MPRA
Munich Personal RePEc Archive

The importance of being informed: forecasting market risk measures for the Russian RTS index future using online data and implied volatility over two decades

Fantazzini, Dean and Shangina, Tamara

2019

Online at https://mpra.ub.uni-muenchen.de/95992/
MPRA Paper No. 95992, posted 12 Sep 2019 17:08 UTC

The importance of being informed: forecasting market risk measures for the Russian RTS index future using online data and implied volatility over two decades

D. Fantazzini, T. Shangina[1]

Applied Econometrics, forthcoming.

This paper focuses on the forecasting of market risk measures for the Russian RTS index future, and examines whether augmenting a large class of volatility models with implied volatility and Google Trends data improves the quality of the estimated risk measures. We considered a time sample of daily data from 2006 till 2019, which includes several episodes of large-scale turbulence in the Russian future market. We found that the predictive power of several models did not increase if these two variables were added, but actually decreased. The worst results were obtained when these two variables were added jointly and during periods of high volatility, when parameters estimates became very unstable. Moreover, several models augmented with these variables did not reach numerical convergence. Our empirical evidence shows that, in the case of Russian future markets, T-GARCH models with implied volatility and student's t errors are better choices if robust market risk measures are of concern.

Ключевые слова: Forecasting, Value-at-Risk, Realized Volatility, Google Trends, Implied Volatility, GARCH, ARFIMA, HAR, Realized-GARCH.
JEL classification: C22**,** C51, C53, G17, G32.

1 Introduction

The market risk is usually defined in the financial literature as the gains and losses on the value of a position or portfolio that can take place due to the movements in market variables (asset prices, interest rates, forex rates), see the Basel Committee on Banking Supervision (2009), Hartmann (2010) and references therein for more details. The most well-known market risk measure is the Value-at-Risk (VaR), which can be defined as the maximum loss over a given time horizon that may be incurred by a position or a portfolio at a given level of confidence. The VaR is recognized by official bodies worldwide as an important market risk measurement tool, see Jorion (2007) and the Basel Committee on Banking Supervision (2013, 2016). The Value-at-Risk (VaR) has been criticized for not being sub-additive so that the risk of a portfolio can be larger than the sum of the stand-alone risks of its components, see e.g. Artzner et al. (1997) and Artzner et al. (1999). For this reason, the Expected Shortfall (ES) has been proposed as an alternative risk measure able to satisfy the property of sub-addittivity and to be a coherent risk measure, see Acerbi and Tasche (2002). The ES measures the average of the worst α losses, where α is the percentile of the returns distribution. Unfortunately, Gneiting (2011) showed that the ES does not satisfy a mathematical property called elicitability (while VaR does have it), and it cannot be backtested. However, Emmer et al. (2015) showed that the ES is elicitable conditionally on the VaR, and that it can be backtested through the approximation of several VaR levels: this idea was further

[1]**Dean Fantazzini.** — Moscow School of Economics, Moscow State University. Email: fantazzini@mse-msu.ru .
Tamara Shangina — Higher School of Economics, Moscow (Russia);

developed by Kratz et al. (2018) who proposed a multinomial test of VaR violations at multiple levels as an intuitive idea of backtesting the ES.

The goal of this paper is to examine whether augmenting a large class of volatility models with implied volatility from option prices and Google Trends data improves the quality of the estimated VaRs at multiple confidence levels for the Russian RTS index future. The RTS index is based on the 50 most liquid stocks of the Russian market, and it is an important indicator for the whole Russian market. Since the RTS index is not tradable, we considered the RTS index future for the purpose of our analysis. Google search data is an indicator of the people behavior and their attention, and it can be a driver of future volatility, see for example Campos et al. (2017) and references therein for a large discussion. Implied volatility from option prices is a standard way to obtain a forward-looking estimate of the volatility which considers investors' beliefs, see the survey by Mayhew (1995) and references therein for more details. Internet searches come mostly from the general public and small investors (Da, Engelberg and Gao (2011), Goddard et al. (2012), Vlastakis and Markellos (2012), and Vozlyublennaia (2014)), whereas implied volatility captures the expectations of institutional investors and market makers who have access to premium and insider information (Martens and Zein (2004), Busch et al. (2011), Bazhenov and Fantazzini (2019)). These two measures of investors' attention and expectations are used to augment a large class of volatility models to forecast the VaR at multiple levels for the Russian RTS index future, using daily data from 2006 till 2019. We considered four class of models: the Threshold-GARCH model by Glosten et al. (1993) and Zakoian (1994), the Heterogeneous Auto-Regressive (HAR) model by Corsi (2009), the AutoRegressive Fractional (ARFIMA) model by Andersen et al. (2003), and the Realized-GARCH model by Hansen et al. (2012). The forecasting performances of these models are compared using the forecasting diagnostics for market risk measurement, such as the tests by Kupiec (1995) and Christoffersen (1998), the multinomial test of VaR violations by Kratz et al. (2018), the asymmetric quantile loss (QL) function proposed by Gonzalez-Rivera et al. (2004), and the Model Confidence Set by Hansen et al. (2011).

The first contribution of this paper is an evaluation of the contribution of both online search queries and options-based implied volatility to the modelling of the volatility of the Russian RTS index future, and how this dependence has changed over almost two decades (from 2006 till 2019). To our knowledge, this analysis has not been done elsewhere. The second contribution is an out-of-sample forecasting exercise of the Value-at-Risk for the RTS index future at multiple confidence levels using several alternative models' specifications, with and without Google data and implied volatility. The third contribution of the paper is a robustness check to measure the accuracy of Value-at-Risk forecasts obtained with a multivariate model.

The rest of this paper is organized as follows. Section 2 briefly reviews the literature devoted to Google Trends and implied volatility, while the methods proposed for forecasting the Value-at-Risk are discussed in Section 3. The empirical results are reported in Section 4, while a robustness check is discussed in Section 5. Section 6 briefly concludes.

2 Literature review

There is an increasing body of the financial literature which examines how implied volatility (IV) from option prices and Google Trends data influence volatility modelling and risk measures.

In the case of implied volatility, past research showed that it provides better forecasts for volatility than traditional GARCH models, see Christensen and Prabhala (1998), Corredor and Santamaría (2004), Martens and Zein (2004), Busch et al. (2011) and Haugom et al. (2014a), just to name a few. However, there are some (few) cases where this is not true: for example, Agnolucci (2009) found that a Component-GARCH model performs better than IV in forecasting the volatility of crude oil futures, while Birkelund et al. (2015) examined the Nordic the power forward market and found that the IV is a biased predictor of the realized volatility. In general, the financial literature usually shows that the best results are obtained when both the IV and other market variables are included in the forecasting model, with intraday volatility information being the only variable able to successfully complement IV, see Taylor and Xu (1997), Pong et al. (2004), and Jeon and Taylor (2013).

While implied volatility forecasts the future volatility well, the results are more mixed when forecasting the future quantiles of the returns' distribution is of concern. Giot (2005) showed that the VaR forecasts based on lagged implied volatility performed similarly to VaR estimates based on GARCH models, while Jeon and Taylor (2013) reported that the implied volatility has explanatory power for the left tail of the conditional distribution of SP500 daily returns. Instead, Chong (2004) found that time series models performed better than the model based on implied volatility when estimating the VaR for exchange rates. Similarly, Christoffersen and Mazzotta (2005) found that, while implied volatility provided the most accurate volatility forecasts for all currency exchange rates and forecast horizons considered, unfortunately, it did not capture the tail behavior of the returns distribution. Barone-Adesi et al. (2019) estimated and backtested the 1%, 2.5%, and 5% WTI crude oil futures value at risk and conditional value at risk for the years 2011–2016 and for both tails of the distribution, and they showed that the option-implied risk metrics are valid alternatives to filtered-historical simulation models.

The work by Bams et al. (2017) is the largest backtesting exercise to date, comparing the Value-at-Risk forecasts from implied volatility and historical volatility models, using more than 20 years of daily data from American markets. Their large-scale backtesting analysis shows that implied volatility based Value-at-Risk tends to be outperformed by simple GARCH based Value-at-Risk, and they explain the poor performance of the former due to the volatility risk premium embedded in implied volatilities. However, even when correcting for the variance risk premium, the VaR forecasts based on the IV cannot outperform the historical volatility models. Bams et al. (2017) explain this result by showing that, while the IV is useful for forecasting the future volatility, it is not useful for forecasting a quantile of the return distribution, due to the complex dependence structure between the volatility risk premium and the extreme returns which influence the quantile forecasting power of the implied volatility.

In the case of Google search data, a vast literature discussed how the Internet search activity captures investor attention and information demand, see Ginsberg et al. (2009), Choi and Varian (2012), Da et al. (2011), Vlastakis and Markellos (2012), Vozlyublennaia (2014), Goddard et al. (2012), and Fantazzini and Toktamysova (2015), just to name a few.

Vozlyublennaia (2014) showed that Google data does have short term and -in some cases- even long term effects on returns, whereas the effects on volatility are less pronounced. However, she does not investigate the predictability of volatility by using Google data. Dimpfl and Jank (2016) were among the first to use daily Google data to forecast daily and weekly realized variances by including it as an additive component in Autoregressive (AR) and Heterogeneous Autoregressive (HAR) models, and they showed that Google searches improved both in- and out-of-sample performances. Campos and Cortazar (2017) evaluated the marginal contribution of Google trends to forecast the Crude Oil Volatility index by using HAR models and several macro-finance variables and showed that Google data has a positive relationship with the oil volatility index, even accounting for macroeconomic variables. They also found that internet search volumes add valuable information to their forecasting models and their predictions generate more economic value than models without them. Xu et al. (2017) showed that Google Trends is a significant source of volatility besides macroeconomic fundamentals and its contribution to forecasting volatility can be enhanced by combining with other macroeconomic variables. Interestingly, they found that Google data with a higher observed frequency tends to be more useful in volatility forecasting than with a lower frequency. Moreover, the higher the stock market volatility is, the more important Google data are for volatility forecasting. Seo et al. (2019) proposed a set of hybrid models based on artificial neural networks with multi-hidden layers, combined with GARCH family models and Google Trends. They reported that their hybrid models with Google data statistically outperform the GARCH models and the hybrid models without Google data when forecasting the volatility of American markets.

