Does trading volume help forecast volatility?

Stanislav Anatolyev*, Sergei Ponomarev†

Abstract

We explore a predictive relationship between volatility and trading volume within the class of HAR type models for realized volatility. We propose a new way to incorporate trading volume into the model of volatility forecasts. In the spirit of the HAR model, we include average trading volumes for the past one, five, and 22 trading days, in various combinations, and for different horizons. We verify whether these averages are able to increase out-of-sample quality of volatility forecasts. We discover that for a one-day horizon, the lags of trading volume significantly enhance the quality of out-of-sample volatility forecasts compared to HAR models without trading volume and with only a one-day trading volume. This result is robust to model specification, choice of loss function, and asset type. In contrast, for both week and month horizons, the quality of volatility forecasts typically does not increase. A plausible reason for forecast enhancement is that the lagged average trading volume represents not only information coming to the market but also signals how fast the incoming information spreads across the market.

Keywords: volatility forecasting; realized volatility; trading volume; heterogeneous autoregressive model.

JEL codes: C22, C53, C58, G17

*Corresponding author. Address: CERGE-EI, Politických vězňů 7, 11121 Prague 1, Czech Republic. E-mail: stanislav.anatolyev@cerge-ei.cz.

†New Economic School. E-mail: sponomarev@nes.ru.

1 Introduction

Volatility forecasting is very important in financial econometrics. Good volatility forecasts are in demand in order to price an option or determine the best trading or hedging strategy. There are numerous other uses for volatility. A lot of academic papers try to find whether the trading volume can increase the quality of volatility forecasts. In this paper, we propose a new way to incorporate trading volume into the model of volatility forecasts. We take as a starting point the heterogeneous autoregressive model for realized volatility (HAR-RV), a popular and attractive workhorse model by Corsi (2009), and add average trading volumes for the past one, five, and 22 trading days, in various combinations, similarly to how the baseline HAR-RV model includes past average realized volatilities. We verify whether these past average volumes are able to improve in-sample and especially out-of-sample quality of volatility forecasts, and how such improvements (if any) vary with the forecast horizon.

Specifically, we use one-minute data on prices and trading volumes for 41 assets contained in the DOW30 index. Using the price data, we compute realized volatility, a consistent measure of integrated volatility. Exploiting the baseline HAR-RV model and its several modifications, we make forecasts of realized volatility for the day, week, and month horizons. After that, we add the corresponding lags of trading volume, in the spirit of the HAR-RV model. Thus, for each version of the HAR model, each forecast horizon, and each of the 41 assets,[1] we compare forecasts produced by four models: one without trading volume, one with one lag of trading volume (i.e., daily volume measure), one with trading volume for one and five trading days (i.e., daily and average weekly volume measures), and one with trading volume for one, five and twenty two trading days (i.e., daily, average weekly, and average monthly volume measures). In all comparisons, we use out-of-sample MSE and QLIKE losses. To make the forecast comparisons, we first informally look at values of the out-of-sample R^2 as a simple measure of forecast quality. Second, and more formally, we construct model confidence sets (MCS) (Hansen, Lunde & Nason, 2011) using values for both loss measures, corresponding to each model.

[1] Even though DOW30 contains 30 components at any given time, the composition has been varying over years, hence the total number 41; for details, see Section 5.

The results are sharply different for different forecast horizons. For the one-day forecast horizon, inclusion of lagged weekly and/or monthly volume into a model with no volume or with only lagged daily volume significantly enhances forecasting quality, as attested by the 90% level MCS for 83% of all assets. In terms of out-of-sample R^2, adding the weekly or weekly together with monthly volume on top of the daily volume improves forecasting quality four times more than adding only the daily volume to the baseline model, and three times more than the difference in forecasting quality among different HAR-RV specifications. For the one-week forecast horizon, there is a similar tendency, albeit weaker: as the 90%-level model confidence sets attest, a similar inclusion increases the quality of the volatility prediction in about 40% cases, and never decreases it. For the one-month forecast horizon, however, the inclusion of the weekly and monthly volume only in less than 27% cases significantly increases volatility forecast quality, and sometimes significantly decreases it. These results are robust to HAR-RV type model specification, choice of loss function, and asset type.

We suggest two explanations for such increases in forecast quality. One is very close to intuition behind the HAR-RV model (Corsi, 2009). It is based on the Heterogeneous Market Hypothesis, which states that traders in the financial market are heterogeneous because of differences in agents' endowments, institutional constraints, risk attitudes, temporal horizons, etc. The financial markets are composed of participants who have a large spectrum of trading frequencies. For example, dealers, market makers, and intra-daily speculators trade at a very high intra-daily frequency. While insurance companies and pension funds trade much less frequently and possibly for larger amounts, the main idea is that agents with different time horizons perceive and react to different news (for which trading volume is a proxy), which results in different trading volumes for different periods. There can also be differences in the speed of information arrival to the market for different types of traders. Thus, trading volume for different time spans should make a difference for volatility prediction.[2]

The second explanation is that the lags of average trading volume represent two types of

[2] Another way to formulate the same intuition is: trading volume and price changes (i.e., volatility) change only when some information comes to the market. As historical volatility for various horizons matters for volatility prediction, the flow of information across various horizons also matters. Thus, since trading volume also reflects an information flow (in a different way than volatility), trading volume for different time spans should also matter for volatility prediction.

information. Firstly, the first lag of average trading volume represents the shocks that came to the market in the previous day. When some information comes to the market, trading volume increases because informed traders who have already received this shock trade with uninformed traders who do not yet know about the shock. Then, after some time, when most of the traders receive the same information, the price changes. Thus, the increase in trading volume yesterday signals that some information has come to the market and that the prices will adjust to this information later, and hence volatility should increase. Secondly, the lags of trading volume are a proxy that describes how fast information spreads in the market. If trading volume is low, information spreads very fast because the prices adjust very fast and the traders do not want to trade intensively when the price has quickly changed. For the same reason, if trading volume is high, information spreads slowly.

When there is no trading volume, such information is not coming to the market. If the model includes only the lag of daily volume, trading volume affects volatility by two different channels. Through one channel, it tends to increase volatility, while through the other channel it tends to decrease volatility. The past daily volume represents an average of impacts though these two opposite channels, and thus is not supposed to significantly increase volatility prediction quality. However, if the model includes the past average weekly trading volume, this predictor represents only the propagation speed of information, as prices are usually adjusted to the shock by less than two days, so the shock spreads fast enough and the weekly volume usually cannot represent the information about the shocks that come to the market. This information is instead represented by the past daily trading volume. Thus, each of these predictors affects volatility through its own single channel, and the quality of volatility prediction increases.

The rest of the paper is organized as follows. Section 2 describes the role of trading volume in forecasting volatility and evidence found in empirical literature. Section 3 discusses our modeling approach, Section 4 lists statistical procedures used, and Section 5 describes the data. Section 6 contains the results and their discussion. Finally, Section 7 concludes.

2 Trading volume

The relationship between trading volume and volatility has received significant attention in the literature. However, there is mixed evidence about whether trading volume can increase the quality of volatility forecasts, and results vary across asset types, data periods, and different horizons of forecasting.

A few empirical studies have found that trading volume does not improve the quality of volatility prediction. For example, Brooks (1998) compares 31 different volatility models and concludes that the predictive power of trading volume is negligible. Kambouroudis & McMillan (2016) discover that trading volume does not improve the accuracy of volatility prediction. However, several studies have discovered that trading volume does have predictive power for volatility. For example, Wang, Wu & Xu (2015) add previous-day trading volume into the HAR-RV model and find that it significantly increases prediction quality. Mougoue & Aggarwal (2011) discover as well that trading volume is able to improve the quality of volatility forecasts.

To the best of our knowledge, no existing study attempts to understand how the term structure of historical trading volume can enhance volatility prediction and uses average trading volume for more lengthy periods as we do, confining to inclusion of only previous-day trading volume as a predictor. Note also that we use a wide set of assets to ensure a high degree of validation, and analyze volatility forecasting at different horizons.

From a theoretical perspective, there are two leading theories about the relationship between trading volume and volatility: the Mixture of Distributions Hypothesis (MDH) and the Sequential Information Arrival Hypothesis (SIAH). MDH suggests that trading volume does not increase the quality of volatility prediction; on the contrary, SIAH suggests that trading volume does increase forecasting quality. For both hypotheses, there are studies that find evidence why these hypotheses hold for some markets, and other studies that show that these hypotheses do not hold. For example, Mougoue & Aggarwal (2011) show that SIAH holds, while Liu, Lee & Choo (2020) find no evidence in its favor. In both theories, trading volume is considered as a proxy for the information flow that comes into financial markets.

