1468-0262.00418.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2004). Analytical evaluation of volatility forecasts. *International Economic Review* **45**(4), 1079–1110. https://doi.org/10.1111/j.0020-6598.2004.00298.x.

Andrei, D., and Hasler, M. (2015). Investor attention and stock market volatility, *Review of Financial Studies* **28**(1), 33–72. https://doi.org/10.1093/rfs/hhu059.

Basak, G. K., Chan, N. H., and Palma, W. (2001). The approximation of long memory by an ARMA model. *Journal of Forecasting* **20**(6), 367–389. https://doi.org/10.1002/for.799.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies* **22**(11), 4463–4492. https://doi.org/10.1093/rfs/hhp008.

Brownlees, C., Engle R., and Kelly, B. (2012). A practical guide to volatility forecasting through calm and storm. *Journal of Risk* **14**(2), 3–22. https://doi.org/10.21314/JOR.2012.237.

Busch, T., Christensen, B. J., and 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**(1), 48–57. https://doi.org/10.1016/j.jeconom.2010.03.014.

Choi, H., and Varian, H. (2012). Predicting the present with Google Trends. *Economic Record* **88**(s1), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x.

Christensen, B. J., and Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics* **50**(2), 125–150. https://doi.org/10.1016/S0304-405X(98)00034-8.

Christiansen, C., Schmeling, M., and Schrimpf, A. (2012). A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics* **27**(6), 956–977. https://doi.org/10.1002/jae.2298.

Chronopoulos, D. K., Papadimitriou, F.I., and Vlastakis, N. (2018). Information demand and stock return predictability. *Journal of International Money and Finance* **80**, 59–74. https://doi.org/10.1016/j.jimonfin.2017.10.001.

Clark, T. E., and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* **105**(1), 85–110. https://doi.org/10.1016/S0304-4076(01)00071-9.

Cole, R. (1969). Data errors and forecasting accuracy, in J. A. Mincer (Ed.) *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, 47–82, NBER.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *Journal of Finance* **66**(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x.

Da, Z., Engelberg, J., and Gao, P. (2015). The sum of all FEARS: Investor sentiment and asset prices. *Review of Financial Studies* **28**(1), 1–32. https://doi.org/10.1093/rfs/hhu072.

Darrat, A. F., Zhong, M., and Cheng, L. T. W. (2007). Intraday volume and volatility relations with and without public news. *Journal of Banking & Finance* **31**(9), 2711–2729. https://doi.org/10.1016/j.jbankfin.2006.11.019.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13**(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599.

Dimpfl, T., and Jank, S. (2016). Can Internet search queries help to predict stock market volatility. *European Financial Management* **22**(2), 171–192. https://doi.org/10.1111/eufm.12058.

Durbin, J. (1960). The fitting of time-series models. *Review of the International Statistical Institute* **28**(3), 233–244. https://doi.org/10.2307/1401322.

Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters* **9**(3), 167–175. https://doi.org/10.1016/j.frl.2011.10.003.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics*, **67**(3), 473–509. https://doi.org/10.1016/S0304-405X(02)00259-3.

French, F. and Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, **17**(1), 5–26. https://doi.org/10.1016/0304-405X(86)90004-8.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., and Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012–1015. https://doi.org/10.1038/nature0763410.1038/nature07634.

Goddard, J., Kita, A., and Wang, Q. (2015). Investor attention and FX market volatility. *Journal of International Financial Markets, Institutions & Money* **38**, 79–96. https://doi.org/10.1016/j.intfin.2015.05.001.

Han, H., and Park, M. D. (2013). Comparison of realized measure and implied volatility in forecasting volatility. *Journal of Forecasting* **32**(6), 522-533. https://doi.org/10.1002/for.2253.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions ASME Journal of Basic Engineering* **82**(1), 35-45. https://doi.org/10.1115/1.3662552.