The studies dealing with market risk measurement and Google Trends are much more limited, instead: Hamid et al. (2015) proposed an empirical similarity approach to forecast the weekly volatility and the VaR for the Dow Jones index by using Google data. They performed an out-of-sample exercise with 260 weekly data and found that their model delivered significantly more accurate forecasts than competing models (HAR and ARFIMA models with and without Google data), while requiring less capital due to fewer over-predictions. Basistha et al. (2018) was the first work that evaluated both the role of the Google data and implied volatility for forecasting the realized volatility of American financial markets, exchanges rates and commodity markets. They found that Google search data play a minor role in predicting the realized volatility once implied volatility is included in the set of regressors. They also performed a small risk management exercise involving the one-week-ahead 1% Value-at-Risk for the SP500 and the DJIA indices and found that adding the implied volatility produces an economically meaningful improvement in market risk measurement, while this is not the case when adding the Google search volume to the forecasting models. Bazhenov and Fantazzini (2019) performed a similar analysis with four Russian stocks for the 2016-2019 time sample, and they found that models including the implied volatility improved their forecasting performances, whereas models including Google Trends data worsened their performances. Interestingly, simple HAR and ARFIMA models without additional regressors often reported the best forecasts for the daily realized volatility and the daily Value-at-Risk at the 1% probability level, thus showing that efficiency gains more than compensate any possible model misspecifications.

3 Methodology

The goal of this paper is to verify whether adding the implied volatility from option prices and Google data to a large set of volatility models improves the quality of the estimated VaRs at multiple confidence levels for the Russian RTS index future. This procedure also has the benefit to indirectly test the quality of the models' Expected Shortfall, following the aforementioned approach by Kratz et al. (2018). Before presenting these market risk measures and the associated backtesting procedures, we briefly review the measures of volatility and the forecasting models that we will use to compute these market risk measures.

3.1 Measures of volatility

3.1.1 Realized Variance

Suppose we have a stochastic process with the following form:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t)$$

where $p(t)$ is the logarithm of instantaneous price, $\mu(t)$ is of finite variation, $\sigma(t)$ is strictly positive and square integrable and $dW(t)$ is a Brownian motion. This is a continuous-time diffusive setting which rules out price jumps and assumes a frictionless market. For this diffusion process, the Integrated Variance associated with day t is defined as the integral of the instantaneous variance over the following one day integral:

$$IV_{t+1} = \int_{t}^{t+1} \sigma^2(s)\,ds$$

It is possible to show that the previous integrated variance can be approximated to an arbitrary precision using the sum of intraday squared returns. This nonparametric estimator is called *Realized Variance* and it is a consistent estimator of the integrated variance as the sampling frequency increases, see Meddahi (2002) and Andersen et al. (2001):

$$RV_{t+1} = \sum_{j=1}^{M} r_{t+j\Delta}^2$$

where $\Delta=1/M$ is the time interval of the intraday prices, M is the number of intraday returns, while $r_{t+j\Delta}$ is the intraday return. The previous estimator considers the daily realized variance, but different time horizons longer than a single day d can be computed. For example, the weekly realized variance w at time t is given by,

$$RV_t^{(w)} = \frac{1}{5}\left(RV_t^{(d)} + RV_{t-1d}^{(d)} + \ldots + RV_{t-4d}^{(d)}\right).$$

where we considered a weekly time interval of five working days.

If we allow for the presence of jumps and consider the following continuous-time jump-diffusion process, $dp(t) = \mu(t)dt + \sigma(t)dW(t) + k(t)dq(t)$, where $q(t)$ is a counting process with $dq(t) = 1$ corresponding to a jump at time t and $dq(t) = 0$ otherwise and $k(t)$ refers to the size of the corresponding jumps, then it is possible to show that the realized volatility converges to the sum of the integrated variance and the cumulative squared jumps, see Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen and Shephard (2006), Andersen et al. (2007):

$$\operatorname*{plim}_{\Delta \to 0} RV_{t+1}(\Delta) = \int_t^{t+1} \sigma^2(s)\,ds + \sum_{t<s\leq t+1} k^2(s)$$

The continuous sample path variation measured by the integrated variance can be estimated non-parametrically using the standardized Realized Bipower Variation measure,

$$BV_{t+1}(\Delta) = \mu_1^{-2} \sum_{j=2}^{1/\Delta} | r_{t+j\Delta} \left| \left| r_{t+(j-1)\Delta} \right| = \mu_1^{-2} \sum_{i=2}^{M} | r_{t,i} \right| \left| r_{t,i-1} \right| = C_{t+1}(\Delta)$$

while the jump component can be consistently estimated by

$$J_{t+1}(\Delta) = max\left[RV_{t+1}(\Delta) - BV_{t+1}(\Delta), 0\right]$$

where the non-negativity truncation on the actual empirical jump measurements was suggested by Barndorff-Nielsen and Shephard (2004b) because the difference between RV and BV can become negative in a given sample. However, Huang and Tauchen (2005) and Andersen et al. (2007) suggested to treat the small jumps as measurement errors and consider them as part of the continuous sample path variation process, whereas only abnormally large values of $RV_{t+1}(\Delta) - BV_{t+1}(\Delta)$ should be associated with the jump component. To achieve this goal, they proposed a test to identify the significant jumps and automatically guarantee that both J_{t+1} and C_{t+1} are positive. We employed this approach in our empirical analysis and we refer to Huang and Tauchen (2005) and Andersen et al. (2007) for the full description of this testing procedure.

Given that we also considered GARCH-type models with daily returns, we had to adjust the previous realized variance for the return in the overnight gap from the market close on day t to the market open on day $t+1$. We scaled up the market-open RV using the unconditional variance estimated with the daily squared returns:

$$RV_{t+1}^{24H} = \left(\frac{\sum_{t=1}^{T} r_t^2}{\sum_{t=1}^{T} RV_t^{OPEN}}\right) RV_{t+1}^{OPEN}$$

where r_t^2 are the daily squared returns computed using the close-to-close daily prices, while RV_{t+1}^{OPEN} is the realized variance computed with intraday data when the RTS future market is open. We chose this approach due to its better results in the financial empirical literature, see Hansen and Lunde (2005), Christoffersen (2012), and Ahoniemi and Lanne (2013).

3.1.2 Implied volatility

An implied volatility index estimates the market expectations for the future volatility implied by the stock index option prices. The first Russian volatility index named RTSVX (Russian Trading System Volatility Index) was introduced on 7 December 2010 and was discontinued on 12 December 2016. It was based on the volatility of the nearby and next option series for the RTS (Russian Trading System) Index futures, see the Moscow Exchange website for more details. The Moscow Exchange makes available a (partially reconstructed) time series of daily closed prices for the RTSVX index starting from January 2006. A new Russian Volatility Index (RVI) was introduced on 16 April 2014: this index measures the market expectations for volatility over a 30 day period, and it is computed using prices of nearby and next RTS Index option series. The RVI is calculated in real time during both day and evening sessions (first values 19:00 – 23:50 Moscow time and then 10:00 – 18:45 Moscow time). There are three aspects where the RVI differs from the RTSVX: it is discrete, it uses actual option prices over 15 strikes, and it calculates a 30-day volatility. The RVI formula is reported below:

$$IV = 100\sqrt{\frac{T_{365}}{T_{30}} * \left| T_1\sigma_1^2 \frac{T_2 - T_{30}}{T_2 - T_1} + T_2\sigma_2^2 \frac{T_{30} - T_1}{T_2 - T_1} \right|},$$

where T_{30} stands for 30 days expressed as a fraction of a calendar year, T_{365} for 365 days expressed as a fraction of a calendar year, T_1 is the time to expiration of the near-series options expressed as a fraction of a calendar year, T_2 is the time to expiration of the far-series options expressed as a fraction of a calendar year, σ_1^2 is the variance of the near-series options and σ_2^2 is the variance of the next-series of options, see http://fs.moex.com/files/6757 for the full description of the RVI methodology. A detailed comparison of the old and new Russian volatility indexes was performed by Caporale et al. (2019) and they found that the differences were minor. For this reason, we built a composite volatility index ranging from January 2006 till April 2019, which allowed us to cover almost two decades[2].

3.2 Volatility Models

3.2.1 TGARCH model

The Generalized Auto-Regressive Conditional Heteroscedasticity (GARCH) models represent an important benchmark in empirical finance, see Hansen and Lunde (2005b) for a large-scale backtesting comparison

[2] We also tried to compute market risk measures using only one of the two volatility indexes, but the results did not change qualitatively, so that we stuck to our composite index. Note that both indexes are expressed in percentage form.

involving more than 330 volatility models. A GARCH(p,q) model for the conditional variance σ_t^2 at time t can be defined as follows:

$$\sigma_{t+1}^2 = \alpha_0 + \sum_{i=0}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=0}^{q} \beta_j \sigma_{t-j}^2$$

The Threshold-GARCH (TGARCH) model proposed by Glosten et al. (1993) and Zakoïan (1994) to take the leverage effect into account can be specified as follows:

$$\sigma_{t+1}^2 = \alpha_0 + \sum_{i=0}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=0}^{q} \beta_j \sigma_{t-j}^2 + \sum_{k=0}^{r} \gamma_k \varepsilon_{t-k}^2 I(\varepsilon_{t-k} < 0)$$

where $I = 1$ if $\varepsilon_{t-k} < 0$. A simple TGARCH (1,1) model with standardized errors following a Student's t-distribution was employed in this work. Moreover, we also considered a TGARCH(1,1) specification including the (implied) volatility index and Google Trends as additional regressors,

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2 + \gamma_1 \varepsilon_t^2 I\left(\varepsilon_t < 0\right) + \delta IV_t + \psi GT_t$$

similarly to the specification proposed for Russian stocks by Bazhenov and Fantazzini (2019).