According to the MDH, prices and volume change simultaneously when information arrives to the market. The MDH assumes that the stock return is drawn from a joint distribution of volume and prices, conditional on the current information. Price and trading volume changes are driven by a simultaneous underlying information arrival process, hence volume and volatility are interrelated. This hypothesis predicts a contemporaneous volume–volatility relation. The MDH does not imply serial dependence in return volatility and volume.

The SIAH argues that each trader observes information sequentially. Traders who have already received new information will increase trading volume, while the asset price will not fully accommodate according to this new information. Thus, an increase in trading volume implies that prices will change in the future, causing volatility to increase. Hence, due to the sequential information flow, the lagged trading volume possesses predictive ability for the current realized volatility.

To summarize, the MDH suggests that trading volume does not provide additional information that improves the accuracy of volatility forecasting. On the contrary, the SIAH states that there is a lead-lag relation between trading volume and volatility.

Note that if trading volume increases the quality of volatility prediction, then the SAIH is not true, whereas if trading volume makes prediction better, this becomes important evidence in favor of SIAH. Several studies reach a conclusion about whether the SIAH is true or false solely based on whether past trading volume is an important predictor for volatility or not; see, for example, Mougoue & Aggarwal (2011) and Liu, Lee & Choo (2020). Because we find that trading volume does matter for volatility prediction, we suggest that there is significant evidence in favor of SIAH, at least for the shares studied.

To forecast realized volatility in day t, we compute the average trading volume for p previous days, where $p \in \{1, 5, 22\}$, and include it (or more, precisely, its log transformation), in various combinations, in HAR-type models of volatility prediction, in the spirit of the baseline HAR-RV model. Formally, we have observations for trading volume TR_t for each day t, and compute average lagged trading volume:

$$TR_t^p = \frac{1}{p} \sum_{i=1}^{p} TR_{t-i}.$$

So, TR_t^p is an average trading volume for p days that are immediately preceding day t. The values $p = 1$, $p = 5$ and $p = 22$ correspond to inclusion of lagged average daily, weekly, and monthly trading volumes, respectively.

We choose a certain restricted number of combinations of lags of trading volume:

$$\bigcup_{p \in \Omega} TR_t^p,$$

where Ω is the set of lags included in a model of volatility forecasting. Specifically, we use $\Omega = \varnothing$, $\Omega = \{1\}$, $\Omega = \{1, 5\}$, and $\Omega = \{1, 5, 22\}$. We use the first two sets as baselines for comparison, i.e. the models with no trading volume and with only lagged daily trading volume, which can be found in HAR-RV type models in the previous literature. In the augmented models, we use the last two sets, which will give evidence on how longer-ranged past volumes – weekly in a more important degree and monthly in a less important degree – improve prediction. In a nutshell, we show that these augmented models do tend to improve volatility forecasts compared to the baselines, although the exact tendencies sharply depend on the forecast horizon – daily, weekly or monthly.

3 HAR type models

The models on which we build is the heterogeneous autoregressive model for realized volatility (HAR-RV) proposed by Corsi (2009), and its modifications. The HAR-RV type models use realized volatility components for the previous day, previous week, and previous month to forecast today's volatility. The intuition of these models is based on the Heterogeneous Market Hypothesis that traders are heterogeneous – agents with different time horizons perceive, react to, and cause different types of volatility components. There are three types of investors that lead to three primary volatility components: short-term traders with a daily or higher trading frequency, medium-term investors who typically rebalance their positions weekly, and long-term agents with a characteristic time of one or more months. As a result, to forecast today's volatility, volatilities from the previous day, week, and month are all important.

There are three reasons for using HAR type models. The first reason is that the HAR models are among the best models to fit the realized volatility process. For example, Liu, Patton & Sheppard (2015) compare the HAR-RV model estimated on five-minute data with nearly 400 other models, and discover that HAR-RV provides one of highest quality forecasts in terms of in- and out-of-sample performance by way of the MCS approach and various other tests. Next, according to Corsi (2009), HAR-RV type models are able, in spite of its simplicity, to reproduce such an important feature of volatility as the long memory property. Finally, HAR-RV type models forecast returns well for different horizons: one day, one week, and one month, while other models usually predict returns only for only one horizon.

In addition, the Heterogeneous Market Hypothesis claiming that traders are heterogeneous is an explanation of why the HAR-RV model works well, and is also an explanation of why trading volumes for different periods have to matter for volatility forecasting. The trading volume and price change only when some new information comes to the market. Since trading volume reflects information flow as well but in a different way than volatility, trading volume over periods of different length should also matter for volatility forecasting. Thus, by intuition of the HAR-RV model, it follows that average trading volumes for periods of different length should be good additional predictors for volatility.

HAR-RV type models are formulated for observable realized volatility (RV), which is a consistent (in the sense of the in-fill asymptotics) measure of integrated volatility, which is true unobserved volatility. To compute RV, we need to use the intradaily stock returns, which are, for intradaily period i of day t, by definition, $r_{t,i} = \ln(p_{t,i}) - \ln(p_{t,i-1})$, where $p_{t,i}$ is a midquote (i.e., the arithmetic average between bid open and ask open prices) at intradaily period i of day t. We use five-minute data for regular trading sessions, so typically we have 79 observations of intradaily prices for one day and thus 78 observations of intradaily returns.

Using these intradaily returns, we calculate realized volatility, RV_t, for day t :

$$RV_t = \frac{1}{N}\sum_{i=1}^{N} r_{t,i}^2,$$

where N is the number of observations of returns during the day. Then, we calculate the average realized volatility for the previous day, week (that is, previous 5 trading days), and month (that is, previous 22 trading days): $RV_t^d = RV_{t-1}$, $RV_t^w = \frac{1}{5}\sum_{k=1}^{5} RV_{t-k}$, and

$RV_t^m = \frac{1}{22}\sum_{k=1}^{22} RV_{t-k}$. Hence, RV_t^d is a previous day realized volatility, RV_t^w is a previous week volatility, and RV_t^m is a previous month volatility.

The HAR-RV model of Corsi (2009) linearly predicts daily volatility RV_t by the daily, weekly and monthly components, RV_t^d, RV_t^w and RV_t^m. In the literature that followed, however, in addition to the original linear-in-levels version, there is another common version of the use of realized volatility in HAR-RV type models – using realized volatility in a log-transformed form. There is evidence that log specification outperforms a specification in levels in terms of quality of volatility forecasting. For example, Corsi, Pirino & Reno(2010) show the logs specification leads to lower out-of-sample losses than the levels specification for the HAR-RV and HAR-RV-CJ (one of extensions, see below) models. Similar evidence is found by Duong and Swanson (2015): log specifications outperform levels specifications in terms of out-of-sample R^2 for the models we use in our analysis. Hence, as we are trying to improve the quality of volatility prediction, we start from the log specification. To simplify notation, let us denote $rv_t = \ln RV_t$, $rv_t^d = \ln RV_t^d$, $rv_t^w = \ln RV_t^w$, $rv_t^m = \ln RV_t^m$; we will follow similar mnemonics for other log-transformed variables throughout. Then, the basic HAR-RV model has the following form:

$$rv_t = a_0 + a_d rv_t^d + a_w rv_t^w + a_m rv_t^m + \epsilon_t.$$

In this and all other regressions that follow, the error ϵ_t is assumed to be an MDS white noise.

To ensure external validity of our conclusions, we consider several extensions from the basic HAR-RV model, and perform the same experiments on these extended HAR-type models. The first extension, HAR-CJ, is the HAR-RV model with the volatility decomposed to the continuous and jump components, which was proposed by Andersen, Bollerslev & Diebold (2007). This is one of the most commonly used extensions of HAR-RV in the related literature. In this extension, the volatility is divided into a jump components – volatility that is much higher than 'average' volatility, and the rest, a continuous component. The intuition for this division is that the effect of lagged volatility on current volatility is non-linear, so the high lagged volatility represented by the jump component has a different impact on current volatility than the continuous component.