Kambouroudis, D. S., McMillan, D. G., and Tsakou, K. (2016). Forecasting stock return volatility: A comparison of GARCH, implied volatility, and realized volatility models. *Journal of Futures Markets* **36**(12), 1127–1163. https://doi.org/10.1002/fut.21783.

Kim, C.-J., and Nelson, C. (1999). State-space models with regime switching. *MIT Press*, Cambridge, MA.

Koreisha, S., and Pukkila, T. (1989). Fast linear estimation methods for vector autoregressive moving-average models. *Journal of Time Series Analysis* **10**(4), 325–339. https://doi.org/10.1111/j.1467-9892.1989.tb00032.x.

Lamoureux, C. G., and Lastrapes, W. D. (1990). Heteroskedasticity in stock return data: Volume versus GARCH effects. *Journal of Finance* **45**(1), 221–229. https://doi.org/10.1111/j.1540-6261.1990.tb05088.x.

Lee, S. S., and Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies* **21**(6), 2535–2563. https://doi.org/10.1093/rfs/hhm056.

Lella, A. (2016, March 16). Comscore Releases February 2016 U.S. Desktop Search Engine Rankings. Retrieved from https://www.comscore.com/Insights/Rankings/comScore-Releases-February-2016-US-Desktop-Search-Engine-Rankings.

Marcellino, M. (2008). A Linear Benchmark for Forecasting GDP Growth and Inflation*?* *Journal of Forecasting* **27**(4), 304-340. https://doi.org/10.1002/for.1059.

Meese, R. A., and Rogoff, K. (1983) Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics* **14**, 3–24. https://doi.org/10.1016/0022-1996(83)90017-X.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review* **62**(5), 902–917.

Patton, A. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* **160**(1), 246–256. https://doi.org/10.1016/j.jeconom.2010.03.034.

Phillips, P. C. B., and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* **75**(2), 335–346. https://doi.org/10.1093/biomet/75.2.335.

Rossi, B., and Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics* **30**(3), 432–453. https://doi.org/10.1080/07350015.2012.693850.

Smith, G. P. (2012). Google Internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters* **9**(2), 103–110. https://doi.org/10.1016/j.frl.2012.03.003.

Szakmary, A., Ors, E., Kim, J. K., and Davidson, W. N. III (2003). The predictive power of implied volatility: Evidence from 35 futures markets. *Journal of Banking & Finance* **27**(11), 2151–2175. https://doi.org/10.1016/S0378-4266(02)00323-0.

Vlastakis, N., and Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance* **36**(6), 1808–1821. https://doi.org/10.1016/j.jbankfin.2012.02.007.

Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance* **41**, 17–35. https://doi.org/10.1016/j.jbankfin.2013.12.010.

Wu, L., and 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
Unit Root Test and Correlations

Panel A. Phillips and Perron (1988) Unit Root Test

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
Google Trends SVI	**-8.10 (0.00)**	**-5.64 (0.00)**	**-4.69 (0.00)**	**-3.95 (0.00)**	**-4.28 (0.00)**	**-3.52 (0.00)**
Realized volatility	**-7.21 (0.00)**	**-7.36 (0.00)**	**-6.40 (0.00)**	**-5.04 (0.00)**	**-9.95 (0.00)**	**-4.87 (0.00)**
Trading volume	**-9.33 (0.00)**	**-8.27 (0.00)**	**-6.87 (0.00)**	**-5.38 (0.00)**	**-5.16 (0.00)**	**-4.66 (0.00)**
Return	**-26.29 (0.00)**	**-26.60 (0.00)**	**-24.16 (0.00)**	**-25.77 (0.00)**	**-25.23(0.00)**	**-25.63 (0.00)**
Implied volatility	**-3.79 (0.00)**	**-3.78 (0.00)**	**-6.02 (0.00)**	**-4.06 (0.00)**	**-5.63 (0.00)**	**-3.19 (0.02)**
N	608	608	608	608	514	514