3.2.2 HAR model

Corsi (2009) proposed a model with a hierarchical process, where the future volatility depends on the past volatility over different lengths of periods. The HAR model has the following form,

$$RV_{t+1} = \beta_0 + \beta_D RV_t + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \epsilon_{t+1},$$

where D,W and M stand for daily, weekly and monthly values of the realized volatility, respectively. Since we want to verify whether adding the implied volatility and Google data improve the estimation of the market risk measures for the RTS index future, we also included these additional regressors using the HAR specification by Bazhenov and Fantazzini (2019):

$$RV_{t+1} = \beta_0 + \beta_D RV_t + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \delta IV_t + \psi GT_t + \epsilon_{t+1} \qquad (1)$$

Andersen et al. (2007) extended the HAR-RV model by decomposing the realized variances into the continuous sample path variability and the jump variation, and they proposed the so-called *HAR-CJ* model which is defined as follows:

$$RV_{t+1} = \beta_0 + \beta_{CD} C_t + \beta_{CW} C_{t-5,t} + \beta_{CM} C_{t-22,t} + \beta_{JD} J_t + \beta_{JW} J_{t-5,t} + \beta_{JM} J_{t-22,t} + \epsilon_{t+1}$$

If we add the IV and Google data, the previous equation will transform into:

$$RV_{t+1} = \beta_0 + \beta_{CD} C_t + \beta_{CW} C_{t-5,t} + \beta_{CM} C_{t-22,t} + \beta_{JD} J_t + \beta_{JW} J_{t-5,t} + \beta_{JM} J_{t-22,t} + \delta IV_t + \psi GT_t + \epsilon_{t+1}$$

Note that a large-scale forecast comparison of volatility models for the Russian stock market was performed by Aganin (2017), and he found that the HAR model showed a statistically significant superior performance compared to all other competing models.

3.2.3 ARFIMA model

The Auto-Regressive Fractional Integrated Moving Average -ARFIMA(p,d,q)- model to forecast the realized volatility was proposed by Andersen et al. (2003), and it can be expressed as follows

$$\Phi(L)(1-L)^d(RV_{t+1}-\mu)=\Theta(L)\varepsilon_{t+1}$$

where L is the lag operator, $\Phi(L)=1-\varphi_1 L-\ldots-\varphi_p L^p$, $\Theta(L)=1+\theta_1 L+\ldots+\theta_q L^q$ and $(1-L)^d$ is the fractional differencing operator defined by:

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

where $\Gamma(\bullet)$ is the gamma function. Similarly to the HAR and GARCH models, we also added the implied volatility and Google Trends as external regressors:

$$\Phi(L)(1-L)^d(RV_{t+1}-\mu)=\delta IV_t+\psi GT_t+\Theta(L)\varepsilon_{t+1}$$

Hyndman and Khandakar (2008) proposed an algorithm for the automatic selection of the optimal ARFIMA model, which is implemented in the R packages *forecast* and *rugarch* and which was used in this work.

3.2.4 Realized-GARCH model

The realized GARCH by Hansen et al. (2012) jointly models the returns and the realized measures of volatility. More specifically, this model connects the realized volatility to the latent volatility via a measurement equation, which also accommodates an asymmetric reaction to shocks. The realized GARCH model with a log-linear specification can be written as follows:

$$r_t=\mu+\sqrt{\sigma_t^2}\cdot z_t,\quad z_t\sim i.i.d.(0,1)$$

$$\log\sigma_t^2=\omega+\sum_{i=1}^{q}\gamma_i\log RV_{t-i}+\sum_{i=1}^{p}\beta_i\log\sigma_{t-i}^2$$

$$\log RV_t=\xi+\psi\log\sigma_t^2+\tau_1 z_t+\tau_2\left(z_t^2-1\right)+u_t,\quad u_t\sim i.i.d.\left(0,\sigma_u^2\right)$$

where the three equations represent the return equation, the volatility equation, and the measurement equation, respectively. The latter equation models the contemporaneous dependence between the latent volatility and the realized measure, while the terms $\tau_1 z_t+\tau_2\left(z_t^2-1\right)$ accommodate potential leverage-type effects. Similarly to previous models, we also added the implied volatility and Google Trends as external regressors. The log-linear specification was used in this work due to its better numerical and statistical properties compared to the linear specification, see Hansen et al. (2012).

3.3 Market Risk Measures

The Value-at-Risk is the maximum market loss of a financial position over a time horizon h with at a pre-defined confidence level (1-α), or alternatively, the minimum loss of the α worst losses over the time horizon h. The VaR is a widely used measure of market risk in the financial sector, and we refer to McNeil et al. (2015) and Fantazzini (2019) for a large discussion at the textbook level. In this work, we considered h = 1.

In the case of GARCH and Realized-GARCH models with student's t errors, the 1-day ahead VaR can be computed as follows,

$$VaR_{t+1,\alpha} = \hat{\mu}_{t+1} + t^{-1}_{\alpha,\upsilon} \cdot \sqrt{(\upsilon-2)/\upsilon} \cdot \sqrt{\hat{\sigma}^2_{t+1}}$$

where $\hat{\mu}_{t+1}$ is the 1-day-ahead forecast of the conditional mean, $\hat{\sigma}^2_{t+1}$ is the 1-day-ahead forecast of the conditional variance, while $t^{-1}_{\alpha,\upsilon}$ is the inverse function of the Student's t distribution with υ degrees of freedom at the probability level α. The term $\sqrt{(\upsilon-2)/\upsilon} \cdot \hat{\sigma}_{t+1}$ is also known as the scale parameter of the Student's t distribution.

In the case of HAR and ARFIMA models, the 1-day ahead VaR can be computed as follows:

$$VaR_{t+1,\alpha} = \Phi^{-1}_{\alpha} \sqrt{RV_{t+1}}$$

where Φ^{-1}_{α} is the inverse function of a standard normal distribution function at the probability level α, while RV_{t+1} is the 1-day-ahead forecast for the realized volatility.

The Expected Shortfall (ES) measures the average of the worst α losses, where α is a percentile of the returns' distribution, and it is computed as follows:

$$ES_{\alpha} = \frac{1}{\alpha}\int_0^{\alpha} F_z^{-1}(X)\,dz = \frac{1}{\alpha}\int_0^{\alpha} VaR_z(X)\,dz$$

where F^{-1} is the inverse function of the returns' distribution, that is the Value-at-Risk.

The ES attracted a lot of attention when in October 2013 the Basel Committee on Banking Supervision issued the revision of the market risk framework, where the 99% VaR (that is, the α=1% probability level VaR) was substituted with the 97.5% ES (see Basel Committee on Banking Supervision (2013), pg. 18). The main drawback of the ES is that it lacks a mathematical property called elicitability, while VaR does have it (see Gneiting (2011)): if a risk measure is elicitable, then it can be used within a scoring function to be minimized for comparative tests on models, thus allowing for the ranking of the risk models' performance. However, Emmer et al. (2015) and Kratz, et al. (2018) showed that the ES is elicitable conditionally on the VaR, and that it can be back-tested through the approximation of several VaR levels. More specifically, Emmer et al. (2015) showed that,

$$ES_{\alpha} \approx \frac{1}{4}\left[q(\alpha)+q(0.75\alpha+0.25)+q(0.5\alpha+0.5)+q(0.25\alpha+0.75)\right]$$

where $q(\gamma)=VaR_{\gamma}$. For example, if α=2.5% as proposed by the Basel Committee, then

$$ES_{2.5\%} \approx \frac{1}{4}\left[VaR_{2.5\%}+VaR_{2.125\%}+VaR_{1.75\%}+VaR_{1.375\%}\right]$$

In this case, a more convenient approximation was proposed by Wimmerstedt (2015):

$$ES_{2.5\%} \approx \frac{1}{5}\left[VaR_{2.5\%}+VaR_{2.0\%}+VaR_{1.5\%}+VaR_{1.0\%}+VaR_{0.5\%}\right]$$

Given these recent results, we decided to take a neutral stance towards the current (hot) debate between VaR and ES proponents: we computed the VaR_{α} at five probability levels (2.5%, 2%, 1.5%, 1%, 0.5%), so that we could also provide an approximate backtesting of the $ES_{2.5\%}$ which will be included in the future Basel 3 agreement (scheduled to be introduced on 1 January 2022).

3.4 Backtesting Methods

The forecasting performance of different VaR models can be checked by comparing the forecasted values of the VaR with the actual returns for each day. The first step is to count the number of violations T_1 when the ex-ante forecasted VaR are smaller than the actual losses, with $T=T_1+T_0$, while T_0 is the the number of no VaR violations. A "perfect VaR model" would show the fraction of actual violations $\hat{\pi}=T_1/T$ equal to α%. The null hypothesis $H_0:\hat{\pi}=\alpha$ can be tested using the ***unconditional coverage test*** by Kupiec (1995). He showed that the test statistic for this null hypothesis is given by,

$$LR_{uc}=-2\ln\left[(1-\alpha)^{T_0}\alpha^{T_1} / \{(1-T_1/T)^{T_0}(T_1/T)^{T_1}\}\right]\overset{H_0}{\sim}\chi_1^2$$

If we want to test the joint null hypothesis that the average number of VaR violations is correct and the violations are independent, then we can resort to the ***conditional coverage test*** by Christoffersen (1998). The main advantage of this test is that it can reject a model that forecasts either too many or too few clustered violations, while its main disadvantage is the need of at least several hundred observations to be accurate. The test statistic is reported below,

$$LR_{cc}=-2\ln\left[(1-\alpha)^{T_0}\alpha^{T_1}\right]+2\ln\left[(1-\pi_{01})^{T00}\pi_{01}^{T01}(1-\pi_{11})^{T10}\pi_{11}^{T11}\right]\overset{H_0}{\sim}\chi_2^2$$

where T_{ij} is the number of observations with value *i* followed by *j* for *i,j* = 0,1 and $\pi_{ij}=T_{ij}/\sum_j T_{ij}$ are the corresponding probabilities.