Note that volatility in the morning is usually higher than during the rest of a day. Boudt, Croux, & Laurent (2011) show that because of this periodicity, jumps will occur more frequently than they should, lowering the quality of HAR model forecasts. In order to prevent this, we use the weighted standard deviation (WSD) approach of Boudt, Croux, & Laurent (2011) to standardize the return and filter out the intradaily diurnal pattern. For the day t and intradaily period i, the filtered returns are denoted as $\hat{r}_{t,i}$. Next, we divide volatility into jump and continuous components:[3]

$$C_t = \sum_{i=1}^{N} r_{t,i}^2 \left(1 - \mathcal{J}_{t,i}\right),$$

$$J_t = \sum_{i=1}^{N} r_{t,i}^2 \mathcal{J}_{t,i},$$

where C_t stands for continuous volatility for day t, and J_t stands for jump volatility for day t. The variable $\mathcal{J}_{t,i}$ is a binary variable that is equal to 1 if the jump occurred in minute i in day t. A return is classified as a jump if it lies in the tail of the volatility distribution. According to Andersen, Bollerslev & Dobrev (2007), the return is classified as a jump if it lies in the top 0.1% of the distribution of absolute values of returns. According to Andersen, Bollerslev & Dobrev (2007), this happens if

$$|\hat{r}_{t,i}| > \Phi\left(1 - \frac{1-(1-10^{-5})^{1/N}}{2}\right)^{-1} \sqrt{\frac{\pi}{2N} \sum_{i=2}^{N} |r_{t,i}|\,|r_{i-1,t}|},$$

where Φ stands for the CDF of the standard normal distribution, and N is a number of observations per day. Note that on the left hand side of this equation, the standardized returns are used. The average jump components of realized volatility for the previous day, J_t^d, previous week, J_t^w, and previous month J_t^m, as well as the average continuous components of realized volatility for the previous day, C_t^d, previous week C_t^w, and previous month C_t^m

[3]There are several ways to estimate the jump and continuous components of volatility. We have decided to use the approach of Andersen, Bollerslev & Dobrev (2007), the most commonly used method that is specifically designed for the intra-daily detection of jumps and thus allows us to divide jumps and continuous components into positive and negative components.

are computed as follows: $J_t^d = J_{t-1}$, $J_t^w = \frac{1}{5}\sum_{k=1}^{5} J_{t-k}$, $J_t^m = \frac{1}{22}\sum_{k=1}^{22} J_{t-k}$; $C_t^d = C_{t-1}$, $C_t^w = \frac{1}{5}\sum_{k=1}^{5} C_{t-k}$, $C_t^m = \frac{1}{22}\sum_{k=1}^{22} C_{t-k}$. According to our mnemonic rule, we denote $j_t^d = \ln(J_t^d{+}1)$, $j_t^w = \ln(J_t^w{+}1)$, ..., $c_t^w = \ln(C_t^w)$, $c_t^m = \ln(C_t^m)$. Note that for the jump component we add unity because the jump component is typically zero. The continuous component is strictly positive in more than 99.999% of day observations; if for some observation it is zero, such an observation is dropped, so in the end, less than 0.01% of observations are excluded. With the notation introduced, the HAR-CJ model of Andersen, Bollerslev, Diebold (2007) reads

$$rv_t = a_0 + a_d c_t^d + a_w c_t^w + a_m c_t^m + b_d j_t^d + b_w j_t^w + b_m j_t^m + \epsilon_t.$$

The second extension is the HAR-RV model, where volatility is decomposed into positive and negative volatility, so called HAR-GB, proposed in Patton & Sheppard (2015). The authors show that this model outperforms the basic HAR-RV model in terms of out-of-sample volatility prediction. They also show that the model that decomposes volatility into positive, negative, jump, and continuous components predicts volatility out-of-sample better than any other commonly used HAR-RV type model.

So, the volatility is divided into positive (i.e., volatility with an increase in price) and negative (i.e., volatility with a decrease in price). The intuition behind this division is that an increase in price represents different news than a decrease in price, and hence will affect volatility in the future differently. Then, for day t, the positive component of volatility, say RV_t^+, and the negative component of volatility, say RV_t^-, are defined as follows:

$$RV_t^- = \frac{1}{N}\sum_{i=1}^{N} r_{t,i}^2 \mathbb{I}_{\{r_{t,i}<0\}},$$

$$RV_t^+ = \frac{1}{N}\sum_{i=1}^{N} r_{t,i}^2 \mathbb{I}_{\{r_{t,i}>0\}},$$

where N is a number of observations on returns during the day, and $\mathbb{I}_{\{\cdot\}}$ is an indicator function. Note that positive and negative volatility is always larger than zero. The average positive realized volatility for the previous day, RV_t^{d+}, previous week RV_t^{w+} and previous month RV_t^{m+}, and the average negative realized volatility for the previous day, RV_t^{d-}, previous week RV_t^{w-} and previous month RV_t^{m-} are computed similarly to previously defined variables: $RV_t^{d+} = RV_{t-1}^+$, $RV_t^{w+} = \frac{1}{5}\sum_{k=1}^{5} RV_{t-k}^+$, $RV_t^{m+} = \frac{1}{22}\sum_{k=1}^{22} RV_{t-k}^+$; $RV_t^{d-} = RV_{t-1}^-$, $RV_t^{w-} = \frac{1}{5}\sum_{k=1}^{5} RV_{t-k}^-$, $RV_t^{m-} = \frac{1}{22}\sum_{k=1}^{22} RV_{t-k}^-$, with obvious mnemonics for corresponding log-transformed variables. The HAR-GB model with decomposed risk premia of Patton & Sheppard (2015) reads

$$rv_t = a_0 + a_d^+ rv_t^{d+} + a_d^- rv_t^{d-} + a_w^+ rv_t^{w+} + a_w^- rv_t^{w-} + a_m^+ rv_t^{m+} + a_m^- rv_t^{m-} + \epsilon_t.$$

An even more general model named HAR-GB-CJ and also described in Patton & Sheppard (2015) includes both divisions: to positive and negative volatility, on the one hand, and to jump and continuous components, on the other. Define the positive jump component of volatility J_t^+, negative jump component of volatility J_t^-, positive continuous component of volatility C_t^+, and negative continuous component of volatility C_t^- as

$$C_t^- = \frac{1}{N}\sum_{i=1}^{N} r_{t,i}^2 \left(1 - \mathcal{J}_{t,i}\right) \mathbb{I}_{\{r_{t,i}<0\}}, \quad C_t^+ = \frac{1}{N}\sum_{i=1}^{N} r_{t,i}^2 \left(1 - \mathcal{J}_{t,i}\right) \mathbb{I}_{\{r_{t,i}>0\}},$$

$$J_t^- = \frac{1}{N}\sum_{i=1}^{N} r_{t,i}^2 \mathcal{J}_{t,i} \mathbb{I}_{\{r_{t,i}<0\}}, \quad J_t^+ = \frac{1}{N}\sum_{i=1}^{N} r_{t,i}^2 \mathcal{J}_{t,i} \mathbb{I}_{\{r_{t,i}>0\}},$$

where $\mathcal{J}_{t,i}$ is the same jump detection as in the HAR-CJ model. Then, for each volatility component we calculate it for the previous day, week, and month: $J_t^{d+} = J_{t-1}^+$, $J_t^{w+} = \frac{1}{5}\sum_{k=1}^{5} J_{t-k}^+$, $J_t^{m+} = \frac{1}{22}\sum_{k=1}^{22} J_{t-k}^+$; $J_t^{d-} = J_{t-1}^-$, $J_t^{w-} = \frac{1}{5}\sum_{k=1}^{5} J_{t-k}^-$, $J_t^{m-} = \frac{1}{22}\sum_{k=1}^{22} J_{t-k}^-$; $C_t^{d+} = C_{t-1}^+$, $C_t^{w+} = \frac{1}{5}\sum_{k=1}^{5} C_{t-k}^+$, $C_t^{m+} = \frac{1}{22}\sum_{k=1}^{22} C_{t-k}^+$; $C_t^{d-} = C_{t-1}^-$, $C_t^{w-} = \frac{1}{5}\sum_{k=1}^{5} C_{t-k}^-$, $C_t^{m-} = \frac{1}{22}\sum_{k=1}^{22} C_{t-k}^-$, where the upper index d represents the daily component, w the weekly component, and m the monthly component. Again, when taking a log transformation for the jump component, unity is added.In terms of our mnemonical variables, the specification of the HAR-GB-CJ model reads

$$rv_t = a_0 + a_d^+ c_t^{d+} + a_d^- c_t^{d-} + a_w^+ c_t^{w+} + a_w^- c_t^{w-} + a_m^+ c_t^{m+} + a_m^- c_t^{m-}$$

$$+ b_d^+ j_t^{d+} + b_d^- j_t^{d-} + b_w^+ j_t^{w+} + b_w^- j_t^{w-} + b_m^+ j_t^{m+} + b_m^- j_t^{m-} + \epsilon_t.$$

The basic HAR-RV model and its extensions described above are the most commonly used HAR-RV type models. We use them to show that the results of adding trading volume do not depend on the exact type of model; so that the conclusions are robust to a specification of the HAR-RV type model.

