Developing Automated Trading Algorithm Using Volatility Forecasts of HARX Model

Ko Adachi*

Purvesh Vasant Shende†

Abstract

This paper develops an automated trading algorithm which utilizes the stock return volatility forecasts of the HARX (Heterogenous Autoregressive with Extra Inputs) model. The HARX model is an extension of the simple HAR model in three aspects: it i) has the higher lag order, ii) utilizes information on the volatility of overall stock market, and iii) considers asymmetry in volatility dynamics. The HARX model outperforms the simple HAR model in both in-sample and out-of-sample forecasting accuracy. Then we develop an automated stock options trading algorithm incorporating the HARX model's predictions using IB-Matlab. We finally run the algorithm on a paper trading account and carry out a preliminary assessment of its profitability.

Keywords: Stock return volatility; HAR model; HARX model; realized volatility; implied volatility; IB-Matlab; automated trading.

1 Introduction

Volatility is a core element in asset pricing. In particular, an option price is closely related to the volatility forecast of the underlying asset. In fact, in the financial industry, buying options is often referred to as buying volatility and selling options is often called selling volatility. Therefore, it is essential to accurately forecast the underlying asset volatility in order to make sustainable profits from options trading.

In academia, a number of stock return volatility forecasting models have been developed over the years. Among the older ones, the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle (1982) and the Generalized ARCH (GARCH) model by Bollerslev (1986) are well known. More recently, the models such as ARFIMA (Autoregressive Fractionally Integrated Moving Average), HAR (Heterogeneous Autoregressive), and UC (Unobserved Component) models have been found to be useful for volatility forecasting (Giot and Laurent [2004], Corsi [2009], etc.), because they can

* M.S. in Financial Economics Candidate. Email: kadachi@wisc.edu

† M.S. in Financial Economics Candidate. Email: pvshende@wisc.edu

successfully reproduce the time-series property of stock return volatility which is called long memory.

In this paper, the HARX (HAR with Extra Inputs) model, an extension of the simple HAR model proposed by Corsi (2009), is developed to forecast stock return volatility. One of the advantages to employ HAR-type models is their estimation simplicity as they have the similar structure as AR models, in spite of the fact that they can successfully depict the time-series characteristics of stock return volatility as mentioned above.

In order to improve forecasting accuracy, we add to the simple HAR model several factors considered useful in forecasting stock return volatility. Specifically, we firstly increase the lag order compared to the HAR model proposed by Corsi (2009). Corsi (2009) predicts asset return volatility from volatility components at three different frequencies: daily, weekly, and monthly. On the other hand, Fernandez et al. (2014), who analyzed a forecasting model for the VIX index, report that volatility components at bi-weekly and quarterly frequencies are also useful in forecasting the VIX index. In this paper, we follow their work and incorporate volatility components at bi-weekly and quarterly frequencies into the simple HAR model.

Second, our HARX model is extended to take advantage of information on overall stock market volatility. Hizmeri et al. (2022) incorporate daily, weekly, and monthly market realized volatilities to the simple HAR model and find statistically and economically significant improvement in predictive power. In addition, the VIX index is known to be extremely useful in forecasting stock return volatility (e.g., Blair et al. [2001], Martens and Zein [2004]). Based on these considerations, we use the lagged variables of the market return volatility and the VIX index.

Third, our model takes into account the asymmetry of volatility dynamics. Asymmetry in volatility dynamics refers to the tendency for stock prices to be more volatile on the day after a decline than on the day after an increase. This tendency has been empirically examined by Giot and Laurent (2004) and others, and they find the tendency is statistically significant. Therefore, we incorporate covariates that capture the volatility dynamics asymmetry into the model.

In estimation part, we find our HARX model estimated using the last 20 years of data outperforms an AR-type model and the Corsi's HAR model in terms of forecast accuracy, both in-sample and out-of-sample. This fact is robustly observed irrespective of the targeted stocks, time horizons, and the forecast accuracy measures.

Finally, we develop an automated stock option trading algorithm that incorporates volatility predictions from the HARX model. The algorithm is built using an application called IB-Matlab, which enables us to fetch financial data from Interactive Brokers (IB) and submit orders to IB automatically through MATLAB. In the algorithm, the HARX model's volatility prediction is compared with the implied volatility (IV) calculated from the corresponding option price, and if the latter exceeds the former, it submits a short sale order for the option to IB. Using the algorithm, we finally traded options on an Interactive Brokers' paper trading account for 5 days and carried out a preliminary assessment of its profitability.

The remainder of this paper is organized as follows. Section 2 describes the HAR model proposed by Corsi (2009) and our HARX model; Section 3 presents the estimation results of the models; Section 4 reports the out-of-sample forecast performance of the models; Section 5 details the automated trading algorithm and examines its profitability; Section 6 concludes.

2 Modeling stock return volatility

2.1 Heterogenous autoregressive model (HAR)

Corsi (2009) considers a financial market where some participants, such as dealers and speculators, trade at a high frequency, while others, such as institutional investors, trade only at a much lower frequency. He shows that, in this situation, the dynamics of stock return volatility depends on volatility components defined over different time horizons as a result of the fact that each market participant face or react to different volatility depending on his or her trading frequency.