Financial regulators are concerned not only with the number of VaR violations, but also with their magnitude. For this reason, we also computed the ***asymmetric quantile loss*** (***QL***) function proposed by Gonzalez-Rivera et al. (2004),

$$QL_{t+1,\alpha} = \left(\alpha - I_{t+1}\left(\alpha\right)\right)\left(r_{t+1} - VaR_{t+1,\alpha}\right) \ ,$$

where $I_{t+1}\left(\alpha\right)=1$ if $r_{t+1} < VaR_{t+1,\alpha}$ and zero otherwise. This loss function penalizes more heavily the realized losses below the α-th quantile level, so that it can be useful to compare the costs of different admissible choices. The quantile loss function was subsequently used together with the ***Model Confidence Set*** (***MCS***) by Hansen, Lunde, and Nason (2011) to select the best VaR forecasting models at a specified confidence level. Given the difference between the QLs of models i and j at time t (that is $d_{i,j,t} = QL_{i,t} - QL_{j,t}$), the MCS approach is used to test the following hypothesis of equal predictive ability, $H_{0,M}$: $\mathrm{E}(d_{i,j,t}) = 0$, *for all* $i,j \in M$, where M is the set of forecasting models. The first step is to compute the following t-statistics,

$$t_{ij} = \frac{\bar{d}_{ij}}{var\left(\bar{d}_{ij}\right)} \quad for\, i, j \in M$$

where $\bar{d}_{ij} = T^{-1}\sum_{t=1}^{T} d_{ij,t}$, and $var\left(\bar{d}_{ij}\right)$ is an estimate of $var\left(\bar{d}_{ij}\right)$. Then, the following test statistic is computed: $TR,M = \ max_{i \in M} \left|t_{ij}\right|$. This statistic has a non-standard distribution, so the distribution under the null hypothesis is computed using bootstrap methods with 5000 replications and a minimum block length equals to 5. If the null hypothesis is rejected, one model is eliminated from the analysis and the testing procedure starts from the beginning.

The ***multinomial VaR test*** by Kratz et al. (2018) implicitly backtests the Expected Shortfall using the previous idea by Emmer et al. (2015) to approximate the ES with several VaR levels. More specifically, Kratz et al. (2018) considered several VaR probability levels $\alpha_1,\ldots,\alpha_N$ defined by $\alpha_j = \alpha + \left[\left(j-1\right)/N\right]\left(1-\alpha\right),\ j=1,\ldots,N$, for some starting level α. If $I_{t,j} = 1_{(L_t > VaR_{\alpha_j,t})}$ is the usual indicator function for a VaR violation at the level α_j and $X_t = \sum_{j=1}^{N} I_{t,j}$, then the sequence $(X_t)_{t=1,\ldots,T}$ counts the number of VaR violations at the level α_j. If we denote with $MN\left(T,\left(p_0,\ldots,p_N\right)\right)$ the multinomial distribution with T trials, each of which may result in one of $N+1$ outcomes $\left\{0,1,\ldots,N\right\}$ according to probabilities $p_0,\ldots,p_N$ that sum to one, while the observed cell counts are denoted by $O_j = \sum_{t=1}^{T} I_{(X_t=j)}\ \ j=0,1,\ldots N$, then, under the assumptions of unconditional coverage and independence as in Christoffersen (1998), it is possible to show that the random vector

$\left(O_0,\ldots,O_N\right)$ will follow the multinomial distribution $\left(O_0,\ldots,O_N\right)\sim MN\left(T,\left(\alpha_1-\alpha_0,\ldots,\alpha_{N+1}-\alpha_N\right)\right)$. Given an estimated multinomial distribution represented by $MN(T,\left(\theta_1-\theta_0,\ldots,\theta_{N+1}-\theta_N\right)$ where θ_j ($j=1,\ldots,N$) are the distribution parameters estimated with the available data sample, Kratz et al. (2018) consider the following null and alternative hypotheses:

$$H_0:\theta_j=\alpha_j,\quad \text{for} \quad j=1,\ldots,N$$

$$H_1:\theta_j\neq\alpha_j,\quad \text{for at least one } j\in\{1,\ldots,N\}$$

The null hypothesis can be tested with several test statistics, and we refer to Cai and Krishnamoorthy (2006) for a large simulations study to verify the exact size and power properties of five possible tests (three of them were later used by Kratz et al. (2018)). We employed the *exact method* in our empirical analysis: this is the fifth test statistic reviewed by Cai and Krishnamoorthy (2006), and it computes the probability of a given outcome under the null hypothesis using the multinomial probability distribution itself:

$$P\left(O_0,O_1,\ldots,O_N\right)=\frac{T!}{O_0!O_1!\ldots O_N!}(\alpha_1-\alpha_0)^{O_0}(\alpha_2-\alpha_1)^{O_1}\ldots(\alpha_{N+1}-\alpha_N)^{O_N}$$

Cai and Krishnamoorthy (2006) found that the exact method performs very well, but it can be time-consuming if the number of cells N and the sample size T are large. In this latter case, simulation methods need be used[3]. A large discussion at the textbook level of all these backtesting methods and many others can be found in Fantazzini (2019) - chapter 11.

4 Empirical Analysis

4.1 Data

Intraday data sampled every 5 minutes for the continuous RTS index future were downloaded from the website *finam.ru*, covering the period from January 2006 till April 2019. We then used the 5-minutes squared log-returns to calculate the daily, weekly and monthly realized variance measures, as previously discussed in Section 3.1.1. We remark that Liu et al. (2015) performed a large-scale forecasting analysis involving more than 400 estimators of realized measures and they found that it is difficult to significantly outperform the 5-minute RV estimator. For this reason, we employed this estimator in this work.

The daily historical data for the implied volatility of the RTS index were downloaded from the Moscow exchange (*moex.com*): as we discussed in section 3.1.2, our volatility index is a composite index which consists of the RTSVX index for the period from January 2006 till November 2016, and the RVI index for the period from December 2016 till April 2019. The volatility index was rescaled to make it a daily volatility index, comparable with the other variables used in our empirical work.

[3] Several tests reviewed by Cai and Krishnamoorthy (2006) are implemented in the R package XNomial available at cran.r-project.org/web/packages/XNomial .

Google Trends tracks the number of search queries for a topic or a keyword over a specific period and a specific region, and creates time-series reporting the relative popularity of the searched queries. More specifically, the amount of searches is divided by the total amount of searches for the same period and region, and the resulting time series is divided by its highest value and multiplied by 100. We remark that Google Trends creates a new time series for every period because its algorithm takes the highest value over the chosen period and normalizes all others to this peak point. Even though Google is not the main search engine in Russia (Yandex is, with a market share close to 56% in 2018 - all platforms), its market share is still very significant (over 40%). Moreover, Yandex search data are available only for the last year (in case of monthly data) or for the last 2 years (in case of weekly data), so that a reliable statistical analysis with these data is not possible. We used Google Trends data for the query "RTS index", both in English and in Russian, and we computed the average of these two series. All search volumes were downloaded from the Google Trends website using the R package "*gtrendsR*". Google trends data are available since 2004, but if a multi-year sample is requested, only monthly data are obtained. To remedy this problem, we downloaded daily data for each month separately, and then we concatenated them in a single series by multiplying the separate daily data with the corresponding monthly data for the whole period.

The daily returns for the RTS index future, the implied volatility index (rescaled to show the daily implied variance), the Google Trends data and the daily realized variance are reported in Figure 1.

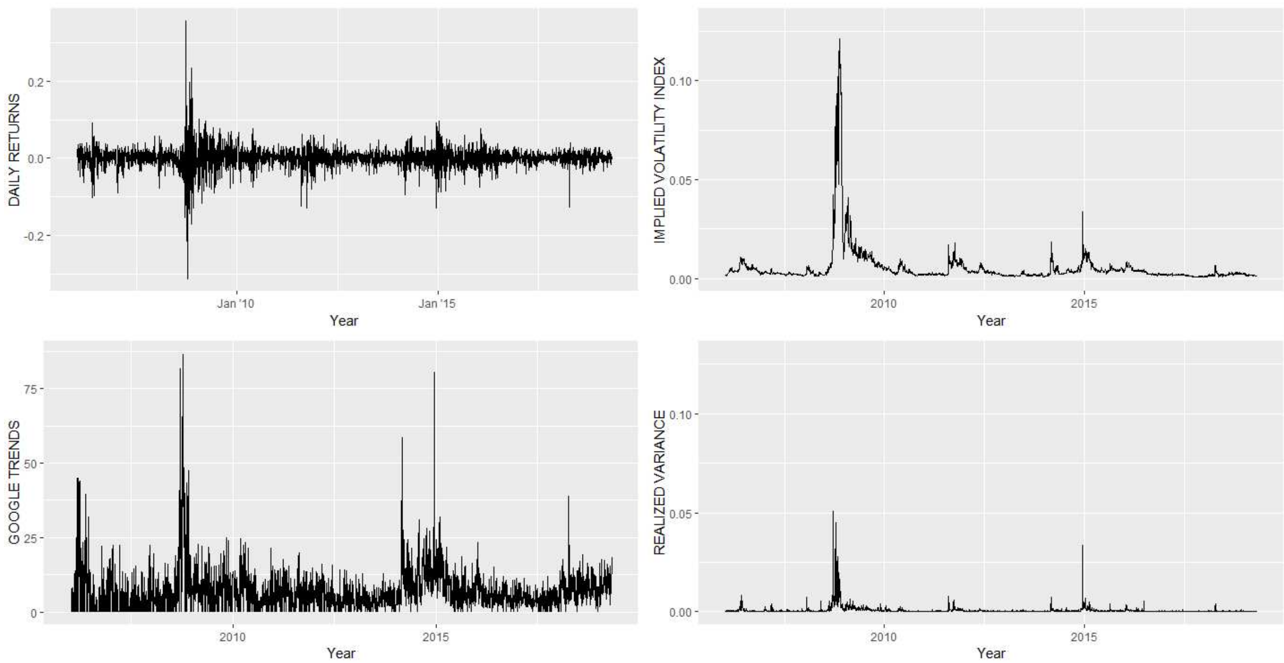

Figure 1: The daily returns, the implied volatility index, the Google Trends data and the daily realized variance for the RTS index future.

4.2 In-sample analysis

Our time sample covers almost 15 years of daily data which includes several episodes of high volatility in the Russian financial market, like the global financial crisis in 2008-2009 and several rounds of sanctions since 2014: for example, Aganin and Peresetsky (2018) found that sanctions initially increased the volatility of the ruble exchange rate, but their impact has then decreased with time. For these reasons, we do not report the models' estimates for the full sample because they would be misleading, being strongly impacted by structural breaks of different nature. Instead, we prefer to show the recursive estimates of the coefficients for the implied volatility and Google Trends in the HAR model of eq. (1), which better convey the changing impacts of these regressors on the realized volatility of the RTS index future.