In addition, as noted above, the HAR-RV type models are commonly used to predict realized volatility at higher than daily horizons, in particular, for forecasting monthly and weekly volatility. We also use these models to predict the monthly and daily components of volatility, by using the same specifications with the left side variable changed to a weekly or monthly average realized volatility appropriately dated: $rv^w_{t+5} = \ln(RV^w_{t+5})$ for weekly and $rv^m_{t+22} = \ln(RV^m_{t+22})$ for month-ahead forecasting are used in place of $rv_t = \ln(RV_t)$.

The trading volume predictors are included in the log-transformed form as well. Define

$$v^p_t = \ln(TR^p_t).$$

Respectively, we sometimes add "-V" for 'volume' to model acronyms in addition to "-RV" for 'realized volatility'. In the table below we summarize the predictive regressions with added trading volume predictors.

model name	predictive part of specification (apart from intercept)
HAR	$a_d rv^d_t + a_w rv^w_t + a_m rv^m_t + \sum_{p\in\Omega} a^p_v v^p_t$
HAR-GB	$a^+_d rv^{d+}_t + a^-_d rv^{d-}_t + a^+_w rv^{w+}_t + a^-_w rv^{w-}_t + a^+_m rv^{m+}_t + a^-_m rv^{m-}_t + \sum_{p\in\Omega} a^p_v v^p_t$
HAR-CJ	$a_d c^d_t + a_w c^w_t + a_m c^m_t + b_d j^d_t + b_w j^w_t + b_m j^m_t + \sum_{p\in\Omega} a^p_v v^p_t$
HAR-GB-CJ	$a^+_d c^{d+}_t + a^-_d c^{d-}_t + a^+_w c^{w+}_t + a^-_w c^{w-}_t + a^+_m c^{m+}_t + a^-_m c^{m-}_t + b^+_d j^{d+}_t + b^-_d j^{d-}_t$ $+ b^+_w j^{w+}_t + b^-_w j^{w-}_t + b^+_m j^{m+}_t + b^-_m j^{m-}_t + \sum_{p\in\Omega} a^p_v v^p_t$

4 Evaluation methodology

A rather informal statistic we use to judge forecasting quality is the out-of-sample R^2 (e.g., Campbell & Thompson, 2008), which produces point estimates of quality of volatility forecasts and helps compare different predictive models. The out-of-sample R^2 measures how well, in terms of the mean squared forecast error (MSFE), the model predicts the data relative to a simple benchmark forecast of realized volatility. As a benchmark forecast, we use the previous day value of volatility; this forecast delivers smaller MSFE than, for example, the historical average, so in a sense our R^2 figures are conservative measures of MSFE-based forecasting quality, and will likely increase if one uses another benchmark. If for some model the out-of-sample R^2 is higher than for other models, such a model in interpreted as forecasting volatility better in terms of MSFE. The out-of-sample R^2 is computed from the following formula:

$$R^2 = 100\% \times \left(1 - \frac{\sum_{t=N/2}^{N} \left(rv_t - \widehat{rv}_t\right)^2}{\sum_{t=N/2}^{N} \left(rv_t - rv_{t-1}\right)^2}\right),$$

where rv_t and rv_{t-1} is the logarithm of (monthly, daily or weekly) realized volatility for days $t-1$ and t, respectively, and $\widehat{rv}_t$ is a forecast of the logarithm of realized volatility for day t.

To formally determine whether the past average daily, weekly and monthly trading volume enhance the quality of volatility forecasts relative to baseline models without volume or only including a lag of daily trading volume, we employ the model confidence set (MCS) approach by Hansen, Lunde & Nason (2011). We use two most commonly used MSE and QLIKE loss functions:

$$\text{MSE: } L\left(\widehat{rv}, rv\right) = \left(\widehat{rv} - rv\right)^2,$$

$$\text{QLIKE: } L\left(\widehat{rv}, rv\right) = -\ln\left(\frac{\widehat{rv}}{rv}\right) + \frac{\widehat{rv}}{rv},$$

where rv is the logarithm of realized volatility from the data, and $\widehat{rv}$ is the predicted logarithm of realized volatility. Both losses are robust to volatility measurement in the sense of Patton (2011), and we use both to verify that the results are robust to a choice of loss functions.

To obtain losses, we estimate each model[4] on the first $\frac{1}{2}$ of data sample (i.e., from $t = 1$ to

[4] We estimate the predictive models by OLS, which minimizes both MSE and QLIKE criteria in our case.

$t = N/2 - 1$) and produce out-of-sample forecasts (i.e., for $t = N/2$ to $t = N$), and compute the forecast error and loss values. We used the recursive estimation scheme, i.e. for each day $t = N/2, ..., N$, we estimate the model using all observations preceding day t and predict realized volatility for day t, and week and month that start on day t.

Now we briefly describe the construction of an MCS. Denote by $\mathcal{M}$ the initial set of competing models with different inclusions of average trading volume, which in our case is 4. Define the loss differential between models m and μ as

$$d_{m,\mu} = L_m - L_\mu,$$

where L_m is the loss of model m and L_μ is the loss of model μ. Then, one computes a t statistic for each loss differential:

$$t_{m,\mu} = \frac{d_{m,\mu}}{\sqrt{\widehat{\text{var}}\left(d_{m,\mu}\right)}} \text{ for all } m, \mu \in \mathcal{M},$$

where $\widehat{\text{var}}$ is a consistent estimate of the loss differential's (asymptotic) variance. The MCS test statistic is given by $\max_{m,\mu\in\mathcal{M}} |t_{m,\mu}|$, The null is that all the models in $\mathcal{M}$ have the same expected loss; under the alternative, there is some model that has an expected loss greater than expected losses of all the other models. If the null hypothesis is rejected, the worst performing model is eliminated. The test is performed iteratively, until no further model can be eliminated. We denote the final set of surviving models by $\mathcal{M}_*$. This final set contains the best forecasting model with the confidence level set in advance. The asymptotic distribution of the max-t test statistic is approximated by bootstrapping.[5]

The standard errors are constructed using the Newey & West (1987) heteroskedasticity and autocorrelation consistent (HAC) estimator, with truncation lag selected according to formula $4(n/100)^{1/3}$ (Newey & West, 1994), where n is the number of observations for a particular asset. On average the number of autocovariances lags is 15. Jump component coefficients, when reported, are multiplied by 10^{-5}, as they are relatively high due to low values of jump components.

[5]The bootstrap block length is set to a square root of a number of loss observations over which the loss is computed, i.e. $\sqrt{N/2}$.

5 Data

We use data on prices and trading volumes for the 45 stocks that were included in the past or are included now in the DOW30 index. The DOW30 is the index that includes 30 biggest U.S. companies in all industries. These companies' stocks are very liquid and highly diversified across industries. The data for daily trading volume is taken from ßYahoo! Finance. All the data on prices is taken from kibot.com.[6] Note that there are published papers analyzing realized volatility that use data from this source (for example, Caporin, 2017, 2022).

For each stock, if for some year more than 1% of observations for regular trading sessions are missing, that year's observations and observations for all previous years are excluded. If the data for some stock shares contain fewer than 2000 daily observations, this stock is excluded from the analysis. As a result, four stocks are excluded.[7] The data have at least 3370 daily observations, and on average 5438 daily observations. The data for different stocks start from different dates. The earliest year is 1998, and the latest year is 2008. Almost all stocks' data end in June 2022.

The raw price data contain prices for each minute of pre-market (8:00-9:30 a.m.), regular (9:30-4:00 p.m.), and aftermarket (4:00-6:30 p.m.) sessions on each of the trading days for each stock. Since realized volatility and its jump component are used, the high volatility of pre-market and after-market sessions may distort volatility measures. Thus, we focus on regular session data, so the data only for transactions from 9:30 a.m. to 4:00 p.m. are used. The price of each stock is a midquote, i.e. an arithmetic average between the open bid and open ask price for each five minutes. We use this price to calculate intradaily returns for each 5 minutes.

[6] The data on prices are available on the payment basis at this kibot.com link.

[7] We used stocks for the following companies: Alcoa Corporation, Apple Inc., American International Group Inc., American Express Company, Boeing Company, Bank of America Corporation, Citigroup Inc., Caterpillar Inc., Salesforce Inc., Cisco Systems Inc., Chevron Corporation, Walt Disney Company, General Electric Company, General Motors Company, Goldman Sachs Group Inc., Home Depot Inc., Honeywell International Inc., HP Inc., International Business Machines Corporation, Intel Corporation, Johnson & Johnson, JP Morgan Chase & Co., Coca-Cola Company, McDonald's Corporation, Mondelez International Inc., 3M Company, Altria Group Inc., Merck & Company Inc., Microsoft Corporation, Nike Inc., Pfizer Inc., Procter & Gamble Company, Raytheon Technologies Corporation, AT&T Inc., The Travelers Companies Inc., United Health Group Incorporated, Visa Inc., Verizon Communications Inc., Walgreens Boots Alliance Inc., Walmart Inc., Exxon Mobil Corporation.