Panel B. Correlations of Google Search Activity with Other Variables

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
Correlation with:						
Realized volatility	**0.33 (0.00)**	**0.75 (0.00)**	**0.65 (0.00)**	**0.51 (0.00)**	**0.59 (0.00)**	**0.79 (0.00)**
Trading volume	**0.16 (0.00)**	**0.14 (0.00)**	**0.49 (0.00)**	**-0.46 (0.00)**	**0.65 (0.00)**	**0.15 (0.00)**
Return	**-0.10 (0.02)**	**-0.13 (0.00)**	**-0.14 (0.00)**	-0.02 (0.69)	-0.05 (0.30)	**-0.18 (0.00)**
Implied volatility	**0.25 (0.00)**	**0.76 (0.00)**	**0.65 (0.00)**	**0.47 (0.00)**	**0.52 (0.00)**	**0.77 (0.00)**
N	608	608	608	608	514	514

Panel A shows results of the Phillips and Perron (1988) unit root test. The null hypotheses is that the variable contains a unit root. Panel B shows correlations of Google search activity with realized volatility, trading volume, returns, and implied volatility. Log transformation is used for the Google Trends search volume index, realized volatility, 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, DJIA, Euro, and Canadian dollar, 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 2
Granger Causality Tests

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
	Panel A: Without IV in VAR					
GT → RV	4.5 (0.10)	**16.1 (0.00)**	**17.7 (0.00)**	**11.0 (0.00)**	**21.4 (0.00)**	**39.1 (0.00)**
RV → GT	5.2 (0.07)	**8.1 (0.02)**	3.2 (0.20)	3.5 (0.17)	**10.9 (0.00)**	**13.1 (0.00)**
GT → Trading volume	**9.2 (0.01)**	1.8 (0.40)	**10.9 (0.00)**	4.0 (0.13)	**22.2 (0.00)**	**15.1 (0.00)**
Trading volume → GT	**11.0 (0.00)**	**13.9 (0.00)**	0.5 (0.77)	**7.5 (0.02)**	5.7 (0.06)	2.7 (0.25)
GT → Return	**6.4 (0.04)**	5.1 (0.08)	2.7 (0.26)	0.5 (0.79)	1.5 (0.47)	**8.8 (0.01)**
Return → GT	**17.2 (0.00)**	**29.8 (0.00)**	**6.2 (0.05)**	0.5 (0.78)	**13.3 (0.00)**	0.2 (0.92)
	Panel B: With IV in VAR					
IV → RV	**149 (0.00)**	**128 (0.00)**	**243 (0.00)**	**162 (0.00)**	**110 (0.00)**	**74.1 (0.00)**
RV → IV	2.3 (0.32)	3.9 (0.14)	**14.1 (0.00)**	**21.4 (0.00)**	**8.6 (0.01)**	**11.1 (0.00)**
GT → RV	4.8 (0.09)	1.0 (0.61)	1.4 (0.50)	3.3 (0.20)	**9.8 (0.01)**	**28.5 (0.00)**
RV → GT	5.6 (0.06)	2.7 (0.26)	**39.8 (0.00)**	**11.9 (0.00)**	**27.2 (0.00)**	**33.6 (0.00)**
GT → IV	4.8 (0.09)	**11.1 (0.00)**	5.6 (0.06)	2.4 (0.30)	**8.6 (0.01)**	**14.3 (0.00)**
IV → GT	**37.6 (0.00)**	**48.3 (0.00)**	**61.0 (0.00)**	**20.4 (0.00)**	**25.0 (0.00)**	**41.7 (0.00)**
GT → Trading volume	**6.5 (0.04)**	**8.0 (0.02)**	4.7 (0.10)	5.7 (0.06)	**15.5 (0.00)**	**8.9 (0.01)**
Trading volume → GT	**16.0 (0.00)**	**12.7 (0.00)**	1.1 (0.59)	**9.8 (0.01)**	**9.7 (0.01)**	1.3 (0.52)
GT → Return	**6.4 (0.04)**	5.2 (0.07)	1.5 (0.47)	0.5 (0.78)	1.5 (0.48)	**8.5 (0.01)**
Return → GT	1.1 (0.59)	2.9 (0.24)	3.3 (0.19)	0.3 (0.85)	**8.9 (0.01)**	1.0 (0.62)
N	606	606	606	606	512	512