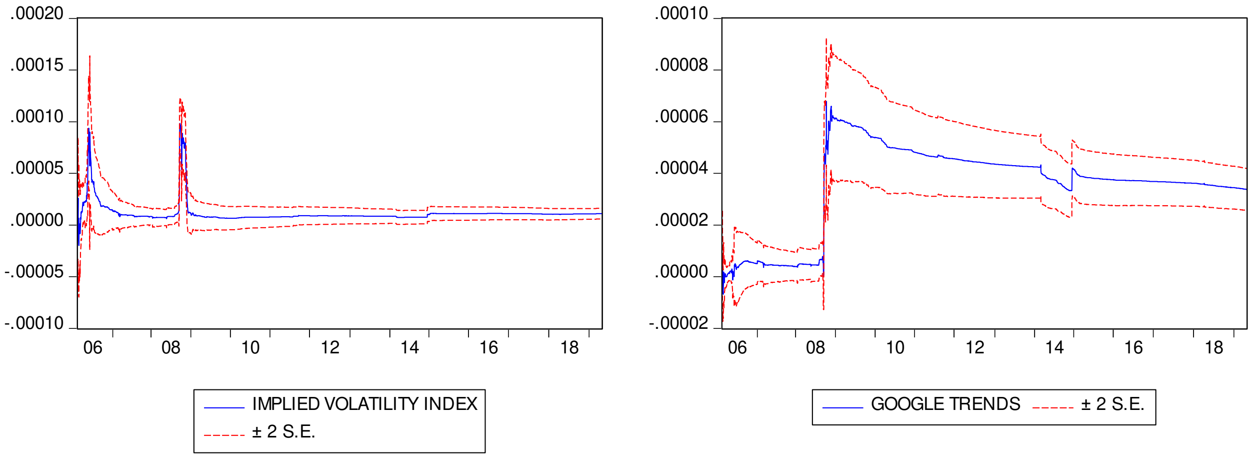

Figure 2: Recursive estimates of the coefficients for the implied volatility and Google Trends in the HAR model of eq. (1).

Figure 2 clearly shows the strong impact of the global financial crisis in 2008-2009, whereas the impact of sanctions since 2014 appear to be minor. Moreover, the (positive) effect of Google search queries seems to have decreased with time.

There is a vast literature dealing with multiple structural breaks in linear regression models, see Zeileis et al. (2002), Zeileis (2005) and Perron (2006) for extensive surveys. Among the several approaches proposed, we decided to employ the methodology based on information criteria proposed by Yao (1988), Liu et al. (1997), Bai and Perron (2003a) and Zeileis et al. (2010), which finds the optimal number of breakpoints by optimizing the Bayesian Information Criterion (BIC) and the modified BIC by Liu et al. (1997), known as the LWZ criterion. This approach has shown to be robust with different model setups and computationally tractable even with large datasets.

The multiple breakpoint test for the HAR model of eq. (1) allowing for a maximum of 5 breaks and a dataset trimming of 15% is reported in Table 1. It employs heteroscedasticity and autocorrelation consistent (HAC) covariances using a Quadratic-Spectral kernel with a Newey-West bandwidth, see Newey and West (1994), Zeileis (2006) and references therein for more details. The estimates of the model coefficients with breaks are reported in Table 2.

Table 1: Multiple breakpoint test for the HAR model of eq. (1).

Breaks	# of Coefs.	Sum of Sq. Resids.	Log-L	Schwarz* Criterion	LWZ* Criterion
0	6	0.006125	17081.35	-13.18175	-13.15066
1	13	0.005198	17351.93	**-13.32864**	**-13.26128**
2	20	0.005136	17371.72	-13.32345	-13.21981
3	27	0.005099	17383.45	-13.31337	-13.17344
4	34	0.005069	17393.26	-13.30213	-13.12591
5	41	0.005060	17396.41	-13.28684	-13.07433

Estimated break dates:
1: 9/23/2008
2: 9/23/2008, 9/17/2010
3: 9/23/2008, 7/18/2014, 7/19/2016
4: 9/23/2008, 9/20/2010, 7/18/2014, 7/19/2016
5: 9/23/2008, 9/17/2010, 9/21/2012, 11/25/2014, 11/21/2016

* The minimum information criterion values are displayed in bold font and with shading.

Table 2: Model estimates for the HAR model of eq. (1) with breaks and HAC standard errors.

	2/09/2006 - 9/22/2008		**9/23/2008 - 4/29/2019**	
Constant	-0.00154	***	-0.00016	
	(0.00042)		(0.00014)	
Realized Volatility (daily)	-0.77159	**	0.49714	***
	(0.27715)		(0.01829)	
Realized Volatility (Weekly)	2.62269	***	-0.06182	
	(0.68857)		(0.04861)	
Realized Volatility (Monthly)	-1.15187	*	0.38608	***
	(0.52013)		(0.07994)	
RVI index	0.00005	***	0.00000	
	(0.00002)		(0.00000)	
Google Trends	0.00004		0.00002	**
	(0.00002)		(0.00001)	

Note: * — $p < 0.05$, ** — $p < 0.01$, *** — $p < 0.001$

Table 1 shows that both the Schwarz and the LWZ information criteria select 1 break, coinciding with the beginning of the global financial crisis. Interestingly, the selected date is just 1 week after the bankruptcy of Lehman Brothers. Table 2 shows that both the sign and the size of the coefficients change significantly between the two time samples, particularly for the volatility components. Instead, the coefficients for the lagged implied volatility and Google Trends remain positive in both samples, but the RVI is statistically significant only in the first sample up to September 2008, whereas Google search queries are statistically significant only in the second sample (which makes sense given that Google was not very used in Russia during the first period). We also computed other tests for detecting breaks, like the sequential tests proposed by Bai (1997) and Bai and Perron (1998), and the global maximizer tests by Bai and Perron (1998, 2003a, 2003b): in these cases, the number of significant breaks was higher -mostly 4 breaks-, identified around September 2008, September

2010, December 2014 and December 2016, which we can be loosely interpreted as the beginning and the end of the global financial crisis in Russia (2008-2010) and the beginning and the end of the crisis related to sanctions and the oil price collapse (2014-2016)[4]. Finally, we remark that the variability of the parameters for the other models (GARCH and ARFIMA models) was even higher, which should not be a surprise, given the greater computational complexity of these models. However, we do not report them for the sake of interest and space, and we prefer to focus on VaR forecasting which is the main goal of this work.

4.3 Value-at-Risk forecasts

The previous empirical evidence of 1 or more breaks suggested to us to use a rolling window of 500 observations to estimate the volatility models and to compute the forecasted VaR. This time window should be a good compromise, given the numerical properties of GARCH models discussed by Hwang and Valls Pereira (2006) and Bianchi et al. (2011), together with the simulation evidence reported by Pesaran and Timmermann (2007), who showed that in a regression with multiple breaks the optimal window for estimation includes all of the observations after the last break, plus a limited number of observations before this break. Moreover, we considered the results for the full out-of-sample validation period (2008-2019) and for a rolling out-of-sample of 250 days (as requested by the Basel agreements), to examine how several structural breaks impacted the backtesting procedure. We employed the following models:

Table 3: Model specifications used in the backtesting analysis.

Model	**NO external regressors**	**IV**	**GT**	**IV+GT**	***Total***	
TGARCH	✓	✓	✓	✓	4	
HAR	✓	✓	✓	✓	4	
HARCJ	✓	✓	✓	✓	4	
ARFIMA	✓	✓	✓	✓	4	
RG	✓	✓	✓	✓	4	
HAR LOG	✓	✓	✓	✓	4	
HARCJ LOG	✓	✓	✓	✓	4	**TOTAL**
ARFIMA LOG	✓	✓	✓	✓	4	32

4.3.1 Full out-of-sample validation

The p-values of the Kupiec and Christoffersen's tests values and the number of violations in % are reported in Table 4, while the models included in the Model Confidence Set (MCS) at the 10% confidence level and their associated asymmetric quantile loss are reported in Table 5. The p-values of the Multinomial VaR test by Kratz et al. (2018) with probability levels α_1=0.5%, α_2=1%, α_3=1.5%, α_4=2% and α_5=2.5% are reported in Table 6. Only models who reached numerical convergence over the full out-of-sample are reported.

[4] These results are not reported for sake of space and are available from the authors upon request.

Table 4: Kupiec tests p-values, Christoffersen's tests p-values and number of violations in %. P-values smaller than 0.05 are in bold font.