Note that we used 5 minute returns. There is a trade-off between using data for higher or lower sampling frequency. The higher the sampling frequency is and thus the larger the sample size of intradaily returns is, the more precise the estimates of daily volatility should become. However, the higher the frequency is, the larger is the effect of microstructure noise that distorts the aggregated realized volatility. According to Bandi & Russell (2006), on average the four minute sampling frequency is enough for liquid stocks[8] to ensure that microstructure noise does not distort realized volatility too much. Since the most frequently used in the literature for stocks of sufficient liquidity is five minute sampling frequency, we too use this exact frequency.

6 Results

6.1 Regression coefficients

First we discuss numerical values, signs, and statistical significance of coefficient estimates corresponding to volume-related predictors. We report aggregated regression results for the basic HAR-RV model specification in Table 1; the results for the extended specifications and assets are similar and reported in similar Tables 4–6 in the Appendix. A figure without brackets is an arithmetical average across point estimates of the corresponding coefficient for all 41 assets, while a figure in square brackets shows a fraction of coefficient estimates that are statistically significant at the 90% significance level. Several important conclusions can be derived from these figures.

[8]The results in Bandi & Russell (2006) are obtained for each stock from the S&P100 index.

Table 1: Aggregated regression results for basic HAR-RV-V model.

predictor	day-ahead forecast				week-ahead forecast				month-ahead forecast			
rv_t^d	0.39	0.38	0.32	0.32	0.27	0.28	0.23	0.23	0.18	0.20	0.17	0.17
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]
rv_t^w	0.30	0.30	0.38	0.37	0.30	0.30	0.36	0.37	0.25	0.25	0.29	0.32
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[98]	[100]
rv_t^m	0.25	0.25	0.24	0.26	0.33	0.32	0.32	0.3	0.39	0.37	0.37	0.34
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[95]
v_t^d		0.04	0.23	0.23		-0.02	0.11	0.11		-0.06	0.03	0.03
		[66]	[100]	[100]		[56]	[98]	[95]		[54]	[37]	[27]
v_t^w			-0.24	-0.20			-0.17	-0.20			-0.11	-0.18
			[100]	[98]			[98]	[85]			[63]	[71]
v_t^m				-0.05				0.03				0.08
				[37]				[22]				[34]

Parameters are estimated by OLS; standard errors are constructed using Newey & West (1987) HAC estimator, with truncation lag selected according to Newey & West (1994).

First, the lags of average realized volatility are always significant at the 10% level. This result is consistent with the intuition behind the HAR-RV model. Second, the coefficient on lagged average realized volatility decreases when the trading volume is added. The reason is that trading volume and volatility are both proxies for the information flow that comes into the market, so the influence of information in RV lagged values decreases when the lagged trading volume is added and hence the corresponding coefficient goes down. Third, for the prediction of daily volatility, in most cases, coefficients on trading volumes for the lagged daily and monthly average volumes are statistically significant. This result is consistent with and is justified by intuition, which we outlines in the Introduction.

Fourth, the coefficient of the lagged daily trading volume is usually positive. Trading volume is a proxy for information that comes to the market; thus, an increase in trading volume implies that some information has appeared in the market. This information will be able to change prices in the future, and hence volatility will increase in the future. The coefficient of the lagged average weekly trading volume is negative and relatively big in absolute value. The intuition behind this evidence is that the average trading volume is a proxy that describes how fast information spreads in the market. If average weekly trading volume is low, information spreads very fast because the prices adjust very fast, and the traders do not want to trade intensively because of the changed price. For the same reason, if trading volume is high for the past week, information spreads slowly. The lagged daily volume carries information about the information that comes to market, not about its propagation speed. Thus, the lagged average weekly trading volume predicts volatility with a negative sign because it measures the speed of information propagation in the market.

Finally, in the models where a lagged average weekly volume is present, the coefficients for the lagged daily volume are always positive and almost always statistically significant. These coefficients are much higher than the same coefficients in the models with only the lagged daily trading volume accompanied with a lower fraction of statistical significance. These results are consistent with the intuition described in the previous paragraph: in the model without lagged average weekly volume, the daily volume measures information that comes to the market and hence increases volatility. But at the same time, since there is no weekly volume predictor, the lagged daily volume predictor measures the information propagation speed and thus decreases volatility. Thus, the average weekly volume affects volatility through two different channels: through one channel, it has a positive effect, while through the other channel it decreases volatility. As a net effect, this decreases the values of coefficients and makes them less statistically significant. When average weekly volume that represents the propagation speed of information is added to the model, the lagged daily volume then represents only information that comes into the market, and as a result, affects today's volatility only through one channel. So the coefficient on the lagged daily volume becomes highly positive and significant.

6.2 Model confidence sets

Now we present the results of the model confidence set procedures. Table 2 reports, for each model, forecast horizon and criterion, a number of assets (out of all 41), for which adding average weekly volume predictor, or average weekly and monthly volume predictors together, produce a 90% level statistically significant improvement in forecasting quality, or, conversely, its worsening. Formally, each figure in the upper panel points at a number of assets, for which the model without trading volume or with only lagged daily trading volume is not included in the 90% level MCS $\mathcal{M}_*$, the set of survived models from the MCS procedure, while one of the volume-augmented model lie inside $\mathcal{M}_*$. Similarly, in the lower panel, each figure points at a number of assets, for which the baseline model lies inside $\mathcal{M}_*$, while one of volume-augmented models lies inside $\mathcal{M}_*$. Several important conclusions can be derived from these figures.

Table 2: Number of assets (out of 41) with MCS significance level of 90%.

forecast horizon	predictive model and criterion							
	MSE				$QLIKE$			
	HAR	HAR-GB	HAR-CJ	HAR-GB-CJ	HAR	HAR-GB	HAR-CJ	HAR-GB-CJ
	increase of forecasting quality							
day	34	36	36	36	32	33	32	35
week	15	16	15	16	15	16	14	15
month	10	10	11	11	10	11	11	11
	decrease of forecasting quality							
day	0	0	0	0	0	0	0	0
week	0	0	0	0	0	0	0	0
month	3	3	3	3	3	3	3	3

First, one can see that at the daily horizon, the models containing average weekly volume predictor or average weekly and monthly volume predictors together typically forecast volatility better and never forecast worse than the baseline models. This result does not significantly depend on a model specification or criterion. At least for 32 assets out of 41 the forecasting quality improves at the 90% significance level, while on average this happens in 34.25 out of 41 cases. There are two possible explanations of this result.

The first is that there are two channels of how past average trading volume can increase the quality of prediction. It adds new information about shocks that come to the market, and it is a proxy that describes how fast information spreads in the market. Through these two channels, past volume affect volatility differently. Through the channel of new information, the average trading volume should increase volatility, while through the channel of the information propagation speed, the average trading volume should decrease volatility. When past volume is not included, this information is missing. When only the daily lag is included, it predicts volatility via both channels – through one, it increases volatility, while through the other it decreases volatility. The lagged daily volume predictor thus shows a net effect of these two opposite channels, and thus may not significantly enhance forecasting quality. However, when one adds average weekly trading volume, which represents only the information propagation speed channel, the lagged daily volume represents only the new information channel. Thus, each predictor affects volatility only through one specific channel, with own coefficient exhibiting that channel's impact, and the forecasting quality increases as a consequence of this division of labor.

The second explanation is similar to the intuition behind the HAR-RV model. It is based on the Heterogeneous Market Hypothesis, which states that traders in financial markets are heterogeneous. The agents with different time horizons perceive and react to different news (for which trading volume is a proxy) and thus cause trading volume differently for different periods. There are differences in the speed of arrival of information to the market for different types of traders as well. In summary, trading volume when averaged over different periods has different volatility predictive ability. At the same time, trading volume represents different information than realized volatility does. Moreover, if a volatility measure is imprecise (for example, due to low data quality), a trading volume averaged over the same past periods as volatility is able to complement past volatility predictors and so enhance volatility forecasting quality. This intuition is supported by the fact that for volatility prediction for some specific horizon, the most important (in terms of a numerical value of corresponding coefficient) is the length of averaging period that coincides with the horizon. Yet another fact that supports this intuition is that when one adds lagged trading volume, the coefficients for past volatility decrease. That is, the information that lagged volume represents coincides, at least partially, with information that the same-lag of realized volatility represents.