The table shows Wald test statistics of VAR Granger causality tests. Logs are used for realized volatility (RV), trading volume, Google Trends search volume index (GT), and implied volatility (IV). The VAR specification includes two lags of all variables. To save space, only the relations most important to our analysis are shown; the other relations are available upon request. *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, DJIA, Euro, and Canadian dollar, 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
MSPE Ratios for Realized Volatility

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
Benchmark model comprising RV, Trading Volume, and Return						
Benchmark model + GT	0.99	0.97	0.97	0.98	0.96	0.93
Benchmark model + IV	0.80	0.80	0.69	0.78	0.80	0.85
Benchmark model + GT + IV	0.79	0.80	0.69	0.77	0.79	0.81
Benchmark model comprising RV, Trading Volume, Return, and IV						
Benchmark model + GT	0.99	1.00	1.00	0.99	0.98	0.95
N	606	606	606	606	512	512

This table shows the ratios of in-sample mean square prediction errors (MSPEs) of alternative models for realized volatility. The denominator of the ratio is the MSPE of realized volatility computed from the restricted (or benchmark) VAR. The numerator of the ratio is the MSPE of realized volatility computed from the unrestricted VAR that, in addition to the variables included in the benchmark model, contains variables listed in the first column of the table. Logs are used for realized volatility (RV), trading volume, Google Trends search volume index (GT), and implied volatility (IV). The beginning of the sample period is January 4, 2004 for S&P 500, DJIA, Euro, and Canadian dollar, 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 4
Encompassing Tests and MSPE Ratios

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
			Panel A: Ave-ENC			
AR2 + return	**23.34**	**24.49**	**2.03**	**4.65**	**2.14**	0.17
AR2 + trading volume	1.52	-1.11	**3.54**	**7.47**	**9.01**	**14.19**
AR2 + GT	-0.16	**4.72**	**8.57**	**5.83**	**6.35**	**18.45**
AR2 + IV	**104.66**	**94.35**	**156.32**	**111.18**	**67.84**	**43.34**
AR2 + IV + GT*	1.91	-0.96	-0.57	**4.20**	**5.10**	**12.29**
			Panel B: Sup-ENC			
AR2 + return	**44.85**	**45.81**	3.24	**8.24**	**8.39**	0.94
AR2 + trading volume	2.69	1.45	**4.85**	**10.38**	**20.74**	**29.74**
AR2 + GT	0.30	**9.02**	**11.83**	**8.56**	**22.58**	**31.92**
AR2 + IV	**161.54**	**148.38**	**233.01**	**161.08**	**118.02**	**74.31**
AR2 + IV + GT*	2.90	-0.30	0.07	**6.18**	**13.14**	**20.83**
			Panel C: MSPE Ratios			
AR2 + return	0.92	0.91	0.99	0.98	1.01	1.00
AR2 + trading volume	0.99	1.02	0.98	0.95	0.97	0.95
AR2 + GT	1.00	0.99	0.96	0.97	1.00	0.91
AR2 + IV	0.70	0.71	0.63	0.71	0.79	0.85
AR2 + IV + GT*	0.99	1.01	1.01	0.99	0.99	0.93

Panel A shows the average (Ave-ENC) and Panel B shows the supremum (Sup-ENC) of recursive encompassing tests on logs of the realized volatilities. Panel C shows the ratio of the unrestricted model average mean square prediction error (MSPE) to the restricted benchmark model average MSPE. The benchmark model used for the first four rows is AR2. * indicates that AR2 + IV was used as the benchmark model to compute the numbers. Logs are used for Google Trends search volume index (GT), implied volatility (IV), realized volatility, and trading volume. All predictors have two 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 two additional variables are 1.315, 2.019, and 3.644, respectively, and the 10%, 5% and 1% critical values for the Sup-ENC tests with two additional variables are 3.122, 4.313, and 7.243, respectively.)