	VaR with α = 0.5%			*VaR with α = 1%*			*VaR with α = 1.5%*			*VaR with α = 2%*			*VaR with α = 2.5%*		
Model	*Kupiec*	*Christ.*	*Violations*	*Kupiec*	*Christ.*	*Violations*	*Kupiec*	*Christ.*	*Violations*	*Kupiec*	*Christ.*	*Violations*	*Kupiec*	*Christ.*	*Violations*
TGARCH	0.00	0.01	0.96	**0.00**	**0.00**	1.82	**0.00**	**0.00**	2.64	**0.00**	**0.00**	3.15	**0.00**	**0.00**	3.72
TGARCH IV	0.13	0.28	0.71	**0.03**	0.06	1.43	**0.04**	0.11	2.00	**0.00**	**0.02**	2.79	**0.04**	0.10	3.15
TGARCH GT	0.13	0.28	0.71	**0.02**	**0.04**	1.47	**0.01**	**0.04**	2.11	**0.00**	**0.00**	2.93	**0.01**	**0.03**	3.29
TGARCH IVGT	0.13	0.28	0.71	**0.00**	**0.01**	1.57	**0.03**	0.08	2.04	**0.01**	0.05	2.68	**0.04**	0.11	3.15
HAR	**0.00**	**0.00**	1.04	**0.01**	**0.02**	1.54	**0.00**	**0.00**	2.22	**0.01**	**0.02**	2.75	0.06	0.08	3.07
HAR IV	**0.00**	**0.00**	4.75	**0.00**	**0.00**	5.58	**0.00**	**0.00**	6.29	**0.00**	**0.00**	6.79	**0.00**	**0.00**	7.29
HAR GT	**0.00**	**0.00**	2.14	**0.00**	**0.00**	2.75	**0.00**	**0.00**	3.25	**0.00**	**0.00**	3.65	**0.00**	**0.00**	4.25
HAR IVGT	**0.00**	**0.00**	5.97	**0.00**	**0.00**	6.75	**0.00**	**0.00**	7.15	**0.00**	**0.00**	7.68	**0.00**	**0.00**	8.04
HARCJ	**0.00**	**0.00**	1.22	**0.00**	**0.00**	1.64	**0.01**	**0.01**	2.14	**0.04**	**0.04**	2.57	0.12	0.20	2.97
HARCJ IV	**0.00**	**0.00**	4.93	**0.00**	**0.00**	5.58	**0.00**	**0.00**	6.11	**0.00**	**0.00**	6.72	**0.00**	**0.00**	7.33
HARCJ GT	**0.00**	**0.00**	1.97	**0.00**	**0.00**	2.43	**0.00**	**0.00**	3.11	**0.00**	**0.00**	3.47	**0.00**	**0.00**	3.93
HARCJ IVGT	**0.00**	**0.00**	5.58	**0.00**	**0.00**	6.33	**0.00**	**0.00**	6.90	**0.00**	**0.00**	7.40	**0.00**	**0.00**	7.97
ARFIMA	**0.00**	**0.00**	1.14	**0.01**	**0.02**	1.54	**0.01**	**0.01**	2.18	**0.01**	**0.01**	2.68	0.05	0.07	3.11
ARFIMA GT	**0.00**	**0.00**	1.86	**0.00**	**0.00**	2.61	**0.00**	**0.00**	3.25	**0.00**	**0.00**	3.75	**0.00**	**0.00**	4.32
RG	**0.00**	**0.00**	1.11	**0.00**	**0.00**	1.82	**0.00**	**0.00**	2.25	**0.01**	**0.03**	2.72	0.06	0.17	3.07
RG IV	**0.00**	**0.00**	1.18	**0.00**	**0.00**	1.75	**0.00**	**0.00**	2.36	**0.00**	**0.02**	2.79	**0.01**	**0.03**	3.29
RG GT	**0.00**	**0.00**	1.11	**0.00**	**0.00**	1.79	**0.00**	**0.00**	2.25	**0.01**	**0.04**	2.68	0.05	0.14	3.11
RG IVGT	**0.00**	**0.00**	1.18	**0.00**	**0.00**	1.86	**0.00**	**0.00**	2.32	**0.01**	**0.04**	2.72	**0.01**	**0.04**	3.25
HAR LOG	**0.00**	**0.00**	1.82	**0.00**	**0.00**	2.61	**0.00**	**0.00**	3.18	**0.00**	**0.00**	3.47	**0.00**	**0.00**	3.93
HAR IV LOG	**0.00**	**0.00**	1.68	**0.00**	**0.00**	2.43	**0.00**	**0.00**	2.93	**0.00**	**0.00**	3.54	**0.00**	**0.00**	4.07
HAR GT LOG	**0.00**	**0.00**	1.72	**0.00**	**0.00**	2.57	**0.00**	**0.00**	3.15	**0.00**	**0.00**	3.57	**0.00**	**0.00**	4.00
HAR IVGT LOG	**0.00**	**0.00**	1.75	**0.00**	**0.00**	2.43	**0.00**	**0.00**	2.93	**0.00**	**0.00**	3.57	**0.00**	**0.00**	4.07
HARCJ LOG	**0.00**	**0.00**	1.79	**0.00**	**0.00**	2.43	**0.00**	**0.00**	2.79	**0.00**	**0.00**	3.22	**0.00**	**0.00**	3.90
HARCJ IV LOG	**0.00**	**0.00**	1.64	**0.00**	**0.00**	2.43	**0.00**	**0.00**	2.82	**0.00**	**0.00**	3.32	**0.00**	**0.00**	3.82
HARCJ GT LOG	**0.00**	**0.00**	1.79	**0.00**	**0.00**	2.43	**0.00**	**0.00**	2.82	**0.00**	**0.00**	3.22	**0.00**	**0.00**	3.93
HARCJ IVGT LOG	**0.00**	**0.00**	1.57	**0.00**	**0.00**	2.39	**0.00**	**0.00**	2.82	**0.00**	**0.00**	3.36	**0.00**	**0.00**	3.82

Table 5: Models included in the MCS at the 10% confidence level and associated asymmetric quantile loss.

VaR with α = 0.5%		*VaR with α = 1%*		*VaR with α = 1.5%*		*VaR with α = 2%*		*VaR with α = 2.5%*	
Models in MCS	*Loss*	*Models in MCS*	*Loss*	*Models in MCS*	*Loss*	*Models in MCS*	*Loss*	*Models in MCS*	*Loss*
TGARCH IVGT	0.00047	TGARCH IVGT	0.00080	TGARCH IV	0.00110	TGARCH IVGT	0.00138	HARCJ	0.00163
TGARCH IV	0.00049	TGARCH IV	0.00081	TGARCH IVGT	0.00110	TGARCH IV	0.00138	TGARCH IVGT	0.00163
TGARCH	0.00049	TGARCH GT	0.00081	TGARCH GT	0.00111	HARCJ	0.00138	HAR	0.00163
TGARCH GT	0.00049	TGARCH	0.00083	HARCJ	0.00112	TGARCH GT	0.00139	ARFIMA	0.00164
HARCJ	0.00053	HARCJ	0.00085	HAR	0.00113	HAR	0.00139	TGARCH IV	0.00164
HAR	0.00054	HAR	0.00085	TGARCH	0.00114	ARFIMA	0.00139	HARCJ IV LOG	0.00164
HARCJ IV LOG	0.00055	ARFIMA	0.00086	ARFIMA	0.00114	HARCJ IV LOG	0.00140	HARCJ IVGT LOG	0.00164
HARCJ IVGT LOG	0.00055	HARCJ IV LOG	0.00087	HARCJ IV LOG	0.00115	HARCJ IVGT LOG	0.00140	TGARCH GT	0.00164
ARFIMA	0.00056	HARCJ IVGT LOG	0.00087	HARCJ IVGT LOG	0.00115	TGARCH	0.00142	HAR IV LOG	0.00166
HAR IV LOG	0.00057	HAR IV LOG	0.00089	HAR IV LOG	0.00116	HAR IV LOG	0.00142	HARCJ GT LOG	0.00166
HARCJ GT LOG	0.00057	HARCJ GT LOG	0.00089	HARCJ GT LOG	0.00117	HARCJ GT LOG	0.00142	TGARCH	0.00166
HARCJ LOG	0.00057	HAR IVGT LOG	0.00090	HAR IVGT LOG	0.00118	RG IVGT	0.00143	RG IV	0.00167
HAR IVGT LOG	0.00058	HARCJ LOG	0.00090	HARCJ LOG	0.00118	RG IV	0.00143	RG IVGT	0.00167
HAR LOG	0.00060	RG IVGT	0.00090	RG IVGT	0.00118	HAR IVGT LOG	0.00143	HARCJ LOG	0.00167
RG IVGT	0.00060	RG IV	0.00091	RG IV	0.00118	HARCJ LOG	0.00143	HAR IVGT LOG	0.00167
HAR GT LOG	0.00060	HAR LOG	0.00091	RG GT	0.00119	RG GT	0.00144	RG GT	0.00168
RG IV	0.00061	RG GT	0.00092	HAR LOG	0.00119	RG	0.00145	RG	0.00168
HARCJ GT	0.00061	HAR GT LOG	0.00092	RG	0.00120	HAR LOG	0.00145	HAR LOG	0.00168
RG GT	0.00061	RG	0.00093	HAR GT LOG	0.00120	HAR GT LOG	0.00146	HAR GT LOG	0.00169
RG	0.00062	HARCJ GT	0.00093	HARCJ GT	0.00121	ARFIMA GT	0.00147	ARFIMA GT	0.00170
ARFIMA GT	0.00063	ARFIMA GT	0.00094	ARFIMA GT	0.00122	HARCJ GT	0.00147	HARCJ GT	0.00171
HAR IVGT	*eliminated*	HAR IVGT	*eliminated*	HAR IVGT	*eliminated*	HAR IVGT	*eliminated*	HAR IVGT	*eliminated*
HARCJ IVGT	*eliminated*	HARCJ IVGT	*eliminated*	HARCJ IVGT	*eliminated*	HARCJ IVGT	*eliminated*	HARCJ IVGT	*eliminated*
HAR IV	*eliminated*	HARCJ IV	*eliminated*	HARCJ IV	*eliminated*	HARCJ IV	*eliminated*	HARCJ IV	*eliminated*
HARCJ IV	*eliminated*	HAR IV	*eliminated*	HAR IV	*eliminated*	HAR IV	*eliminated*	HAR IV	*eliminated*
HAR GT	*eliminated*	HAR GT	*eliminated*	HAR GT	*eliminated*	HAR GT	*eliminated*	HAR GT	*eliminated*

Table 6: Multinomial VaR test with probability levels α_1=0.5%,α_2=1%,α_3=1.5%,α_4=2%,α_5=2.5%. P-values smaller than 0.05 are in bold font.

Model	*P-values*	***Model***	*P-values*
TGARCH	**0.00**	ARFIMA GT	**0.00**
TGARCH IV	0.08	RG	**0.00**
TGARCH GT	**0.03**	RG IV	**0.00**
TGARCH IVGT	0.10	RG GT	**0.00**
HAR	**0.01**	RG IVGT	**0.00**
HAR IV	**0.00**	HAR LOG	**0.00**
HAR GT	**0.00**	HAR IV LOG	**0.00**
HAR IVGT	**0.00**	HAR GT LOG	**0.00**
HARCJ	**0.00**	HAR IVGT LOG	**0.00**
HARCJ IV	**0.00**	HARCJ LOG	**0.00**
HARCJ GT	**0.00**	HARCJ IV LOG	**0.00**
HARCJ IVGT	**0.00**	HARCJ GT LOG	**0.00**
ARFIMA	**0.00**	HARCJ IVGT LOG	**0.00**

These tables show that only TGARCH models were able to pass the Kupiec and Christoffersen's tests for most quantiles, and to provide VaR violations in % close to the theoretical probability levels. TGARCH models also reported the lowest asymmetric losses for all quantiles up to the 2% probability level, and they provided the most precise VaR forecasts for the most extreme quantiles (0.5% and 1%), which are the most important quantiles for regulatory purposes. However, only the TGARCH model with implied volatility managed to pass most of the VaR tests, including the multinomial back-test with five quantiles, whereas TGARCH models with Google Trends or without any external regressors performed worse.