Note that since the lags of trading volume increase the quality of volatility predictions, we have evidence that the SIAH holds for the assets under consideration, even if the forecasting quality increases only for the daily horizon. The reasoning is given in Section 2.

Second, an increase in weekly horizon forecasting quality from adding the lagged average weekly or weekly and monthly volumes is heterogeneous across assets. For 14–16 out of 41 assets, both the weekly and weekly and monthly volume-augmented models forecast volatility better than both the baseline models, at the 90% significance level. This result almost does not depend on a model specification or criterion. Moreover, there is no asset for which adding these lagged predictors worsens the forecast. The coefficient on the daily volume lag when the weekly volume lag is added in the weekly horizon model is on average more than two times lower than the same coefficient in the daily horizon model. After adding the lagged weekly volume to the model, the lagged daily volume predictor represents information that comes to the market, and so, for weekly forecasts, this information does not affect volatility as much as it does for daily forecasts. Intuitively, information that comes to the market spreads out fast enough to not significantly affect volatility for the following week, but it spreads out slowly enough to affect volatility the following day. Thus, an increase in forecasting quality from adding the lagged weekly volume is lower for weekly horizon forecasts.

Finally, at the monthly horizon, adding weekly or weekly and monthly average volume predictors to the two baseline models rarely increases and rarely decreases the volatility forecasting quality at the 90% significance level. This result holds for both criteria and all the predictive models. There are two possible explanations for this result.

The first explanation is the same as described before, related to the split of information and division of labor among the lagged daily volume and averages over longer periods. The value of the coefficient corresponding to the lagged daily volume predictor when the lagged average weekly volume is added in the month-ahead forecasting model is on average close to zero and statistically insignificant at the 90% level. This means that information that comes to the market represented by the past daily volume, does not affect volatility as much as it does for day-ahead or week-ahead forecasting. Thus, when the lagged average weekly volume is added, the forecasting quality of prediction does not increase significantly.

The second explanation is that sometimes, trading volume may not give any information about too distant future monthly volatility and hence does not improve volatility forecasts whatsoever. This intuition is supported by the fact that the coefficient on almost all lags of average trading volume is always closer to zero and less statistically significant compared to daily and weekly volatility forecasting models.

Next, we present the results of the MCS procedure that compares only two volume-augmented models – one with past average volumes for periods of all three lengths (daily, weekly and monthly), and with past average volumes for periods of only two shortest lengths (daily and weekly). Table 3 reports, for each model, forecast horizon and criterion, a number of assets (out of all 41), for which one of these two models produces a significant, at the 90% significance level, improvement in forecasting quality relative to the other model. Formally, these figures show a number of assets, for which the given model is not included in the final set of surviving models, $\mathcal{M}_*$. The purpose of this evidence is to understand which of the two models – with the lagged monthly volume or without it, produces the highest increase in forecasting quality, on top of inclusion of all smaller-length averaging periods (as a significant increase in forecasting quality is not found at the monthly horizon, this horizon is omitted from the analysis). For daily and weekly horizon volatility predictions, very rarely either model outperforms the other. In most cases, the MCS technology cannot distinguish these two models, at least (or even) at the 90% significance level. This result holds for all model specifications and both criteria.

Table 3: Number of assets (out of 41) with MCS significance level of 90%.

forecast horizon	predictive model and criterion							
	MSE				*QLIKE*			
	HAR	HAR-GB	HAR-CJ	HAR-GB-CJ	HAR	HAR-GB	HAR-CJ	HAR-GB-CJ
	daily and weekly volume predict better than daily, weekly and monthly volume							
day	3	5	5	4	2	4	2	3
week	2	0	1	0	1	1	1	1
	daily, weekly and monthly volume predict better than daily and weekly volume							
day	3	1	2	3	0	1	0	0
week	1	0	0	1	1	1	0	1

This result gives support to both intuitive explanations of improvement in forecasting quality from adding the lagged average weekly or weekly and monthly trading volumes. According to the first intuitive explanation, the forecasting quality increases from adding the average weekly volume because it allows to distinguish two effects of trading volume on volatility: new information about shocks that come to the market, and how fast information spreads in the market. This suggests that adding past average weekly volume increases forecasting quality. The second intuitive explanation is similar to the intuition behind the HAR-RV model and suggests that adding both average weekly and monthly volumes should increase forecasting quality more than solely average weekly volume. Since both models with past volumes for periods of only two shortest lengths, on the one hand, and past volumes for periods of all three lengths, on the other, increase forecasting quality and the MCS cannot distinguish between these two models, both intuitive explanations are valid – otherwise, if only one was true, one of these models would outperformed the other: in case of the former explanation it would be the model with past average weekly and monthly volumes, and in case of the latter explanation, it would be the model with past daily, average weekly and average monthly trading volume predictors.

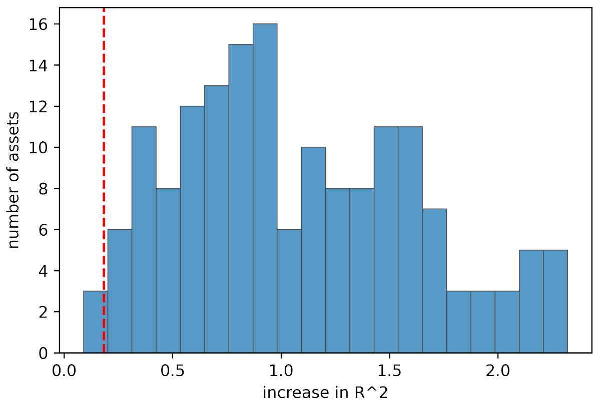
6.3 Quality of out-of-sample fit

Now we discuss increases in the out-of-sample R^2 measure, whose values are reported in Figure 1. The figure shows a distribution of a rise in R^2 from adding the lagged average weekly or weekly and monthly trading volumes to one of the baseline models with the same forecasting horizon. Formally, the variable whose histogram is depicted is a difference in R^2 between, on the one hand, the better model between one including daily, weekly and monthly volumes and one including daily and weekly volumes, and on the other hand, the better model between one including only daily volume and one not including any volume predictor.[9] This R^2 differential is yet another measure showing whether and by how much adding past weekly or monthly trading volume can increase the forecasting quality and providing insight about economic significance of this increase. The red line represents an average increase from adding the past daily trading volume to the model without volume at all.

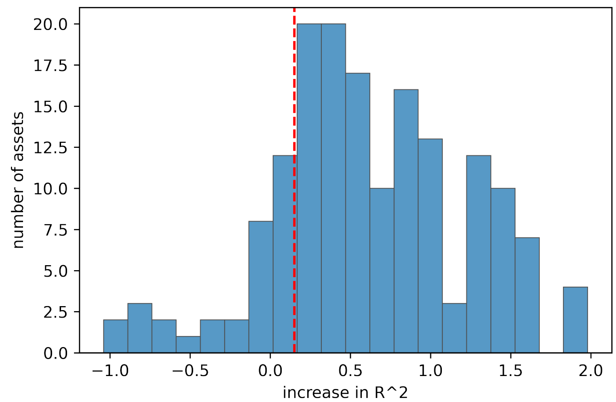
[9] Schematically, this can be written as $\max\{R^2_{d,w,m}, R^2_{d,w}\} - \max\{R^2_d, R^2_\varnothing\}$, with self-explanatory notation.

Figure 1: R^2 differential pooled across models for different forecast horizons

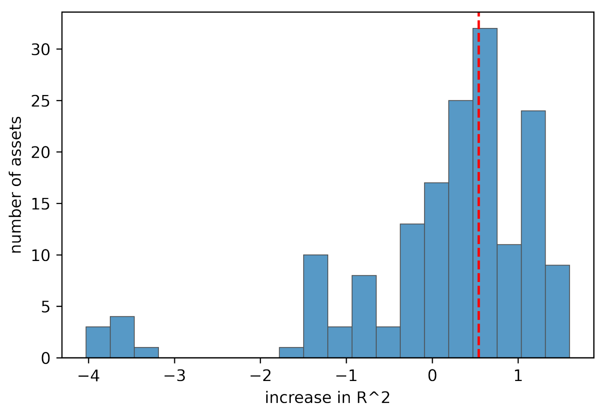
(a) day-ahead forecasting



(b) week-ahead forecasting



(c) month-ahead forecasting



Note that in Figure 1 the out-of-sample R^2 differentials are pooled across all four HAR-RV specification. Figure 2 in the Appendix reports similar R^2 differentials for each HAR-RV specification separately, for the daily forecast horizon. We find that there is no significant difference in the distribution of the R^2 differential across different models; the results for the other forecast horizons are similar.