Table 5
Encompassing Tests and MSPE Ratios with ARMA(1,1)

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
			Panel A: Ave-ENC			
ARMA + return	**25.10**	**25.92**	**2.29**	**4.83**	1.07	0.83
ARMA + trading volume	0.99	-1.25	**4.04**	**5.34**	**8.66**	**16.61**
ARMA + GT	0.05	**4.10**	**9.93**	**5.48**	**7.48**	**20.50**
ARMA + IV	**101.98**	**92.08**	**146.84**	**108.15**	**64.03**	**39.99**
ARMA + IV + GT*	1.91	-0.95	-0.34	**4.07**	**5.30**	**13.36**
			Panel B: Sup-ENC			
ARMA + return	**48.18**	**48.48**	3.44	**8.50**	**6.96**	1.66
ARMA + trading volume	2.03	0.97	**5.31**	**7.57**	**21.20**	**33.54**
ARMA + GT	0.50	**7.94**	**13.70**	**8.27**	**24.53**	**35.16**
ARMA + IV	**157.19**	**144.54**	**216.22**	**153.24**	**111.79**	**69.08**
ARMA + IV + GT*	2.96	-0.33	0.23	**6.05**	**13.48**	**22.35**
			Panel C: MSPE Ratios			
ARMA + return	0.91	0.90	0.99	0.98	1.02	0.99
ARMA + trading volume	1.00	1.02	0.97	0.97	0.97	0.94
ARMA + GT	1.00	0.99	0.95	0.97	1.00	0.90
ARMA + IV	0.70	0.72	0.64	0.71	0.80	0.86
ARMA + IV + GT*	0.99	1.01	1.00	0.99	0.99	0.93

Panel A shows the average (Ave-ENC) and Panel B shows the supremum (Sup-ENC) of recursive encompassing tests on logs of the realized volatilities. Panel C shows the ratio of the unrestricted model average mean square prediction error (MSPE) to the restricted benchmark model average MSPE. The benchmark model used for the first four rows is ARMA(1,1). * indicates that ARMA(1,1) + IV was used as the benchmark model. Logs are used for Google Trends search volume index (GT), implied volatility (IV), realized volatility, and trading volume. All predictors have two 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 two additional variables are 1.315, 2.019, and 3.644, respectively, and the 10%, 5% and 1% critical values for the Sup-ENC tests with two additional variables are 3.122, 4.313, and 7.243, respectively.)

Table 6
Volatility Error Reduction

	S&P 500	DJIA	Euro	Canadian Dollar	Gold	Crude Oil
Model 1: AR2 + GT Model 2: AR2 + IV	(13.70%, 18.05%)	(12.21%, 16.04%)	(17.11%, 20.09%)	(11.66%, 13.57%)	(11.11%, 13.62%)	(2.87%, 3.47%)
Model 1: AR2 + IV Model 2: AR2 + IV + GT	(0.45%, 0.60%)	(-0.25%, -0.33%)	(-0.43%, -0.51%)	(0.94%, 1.10%)	(0.38%, 0.47%)	(2.55%, 3.06%)

This table shows reductions in forecast error for realized volatility computed as: Volatility error reduction$_i = 1 - \left|\frac{x_{\text{model2},i} - \hat{\sigma}_{\text{mean}}}{x_{\text{model1},i} - \hat{\sigma}_{\text{mean}}}\right|$, where $i \in$ {under, over}, Model 1 and Model 2 are alternative volatility forecasting models, x represents typical underestimate or overestimate of volatility based on the given model, and $\hat{\sigma}_{\text{mean}}$ is the average annualized realized standard deviation for the second half of the sample period. GT is the Google Trends search volume index, and IV is implied volatility. The beginning of the sample period is January 4, 2004 for S&P 500, DJIA, Euro and Canadian dollar, and October 23, 2005 for gold and crude oil. The end of the sample period is August 28, 2015 for all markets.