In general, we found that when both the implied volatility and Google data are added jointly, the parameters estimates of several models became very unstable (see the next section 4.3.2 for more details), while six models out of 32 simply did not reach numerical convergence (four ARFIMA models and two realized-GARCH models). With the exception of TGARCH models, these results highlight that simpler models with no external regressors are a better choice when out-of-sample forecasting is the main concern, thanks to more efficient estimates. Our empirical evidence complements the results provided by Bams et al. (2017), who showed that implied volatility based Value-at-Risk could not outperform simple GARCH based Value-at-Risk due to the complex dependence structure between implied volatility, realized volatility and extreme returns. This is particularly true for Russian financial markets, where extreme returns take place more often than in American markets and they are caused by different type of shocks, from energy economics to geopolitics, see e.g. Malakhovskaya and Minabutdinov (2014), and Aganin and Peresetsky (2018). However, in the case of Russian markets, GARCH models augmented with IV do provide more precise VaR forecasts than simple GARCH models. Moreover, our results reveals another important factor that a model need to possess for successful VaR forecasting: computational robustness in case of frequent and extreme market returns.

4.3.2 Rolling out-of-sample of 250 days

After the previous results, we wanted to verify how the backtesting performance of the competing models changed over time. To achieve this goal, we computed the VaR violations in % for all competing models using a rolling out-of-sample of 250 days (as requested by the Basel agreements). The full color figure reporting the violations in % for the forecasted VaR at the 1% probability level can be found in the supplementary materials posted on the corresponding author's website.

First, the performance of TGARCH models remained remarkably stable over the full period, ranging between 1% and 2%, despite the several episodes of strong volatility in the RTS index future (see Figure 1). Secondly, the HAR models with additional regressors performed very poorly and clearly suffered from computational problems, which resulted in VaR forecasts being strongly underestimated: particularly, the HAR models with both IV and Google data, and the HAR models with IV showed empirical violations higher than 5% and, after 2016, even higher than 15%. Using variables in logarithms solved this numerical problem, but the models' VaR violations were still quite high (between 2% and 4%) and unable to pass the Kupiec and Christoffersen tests. The few realized-GARCH and ARFIMA models which managed to reach numerical convergence behaved similarly to HAR models with variables in logs. One of the main messages that our backtesting analysis conveys is to check the computational robustness of the model used to forecast the VaR or any other risk measure. This is important not only for the Russian market, but also for all emerging markets which may be subject to sudden market crashes due to a variety of reasons.

4.4 Robustness Check: a Hierarchical VAR model with LASSO

We wanted to check how our previous results changed with a multivariate model able to both accommodate a large number of regressors and to improve the model estimation and its forecasting performances. To achieve this goal, we employed the *Hierarchical Vector Autoregression* (HVAR) model estimated with the Least Absolute Shrinkage and Selection Operator (LASSO) proposed by Nicholson et al. (2018). Let us consider the following vector autoregression,

$$Y_t = \boldsymbol{v} + \sum_{l=1}^{22} \Phi^l Y_{t-l} + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \sim WN\left(0, \Sigma_u\right)$$

where Y_t is a 4×1 vector containing the daily returns, the daily realized volatility, the implied volatility and the Google data, $\boldsymbol{v}$ is an intercept vector, while Φ^l are the usual coefficient matrices.

The HVAR approach proposed by Nicholson et al. (2018) adds structured convex penalties to the least squares VAR problem, so that the optimization problem is given by,

$$\min_{\boldsymbol{v},\Phi} \sum_{t=1}^{T} \left\| Y_t - \boldsymbol{v} - \sum_{l=1}^{22} \Phi^l Y_{t-l} \right\|_F^2 + \lambda\left(\mathcal{P}_Y\left(\Phi\right)\right)$$

where $\|A\|_F$ denotes the Frobenius norm of matrix A (that is, the elementwise 2-norm), $\lambda \geq 0$ is a penalty parameter, while $\mathcal{P}_Y(\Phi)$ is the group penalty structure on the endogenous coefficient matrices. The HVAR class of models solves the problem of an increasing maximum lag order by including the lag order into hierarchical group LASSO penalties, which induce sparsity and a low maximum lag order.

For our empirical work, we employed the *elementwise penalty function,*

$$\mathcal{P}_Y(\Phi) = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{l=1}^{22}\left\|\Phi_{ij}^{l:22}\right\|_2$$

which is the most general structure, because every variable in every equation is allowed to have its own maximum lag resulting in 4^2 possible lag orders. The penalty parameter λ is estimated by sequential cross-validation, see Nicholson et al. (2018) for the full details. The p-values of the Kupiec and Christoffersen's tests, the number of violations in %, and the p-values of the Multinomial VaR test by Kratz et al. (2018) with probability levels α_1=0.5%, α_2=1%, α_3=1.5%, α_4=2% and α_5=2.5% are reported in Table 7, while the VaR violations in % for the forecasted $\text{VaR}_{1\%}$ using a rolling out-of-sample of 250 days are reported in Figure 3.

Table 7: Kupiec tests p-values, Christoffersen's tests p-values, and Multinomial VaR test with probability levels α_1=0.5%,α_2=1%,α_3=1.5%,α_4=2%,α_5=2.5%. *P-values smaller than 0.05 are in bold font.*

	Kupiec t. (p-value)	*Christ. T. (p-value)*	*Violations %*	*Multinomial VaR test (p-value)*
VaR with α = 0.5%	**0.00**	**0.00**	1.67	**0.00**
VaR with α = 1%	**0.00**	**0.00**	2.02	
VaR with α = 1.5%	**0.00**	**0.00**	2.30	
VaR with α = 2%	**0.01**	**0.00**	2.69	
VaR with α = 2.5%	**0.02**	**0.00**	3.19	

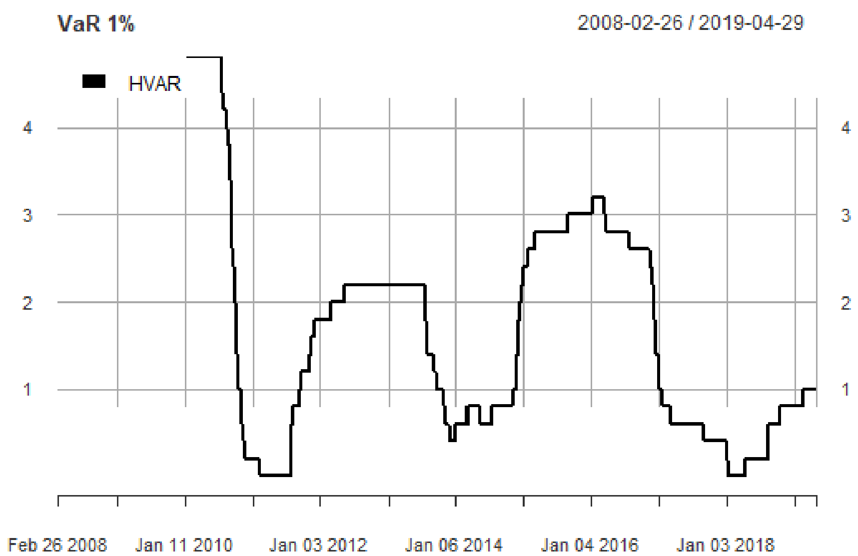

Figure 3: Violations in % for the forecasted $VaR_{1\%}$ *using a rolling out-of-sample of 250 days.*

The HVAR model solves the numerical problems of the HAR model with additional variables in levels, but it has a backtesting performance similar to the HVAR models with variables in logarithms: that is, it underestimates the VaR following episodes of extremely high volatility.

5 Conclusions

We evaluated the contribution of both online search intensity and options-based implied volatility to the modelling of the volatility of the Russian RTS index future, and we examined how this dependence changed over almost two decades. We found that both the sign and the size of their coefficients changed significantly, particularly in the periods following the beginning of the global financial crisis in 2008 and (to a lower degree) after the introduction of sanctions in 2014.

We then performed a backtesting analysis involving the forecasting of the Value-at-Risk for the RTS index future at multiple confidence levels using several alternative models specifications, with and without Google data and implied volatility. We found that only TGARCH models were able to pass the Kupiec and Christoffersen's tests for most quantiles, and they also reported the lowest asymmetric losses for all quantiles up to the 2% probability level. However, only the TGARCH model with implied volatility managed to pass almost all back-tests, including the multinomial test with five quantiles needed to back-test the expected shortfall, whereas TGARCH models with Google Trends or without any external regressors did not. We noticed that when both the implied volatility and Google data were added jointly, the parameters estimates of several models became very unstable and several models did not reach numerical convergence (particularly, ARFIMA and realized-GARCH models). Moreover, with the exception of TGARCH models, our results highlighted that simpler models with no additional regressors provided better VaR forecasts than augmented models. This empirical evidence complements the results provided by Bams et al. (2017), who showed that forecasting the volatility is different from forecasting a certain quantile of the return distribution, hence models forecasting well the former may not forecast well the latter. However, in the case of Russian markets, TGARCH models augmented with IV did provide better VaR forecasts than TGARCH models without it. We also evaluated the backtesting performance of the competing models using a rolling out-of-sample of 250 days: we found that the performance of TGARCH models remained remarkably stable over the full evaluation period, whereas HAR models with additional regressors performed very poorly and clearly suffered from computational problems, which resulted in VaR forecasts being strongly underestimated. Using variables in logarithms solved this numerical problem, but the models' VaR violations were still quite high and unable to pass the usual Kupiec and Christoffersen tests. The few realized-GARCH and ARFIMA models which managed to reach numerical convergence behaved similarly to HAR models with variables in logs. Therefore, one of the main guidance that emerged from our backtesting analysis is to check the computational robustness of the model employed to forecast the VaR (or any other risk measure) in case of extreme and sudden market crashes. Finally, we also performed a robustness check to verify how our previous results changed with a hierarchical-VAR model with LASSO able to both accommodate a large number of regressors and to improve the model estimation and its forecasting performances. The HVAR model solved the numerical problems of the HAR models with additional variables in levels, but it still underestimated the VaR in the periods following episodes of extremely high volatility and abrupt market changes.