One can see that an increase in R^2 is usually consistent with results from MCS. For day-ahead forecasts, the out-of-sample R^2 always increases as a result of adding longer-period volume predictors. For week-ahead forecasts, the out-of-sample R^2 usually increases and rarely decreases. On average, such an increase is lower than in case of day-ahead forecasting. For month-ahead forecasts, adding these volume predictors sometimes increases and sometimes decreases the out-of-sample R^2. The increase is on average lower than for the day or week forecast horizon. On average, there is no improvement in the quality of volatility forecasting for the monthly horizon.

All these results are consistent with those produced by the MCS procedure, but give an idea of the size of forecast improvements, which are pretty impressive, and some insight into economic significance of how much past average trading volume can improve forecasting quality on the daily or weekly scale. First, on average, adding longer-period volume predictors to one of the baseline models increases the out-of-sample R^2 by 1.1% for one-day horizon and by 0.62% for week-ahead horizon, which corresponds to a relative increase of 5.0% and 2.7%, respectively, given that the average R^2 is 22.4%. Second, on average, adding past daily volume to the model without volume predictors increases the out-of-sample R^2 by 0.25% for day-ahead and week-ahead forecasting horizons. Thus, on average, forecasting quality from adding only past daily volume (as usually done in the literature to improve prediction quality) is more than four times lower for day-ahead forecasting and more than two times lower for week-ahead forecasting than from adding past average weekly volume or weekly and monthly volumes on top of that.

Third, on average across assets, the out-of-sample R^2 differential between the best and the worst HAR-RV specifications[10] is 0.37% for one-day and 0.6% for one-week forecasting. Thus, on average, an increase in forecasting quality from adding past average weekly volume or weekly and monthly volumes is three times higher for day-ahead forecasting and similar for week-ahead forecasting than the difference between the best and the worst performing HAR-RV specifications.

To summarize, since for a day-ahead prediction the difference in the out-of-sample R^2 between the best and the worst HAR-RV specifications is three times lower, and the difference between the model with no volume and one with past daily volume is four times lower than an increase in the out-of-sample R^2 from adding past average weekly volume or weekly and monthly volumes, an economically highly significant amount.

[10] This differential is computed as $\frac{1}{41}\sum_{a=1}^{41}(\max\{R^2_{a,\text{HAR-RV}}, R^2_{a,\text{HAR-RV-CJ}}, R^2_{a,\text{HAR-RV-GB}}, R^2_{a,\text{HAR-RV-GB-CJ}}\} - \min\{R^2_{a,\text{HAR-RV}}, R^2_{a,\text{HAR-RV-CJ}}, R^2_{a,\text{HAR-RV-GB}}, R^2_{a,\text{HAR-RV-GB-CJ}}\})$, where $R^2_{i,s}$ is the value of out-of-sample R^2 for asset a and specification s.

7 Conclusion

Within the framework of HAR-RV type models we have investigated whether utilizing, in the spirit of HAR modeling, past average trading volume for long periods – week and/or month – are able to improve the quality of volatility forecasting. We compare models with past average volume for a day, week and month and models with past average volume for a day and week, on the one hand, and the model without volume and the model with past daily trading volume only. We employ an out-of-sample R^2 and the model confidence set procedure driven by MSE or QLIKE loss functions. We find that typically, for the one-day forecasting horizon, including past average week volume or average week and month volumes improves, statistically and economically significantly, and never worsens, forecasting quality relative to the benchmark models. The improvements are weaker at the one-week forecasting horizon, and rarely, if ever, appear at the one-month horizon. We have suggested two explanations of these results, one based on different flows of information coming to the market in periods of different length, and the other based on the division of labor among different-length-periods average volumes in containing different types of information – shocks coming to the market and how fast information spreads out in the market. All our results seem to be robust to HAR-RV type model specification, a choice of loss function, and asset type.

References

[1] Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. Review of Economics and Statistics, 89(4), 701–720.

[2] Andersen, T. G., Bollerslev, T. & Dobrev, D. (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. Journal of Econometrics, 138(1), 125–180.

[3] Bandi, F. M. & Russell, J. R. (2006). Separating microstructure noise from volatility. Journal of Financial Economics, 79(3), 655–692.

[4] Boudt, K., Croux, C. & Laurent, S. (2011). Robust estimation of intraweek periodicity in volatility and jump detection. Journal of Empirical Finance, 18(2), 353–367.

[5] Brooks, C. (1998). Predicting stock index volatility: Can market volume help? Journal of Forecasting, 17, 59–80.

[6] Campbell, J. Y. & Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Review of Financial Studies, 21(4), 1509–1531.

[7] Caporin, M. (2022). The Role of Jumps in Realized Volatility Modeling and Forecasting. Journal of Financial Econometrics, published online.

[8] Caporin, M., Kolokolov, A. & Reno, R. (2017). Systemic co-jumps. Journal of Financial Economics, 126(3), 563–591.

[9] Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. Journal of Financial Econometrics, 7(2), 174–196.

[10] Corsi, F., Pirino, D. & Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics, 159(2), 276–288.

[11] Desboulets, L. (2018). A Review on Variable Selection in Regression Analysis. Econometrics, 6(4), 45.

[12] Duong, D. & Swanson, N. R. (2015). Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. Journal of Econometrics, 187(2), 606–621.

[13] Hansen, P. R., Lunde, A. & Nason, J. M. (2011). The model confidence set. Econometrica, 79(2), 453–497.

[14] Kambouroudis, D. S. & McMillan, D. G. (2016). Does VIX or volume improve GARCH volatility forecasts? Applied Economics, 48(13), 1210–1228.

[15] Liu, M., Lee, C. & Choo, W. (2020). An empirical study on the role of trading volume and data frequency in volatility forecasting. Journal of Forecasting, 40(5), 792–816.

[16] 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.

[17] Mougoue, M. & Aggarwal, R. (2011). Trading volume and exchange rate volatility: Evidence for the sequential arrival of information hypothesis. Journal of Banking & Finance, 35(10), 2690–2703.

[18] Newey, W. K. & West, K. D. (1987). A simple, positive semi-de.nite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55, 703–708.

[19] Newey, W. K. & West, K. D. (1994). Automatic Lag Selection in Covariance Matrix Estimation. The Review of Economic Studies, 61(4), 631–653.

[20] Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics, 160(1), 246–256.

[21] Patton, A. J. & Sheppard, K. (2015). Good Volatility, Bad Volatility: Signed Jumps and The Persistence of Volatility. Review of Economics and Statistics, 97(3), 683–697.

[22] Wang, X., Wu, C. & Xu, W. (2015). Volatility forecasting: The role of lunch-break returns, overnight returns, trading volume and leverage effects. International Journal of Forecasting, 31(3), 609–619.

Appendix

[See the following pages...]

Table 4: Aggregated regression results for HAR-GB-V model.

predictor	daily forecast				weekly forecast				monthly forecast			
rv_t^{d+}	0.15	0.15	0.11	0.11	0.10	0.11	0.08	0.08	0.07	0.08	0.06	0.06
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]
rv_t^{w+}	0.09	0.09	0.14	0.13	0.08	0.08	0.12	0.12	0.05	0.05	0.07	0.09
	[76]	[73]	[93]	[90]	[49]	[54]	[68]	[68]	[15]	[20]	[34]	[44]
rv_t^{m+}	0.10	0.10	0.10	0.11	0.14	0.14	0.14	0.13	0.26	0.25	0.25	0.23
	[46]	[49]	[46]	[49]	[44]	[41]	[41]	[39]	[54]	[51]	[51]	[49]
rv_t^{d-}	0.23	0.23	0.19	0.19	0.16	0.16	0.14	0.14	0.11	0.12	0.10	0.10
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]
rv_t^{w-}	0.22	0.22	0.26	0.25	0.22	0.22	0.25	0.26	0.20	0.20	0.22	0.24
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]
rv_t^{m-}	0.15	0.15	0.14	0.15	0.19	0.18	0.18	0.17	0.13	0.12	0.12	0.10
	[63]	[63]	[61]	[61]	[51]	[51]	[51]	[49]	[27]	[27]	[27]	[29]
v_t^d		0.04	0.24	0.24		-0.01	0.12	0.12		-0.06	0.04	0.03
		[68]	[100]	[100]		[54]	[98]	[98]		[56]	[56]	[51]
v_t^w			-0.26	-0.21			-0.18	-0.20			-0.12	-0.19
			[100]	[98]			[98]	[85]			[71]	[73]
v_t^m				-0.06				0.03				0.08
				[41]				[22]				[34]

Parameters are estimated by OLS; standard errors are constructed using Newey & West (1987) HAC estimator, with truncation lag selected according to Newey & West (1994).