Table 7
Variances and Shares of Implied Volatility and Google Trends Search Volume Residuals

Market	$V(u_{IV})$	$V(u_{GT})$	$Corr(u_{IV}, u_{GT})$	$\frac{V(u^c)}{V(u_{IV})}$	$\frac{V(u^c)}{V(u_{GT})}$
S&P 500	0.017	0.020	0.161	0.179	0.145
DJIA	0.016	0.043	0.375	0.622	0.226
Euro	0.020	0.007	0.404	0.246	0.663
Canadian Dollar	0.014	0.011	0.269	0.235	0.309
Gold	0.023	0.016	0.433	0.360	0.520
Crude Oil	0.017	0.031	0.378	0.517	0.277

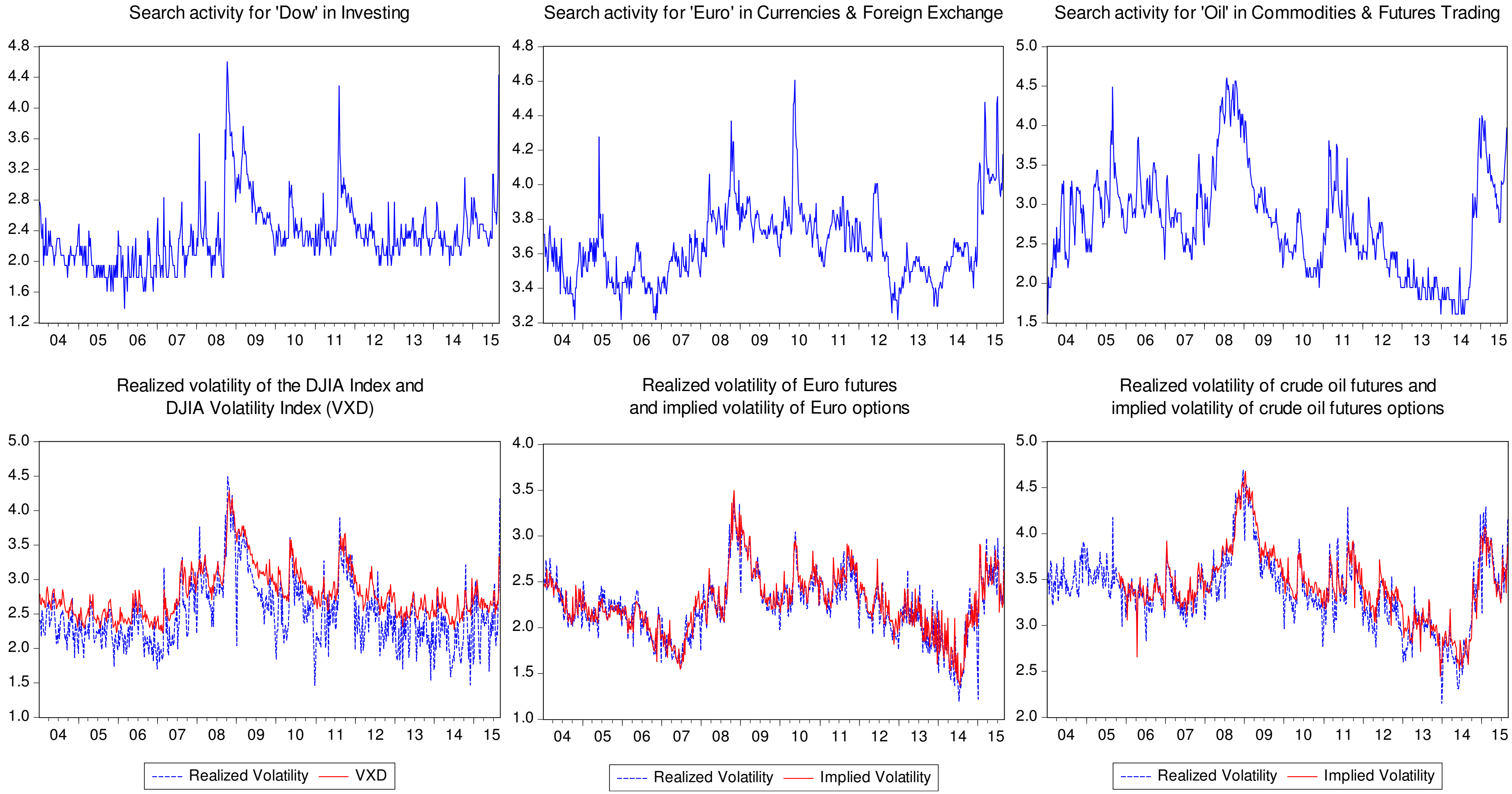
$V(u_{IV})$ shows the variances of residuals of implied volatility from the five-variable VAR with two lags described in Section 3.1. Similarly, $V(u_{GT})$ shows the variances of residuals of Google Trends search volume index from the same VAR. $Corr(u_{IV}, u_{GT})$ shows the correlations of the two residuals. The last two columns show the variance of the common component of the two residuals as a percentage of the variance of the implied volatility residuals and Google Trends search volume residuals, respectively. The beginning of the sample period is January 4, 2004 for S&P 500, DJIA, Euro, and Canadian dollar, and October 23, 2005 for gold and crude oil. The end of the sample period is August 28, 2015 for all markets.