In general, models with implied volatility performed better than models with Google Trends data, thus confirming similar evidence reported by Basistha et al. (2018) and Bazhenov and Fantazzini (2019). These authors suggested two possible explanations for these results: first, the informational content included in Google search activity is also present in the implied volatility, but the opposite is not true, due to the fact that implied volatility is a forward-looking measure based on the expectations of large investors who have access to premium and insider information, while Google Trends data are mainly based on the expectations of small investors and un-informed traders. A second simpler explanation is that Yandex is the main search engine in Russia, so that Google Trends may not be the best proxy for Russian investors' interest and behavior. If Yandex will make available online search data at the daily frequency and for long periods, then this issue will definitely be an interesting avenue of future research.

References

Acerbi C., Tasche D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, 26 (7), 1487–1503.

Aganin A. (2017). Forecast comparison of volatility models on Russian stock market. *Applied Econometrics*, 48, 63-84.

Aganin A., Peresetsky A. (2018). Volatility of ruble exchange rate: oil and sanctions. *Applied Econometrics*, 52, 5-21.

Agnolucci P. (2009). Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, 31 (2), 316-321.

Ahoniemi K., Lanne M. (2013). Overnight stock returns and realized volatility. *International Journal of Forecasting*, 29 (4), 592-604.

Andersen T. G., Bollerslev T., Diebold F. X., Ebens H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61 (1), 43–76.

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

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 and Statistics*, 89 (4), 701–720.

Artzner P., Delbaen F., Eber J. M., Heath D. (1997). Thinking coherently. *Risk*, 10 (11), 68–71.

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

Bai J. (1997). Estimating multiple breaks one at a time. *Econometric Theory*, 13, 315–352.

Bai J., Perron P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, 47–78.

Bai J., Perron P. (2003a). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics,* 18, 1–22.

Bai J., Perron P. (2003b). Critical values for multiple structural change tests. *Econometrics Journal*, 18, 1–22.

Bams D., Blanchard G., Lehnert T. (2017). Volatility measures and Value-at-Risk. *International Journal of Forecasting*, 33 (4), 848-863.

Barndorff-Nielsen O. E., Shephard N. (2004). Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72 (3), 885–925.

Barndorff-Nielsen O. E., Shephard N. (2004b). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2 (1), 1–37.

Barndorff-Nielsen O. E., Shephard N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4 (1), 1–30.

Barone-Adesi G., Finta M. A., Legnazzi C., Sala C. (2019). WTI crude oil option implied VaR and CVaR: an empirical application. *Journal of Forecasting*, forthcoming.

Basel Committee on Banking Supervision (2009). *Findings on the interaction of market and credit risk*. Bank for International Settlements, Working paper n. 16, May.

Basel Committee on Banking Supervision (2013). *Fundamental review of the trading book: A revised market risk framework*. Consultative Document, October.

Basel Committee on Banking Supervision (2016). *Minimum capital requirements for market risk*. Consultative Document, January

Basistha A., Kurov A., Wolfe M. (2018). Volatility forecasting: the role of internet search activity and implied volatility. West Virginia University working paper.

Bazhenov T., Fantazzini D. (2019). Forecasting realized volatility of Russian stocks using Google Trends and implied volatility. *Russian Journal of Industrial Economics*, 12 (1), 79-88.

Bianchi C., De Giuli M. E., Fantazzini D., Maggi M. (2011). Small sample properties of copula-GARCH modelling: a Monte Carlo study. *Applied Financial Economics*, 21 (21), 1587-1597.

Birkelund O. H., Haugom E., Molnár P., Opdal M., Westgaard S. (2015). A comparison of implied and realized volatility in the Nordic power forward market. *Energy Economics*, 48, 288-294.

Busch T., Christensen B. J., Nielsen M. (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.

Cai Y., Krishnamoorthy K. (2006). Exact size and power properties of five tests for multinomial proportions. Communications in Statistics, Simulation and Computation, 35 (1):149–160.

Campos I., Cortazar G., Reyes T. (2017). Modeling and predicting oil VIX: internet search volume versus traditional variables. *Energy Economics*, 66, 194-204.

Caporale G. M., Luis A., Trilochan T. (2019). Volatility persistence In the Russian stock market. *Finance Research Letters*, forthcoming.

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

Chong J. (2004). Value at risk from econometric models and implied from currency options. *Journal of Forecasting*, 23 (8), 603–620.

Christensen B. J., Prabhala N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, 50 (2), 125–150.

Christoffersen P. (1998). Evaluating interval forecasts. *International economic review*, 39, 841-862.

Christoffersen P. (2012). *Elements of financial risk management*. Academic Press.

Christoffersen P., Mazzotta S. (2005). The accuracy of density forecasts from foreign exchange options. *Journal of Financial Econometrics*, 3 (4), 578–605.

Corredor P., Santamaría R. (2004). Forecasting volatility in the Spanish option market. *Applied Financial Economics*, 14 (1), 1-11.

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

Da Z., Engelberg J., Gao P. (2011). In search of attention. *Journal of Finance,* 66 (5), 1461–1499.

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

Emmer S., Kratz M., Tasche D. (2015).What Is the Best Risk Measure in Practice? *Journal of Risk,* 18,31-60.

Fantazzini D. (2019) *Quantitative Finance with R and Cryptocurrencies*. Amazon KDP, ISBN-13: 978-1090685315.

Fantazzini D., Toktamysova Z. (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics*, 170, 97-135.

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

Giot P. (2005). Implied volatility indexes and daily value at risk models. *The Journal of Derivatives*, 12 (4), 54–64.

Glosten L., Jagannathan R., Runke D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance,* 48, 1779–1801

Gneiting T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106 (494), 746-762.

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

González-Rivera G., Lee T. H., Mishra S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of forecasting*, 20 (4), 629-645.

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

Hansen P. R., Lunde A. (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 3, 525–554

Hansen P. R., Lunde A. (2005b). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, 20 (7), 873-889.

Hansen P. R., Lunde A., Nason J.M. (2011). The model confidence set. *Econometrica*, 79 (2), 453-497.

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

Haugom E., Langeland H., Molnár P., Westgaard S. (2014). Forecasting volatility of the US oil market. *Journal of Banking and Finance*, 47, 1-14.

Huang X., Tauchen G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3 (4):456–499.

Hyndman R. J., Khandakar Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27 (3), 1-22.

Hwang S., Valls Pereira P. L. (2006). Small sample properties of GARCH estimates and persistence. *The European Journal of Finance*, 12 (6-7), 473-494.

Jeon J., Taylor J.W. (2013). Using CAViaR models with implied volatility for Value-at-Risk estimation. *Journal of Forecasting*, 32 (1), 62-74.

Jorion, P. (2007). *Financial Risk Manager Handbook*. Vol. 406. Wiley.

Kratz M., Lok Y. H., McNeil A. J. (2018). Multinomial VaR backtests: a simple implicit approach to backtesting expected shortfall. *Journal of Banking and Finance*, 88, 393-407.

Kupiec P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3 (2), 73-84.

Liu J., Wu S., Zidek J. V. (1997). On segmented multivariate regression. *Statistica Sinica,* 7, 497–525

Liu L. Y., Patton A. J., Sheppard K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187 (1), 293-311.

Malakhovskaya O., Minabutdinov A. (2014). Are commodity price shocks important? A Bayesian estimation of a DSGE model for Russia. *International Journal of Computational Economics and Econometrics*, 4 (1/2), 148-180.

Martens M., Zein J. (2004). Predicting financial volatility: high-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets*, 24 (11), 1005-1028.

Mayhew S. (1995). Implied volatility. *Financial Analysts Journal*, 51 (4), 8-20.

McNeil A. J., Frey R., Embrechts P. (2015). *Quantitative risk management: concepts, techniques and tools.* Revised edition. Princeton university press.

Meddahi N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17 (5), 479–508

Newey W. K., West K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61 (4), 631-653.

Nicholson W. B., Wilms I., Bien J., Matteson D. S. (2018). High dimensional forecasting via interpretable vector autoregression. arXiv preprint arXiv:1412.5250.

Perron P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics*, 1 (2), 278-352.

Pesaran M. H., Timmermann A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137 (1), 134-161.

Pong S., Shackleton M. B., Taylor S. J., Xu X. (2004). Forecasting currency volatility: a comparison of implied volatilities and AR(FI)MA models. *Journal of Banking and Finance*, 28 (10), 2541–2563.

Seo M., Lee S., Kim G. (2019). Forecasting the Volatility of Stock Market Index Using the Hybrid Models with Google Domestic Trends. *Fluctuation and Noise Letters*, 18 (01), 1950006.

Taylor S. J., Xu X. (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance*, 4 (4), 317–340.

Vlastakis N., Markellos R. N. (2012). Information demand and stock market volatility. *Journal of Banking and Finance*, 36 (6), 1808-1821.

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

Xu Q., Bo Z., Jiang C., Liu Y. (2019). Does Google search index really help predicting stock market volatility? Evidence from a modified mixed data sampling model on volatility. *Knowledge-Based Systems*, 166, 170-185.

Wimmerstedt L. (2015). Backtesting expected shortfall: the design and implementation of different backtests. Technical report, Swedish Royal Institute of Technology.

Yao Y. C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters*, 6 (3), 181-189.

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

Zeileis A., Leisch F., Hornik K., Kleiber C. (2002). strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7(2), 1–38

Zeileis A. (2005). A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals. *Econometric Reviews*, 24 (4), 445-466.

Zeileis A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, 16 (9), 1–16.

Zeileis A., Shah A., Patnaik I. (2010). Testing, monitoring, and dating structural changes in exchange rate regimes. *Computational Statistics and Data Analysis*, 54 (6), 1696-1706.