Table 5: Aggregated regression results for HAR-CJ-V model.

predictor	daily forecast				weekly forecast				monthly forecast			
c_t^d	0.39	0.38	0.31	0.31	0.27	0.27	0.23	0.23	0.18	0.19	0.17	0.17
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]
c_t^w	0.29	0.29	0.37	0.36	0.29	0.29	0.35	0.36	0.24	0.24	0.28	0.31
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]
c_t^m	0.26	0.27	0.26	0.28	0.35	0.34	0.33	0.32	0.41	0.39	0.38	0.35
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]
j_t^d	1.20	1.17	0.91	0.93	-0.14	-0.07	-0.25	-0.24	-0.58	-0.40	-0.54	-0.52
	[27]	[27]	[20]	[20]	[22]	[22]	[22]	[22]	[10]	[12]	[10]	[10]
j_t^w	-4.88	-4.93	-4.58	-4.72	-5.61	-5.41	-5.14	-5.13	-9.98	-9.49	-9.24	-9.13
	[39]	[37]	[32]	[34]	[34]	[34]	[34]	[34]	[59]	[59]	[61]	[56]
j_t^m	-6.86	-6.78	-6.33	-6.35	-8.26	-7.60	-7.21	-7.0	-5.35	-4.16	-3.81	-3.52
	[29]	[32]	[29]	[29]	[29]	[27]	[29]	[27]	[22]	[24]	[27]	[22]
v_t^d		0.04	0.23	0.23		-0.02	0.11	0.11		-0.06	0.03	0.03
		[71]	[100]	[100]		[54]	[98]	[98]		[61]	[29]	[27]
v_t^w			-0.24	-0.20			-0.17	-0.19			-0.12	-0.18
			[100]	[98]			[98]	[83]			[68]	[68]
v_t^m				-0.06				0.03				0.08
				[41]				[20]				[34]

Parameters are estimated by OLS; standard errors are constructed using Newey & West (1987) HAC estimator, with truncation lag selected according to Newey & West (1994). Jump component coefficients are multiplied by 10^{-5}.

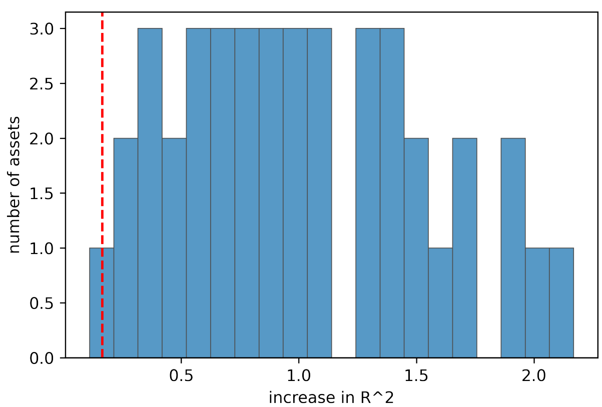
Table 6: Aggregated regression results for HAR-GB-CJ-V model.

predictor	daily forecast				weekly forecast				monthly forecast			
c_t^{d+}	0.15	0.14	0.11	0.11	0.10	0.10	0.08	0.08	0.07	0.07	0.06	0.06
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]
c_t^{w+}	0.09	0.08	0.13	0.12	0.08	0.08	0.11	0.11	0.04	0.05	0.07	0.08
	[71]	[66]	[90]	[85]	[41]	[39]	[63]	[63]	[10]	[17]	[29]	[41]
c_t^{m+}	0.11	0.11	0.11	0.12	0.16	0.15	0.15	0.14	0.26	0.25	0.25	0.23
	[54]	[51]	[51]	[51]	[44]	[41]	[41]	[39]	[51]	[51]	[49]	[49]
j_t^{d+}	0.66	0.60	0.22	0.25	-0.46	-0.38	-0.64	-0.62	-1.09	-0.86	-1.06	-1.03
	[20]	[20]	[20]	[20]	[22]	[20]	[22]	[22]	[7]	[2]	[5]	[5]
j_t^{w+}	-4.20	-4.36	-3.94	-4.25	-4.87	-4.67	-4.35	-4.37	-12.23	-11.59	-11.16	-10.9
	[17]	[17]	[7]	[7]	[12]	[12]	[12]	[12]	[34]	[32]	[29]	[32]
j_t^{m+}	-5.00	-5.19	-4.77	-4.76	-4.55	-4.59	-4.24	-4.24	9.46	9.07	8.91	8.47
	[20]	[20]	[15]	[17]	[15]	[15]	[15]	[15]	[29]	[27]	[29]	[32]
c_t^{d-}	0.23	0.22	0.19	0.19	0.16	0.16	0.14	0.14	0.11	0.11	0.10	0.10
	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]	[100]
c_t^{w-}	0.21	0.21	0.26	0.25	0.22	0.22	0.25	0.25	0.20	0.20	0.22	0.23
	[100]	[100]	[100]	[100]	[98]	[98]	[100]	[100]	[98]	[98]	[98]	[98]
c_t^{m-}	0.15	0.15	0.15	0.16	0.19	0.19	0.19	0.18	0.14	0.14	0.13	0.12
	[66]	[68]	[63]	[66]	[56]	[51]	[54]	[49]	[27]	[29]	[29]	[29]
j_t^{d-}	2.10	2.06	1.84	1.84	0.23	0.29	0.12	0.14	-0.21	-0.07	-0.19	-0.16
	[32]	[32]	[24]	[22]	[20]	[22]	[22]	[20]	[7]	[7]	[7]	[5]
j_t^{w-}	-5.80	-5.85	-5.32	-5.40	-5.99	-5.77	-5.37	-5.33	-7.43	-6.95	-6.72	-6.56
	[24]	[27]	[24]	[24]	[20]	[20]	[20]	[20]	[27]	[22]	[24]	[24]
j_t^{m-}	-9.78	-9.71	-8.41	-8.34	-14.17	-12.51	-11.40	-11.08	-26.32	-22.87	-21.56	-21.0
	[22]	[22]	[24]	[27]	[17]	[15]	[15]	[12]	[24]	[20]	[20]	[20]
v_t^{d}		0.04	0.24	0.24		-0.01	0.12	0.12		-0.06	0.04	0.04
		[71]	[100]	[100]		[56]	[98]	[98]		[59]	[59]	[56]
v_t^{w}			-0.26	-0.21			-0.18	-0.19			-0.12	-0.18
			[100]	[98]			[98]	[88]			[73]	[68]
v_t^{m}				-0.06				0.02				0.07
				[41]				[17]				[29]

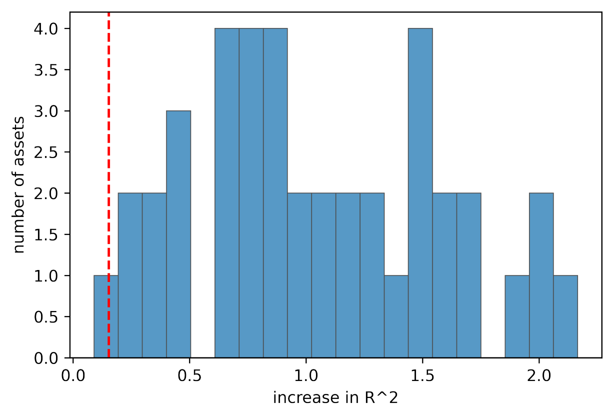
Parameters are estimated by OLS; standard errors are constructed using Newey & West (1987) HAC estimator, with truncation lag selected according to Newey & West (1994). Jump component coefficients are multiplied by 10^{-5}.

Figure 2: R^2 differential for different models for daily forecast horizon.

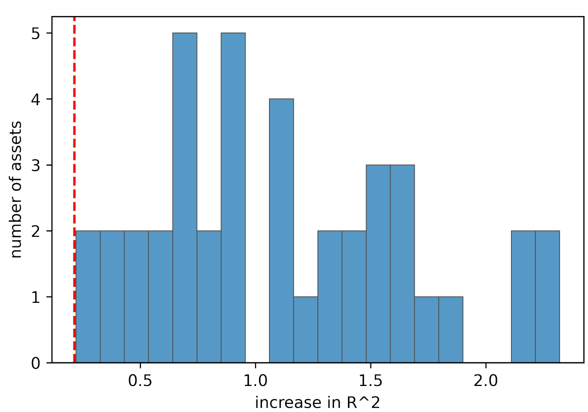
(a) HAR-RV-V



(b) HAR-RV-CJ-V



(c) HAR-RV-GB-V



(d) HAR-RV-GB-CJ-V