Table 8
Fit of the Predictive Regressions Using Unobserved Components

Market	$R^2(u^c)$	$R^2(u^i_{IV})$	$R^2(u^i_{GT})$
	Panel A: Without Additional Predictors		
S&P 500	0.188	0.200	0.008
DJIA	0.232	0.176	0.005
Euro	0.088	0.107	0.008
Canadian Dollar	0.090	0.109	0.003
Gold	0.167	0.098	0.023
Crude Oil	0.109	0.048	0.023
	Panel B: With Three Additional Predictors		
S&P 500	0.717	0.727	0.702
DJIA	0.697	0.702	0.681
Euro	0.755	0.788	0.745
Canadian Dollar	0.830	0.841	0.812
Gold	0.679	0.689	0.658
Crude Oil	0.839	0.831	0.826

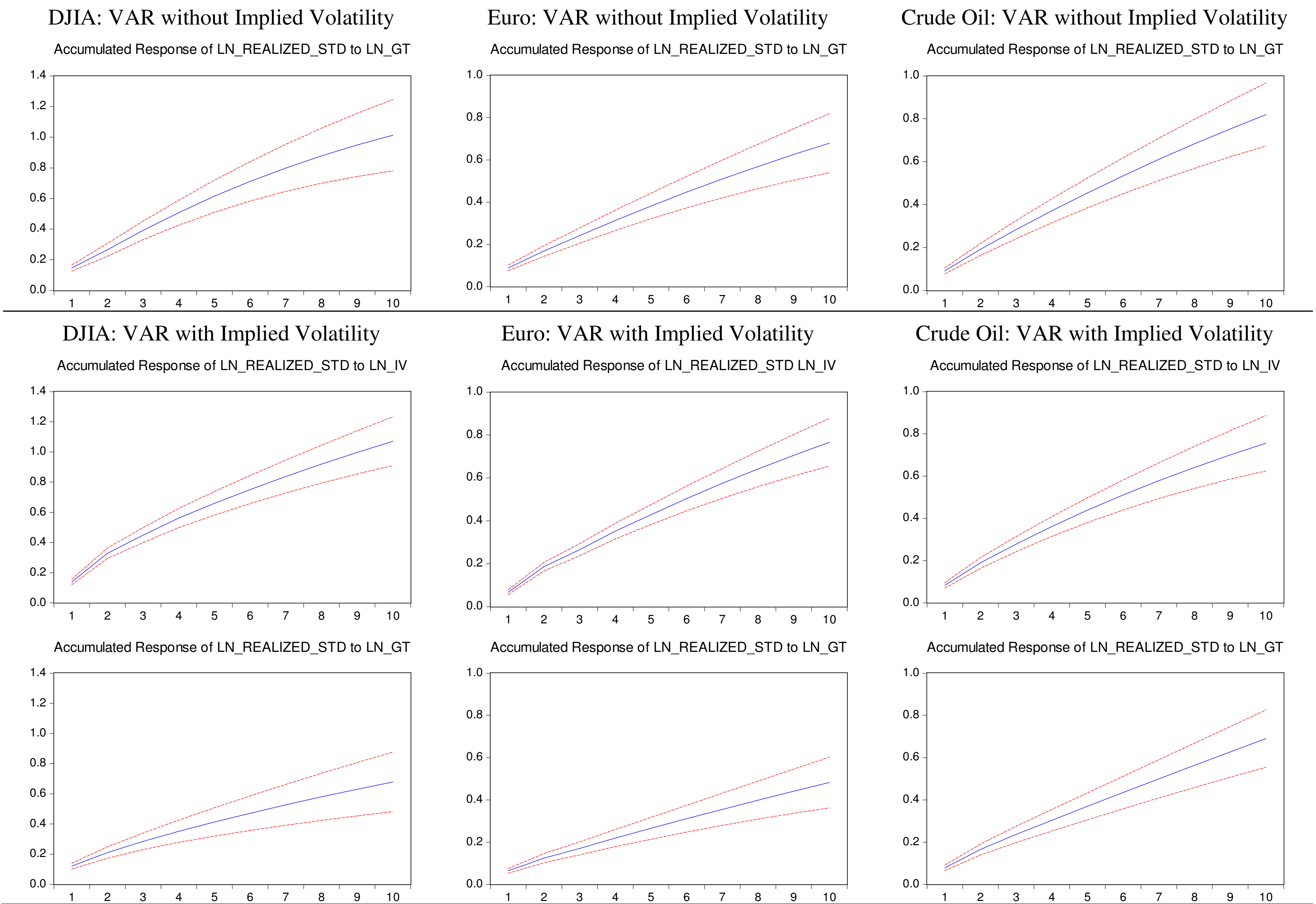
Panel A shows the fit, measured by R^2, for three predictive regressions of realized volatility using each of the three unobserved components u^c, u^i_{IV}, and u^i_{GT} defined in Section 3.3. Two lags of each component are used in these regressions. Panel B shows the fit of the three predictive regressions that additionally include two lags of realized volatility, returns, and log of trading volume. See equation (11) for the specification. The beginning of the sample period is January 4, 2004 for S&P 500, DJIA, Euro, and Canadian dollar, and October 23, 2005 for gold and crude oil. The end of the sample period is August 28, 2015 for all markets.

Figure 1
Google Search Volume, Realized Volatility, and Implied Volatility for DJIA, Euro, and Crude Oil



The top panels show the Google Trends search volume index for ‘dow’, ‘euro’ and ‘oil’ search terms in the U.S. region in the Finance category in the Investing, Currencies & Foreign Exchange, and Commodities & Futures Trading sub-categories, respectively. The bottom panels show realized and implied volatility. All variables are log-transformed. The sample period is from January 4, 2004 to August 28, 2015 for all markets. Data for implied volatility of crude oil futures options is available since October 28, 2005.

Figure 2
Impulse Responses



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