More specifically, Corsi (2009) assumes that there are three types of traders in the stock market: short-term traders who trade on a daily or higher frequency, medium-term investors who rebalance their positions on a weekly basis, and long-term agents who rebalance on a monthly or multi-monthly basis. He then shows that under this assumption, the volatility of daily stock returns is affected by three lagged variables of daily, weekly, and monthly volatility components (Equation (1)).

$$\sqrt{RV_{t+1d}^{(d)}} = \alpha + \beta^{(d)}\sqrt{RV_t^{(d)}} + \beta^{(w)}\sqrt{RV_t^{(w)}} + \beta^{(m)}\sqrt{RV_t^{(m)}} + \omega_{t+1d}, \tag{1}$$

where $RV_t^{(d)}, RV_t^{(w)},$and $RV_t^{(m)}$ denote the daily, weekly, and monthly realized volatility of the stock returns, respectively ($RV_{t+1d}^{(d)}$ is the one-day-ahead daily realized volatility). The realized volatilities are defined as follows.

$$RV_t^{(d)} = \sum_{j=0}^{M-1} r_{t-j\bullet\Delta}^2, \tag{2}$$

$$RV_t^{(w)} = \left\{\frac{1}{5}\left(\sqrt{RV_t^{(d)}} + \sqrt{RV_{t-1d}^{(d)}} + \cdots + \sqrt{RV_{t-4d}^{(d)}}\right)\right\}^2, \tag{3}$$

$$RV_t^{(m)} = \left\{\frac{1}{22}\left(\sqrt{RV_t^{(d)}} + \sqrt{RV_{t-1d}^{(d)}} + \cdots + \sqrt{RV_{t-21d}^{(d)}}\right)\right\}^2, \tag{4}$$

where $\Delta = 1d/M$, and $r_{t-j\bullet\Delta} = p(t - j \bullet \Delta) - p(t - (j+1) \bullet \Delta)$ defines continuously compounded Δ-frequency returns, that is, intraday returns sampled at time interval Δ—— (here, the subscript t indexes the day, while j indexes the time within the day t).

Realized volatility is a volatility estimator independent of theoretical models which is proposed by Andersen and Bollerslev (1998). As shown in Equation (2), it is obtained by summing the squares of intraday returns. Corsi (2009) calculates the realized volatility using 2-hour data (i.e., $M = 12$), while it is common to use 5-minute data to compute it.

However, in our environment, it is difficult to measure realized volatility using high-frequency data. Although IB-Matlab, which we utilize in building the automated trading algorithm, provides a function to obtain intraday data, the length of data that can be retrieved from IB is much shorter in comparison to daily data. For example, we can obtain 5-minute data only for the last one year.

Given this data constraint, in this paper, we use the squared value of daily log returns instead of the realized volatility. The squared value of daily log returns is equivalent to the realized volatility with $M = 1$, and defined as follows.

$$DV_t = \{100 \times (\log(P_t) - \log(P_{t-1}))\}^2, \tag{5}$$

where P_t denotes the closing price on day t.

There is a rationale for using the squared value of daily log returns instead of the realized volatility in the HAR model. Specifically, Allen (2020) compares the predictive performance of the GARCH model which uses the realized volatility derived from 5-minute data, and the AR model which uses the squared value of daily log returns, and concludes that there is no significant difference between them. Since HAR models have the similar structure with AR models as mentioned above, it would be valid to use the squared value of daily log returns in our HAR model as well.

In addition, for the sake of comparing the predicted volatility from the model with the option IVs in the automated trading algorithm, we design the HAR model so that the time horizon of dependent variable can be varied according to the maturity of the compared option (Equation (6)).

$$\sqrt{DV_{t+hd}^{(h)}} = \alpha + \beta^{(d)}\sqrt{DV_t^{(d)}} + \beta^{(w)}\sqrt{DV_t^{(w)}} + \beta^{(m)}\sqrt{DV_t^{(m)}} + \omega_{t+hd} \tag{6}$$

where $DV_{t+hd}^{(h)}$ represents the average stock return volatility for the period from one day to h days ahead which is given below.

$$DV_{t+hd}^{(h)} = \left\{\frac{1}{h}\left(\sqrt{DV_{t+hd}^{(d)}} + \sqrt{DV_{t+(h-1)d}^{(d)}} + \cdots + \sqrt{DV_{t+1d}^{(d)}}\right)\right\}^2 \tag{7}$$

2.2 Heterogenous autoregressive model with extra inputs (HARX)

In order to improve the forecasting accuracy, we refine the original HAR model in three aspects. The first refinement is extending HAR (3) to HAR (5). Fernandez et al. (2014) estimate an HAR model with bi-weekly (10-day) and quarterly (66-day) as well as daily, weekly, and monthly volatility components. They conclude that the model successfully explains the characteristics of the VIX index, such as its long memory. Following them, we include the bi-weekly and quarterly volatility components ($\sqrt{DV_t^{(bw)}}$ and $\sqrt{DV_t^{(q)}}$, respectively) in the covariates.

The second is to take advantage of the market volatility components. Hizmeri et al. (2022) find that incorporating daily, weekly, and monthly market realized volatility into the simple HAR model yields statistically and economically significant improvements in predictive power. Besides that, the VIX index is known to be extremely useful in forecasting stock return volatility (e.g., Blair et al. [2001], Martens and Zein [2004]). Based on these considerations, we use the daily to quarterly volatility components of the S&P 500 index ($\sqrt{MDV_t^{(d)}}$, $\sqrt{MDV_t^{(q)}}$, $\sqrt{MDV_t^{(w)}}$, $\sqrt{MDV_t^{(m)}}$, and $\sqrt{MDV_t^{(q)}}$) and the VIX index (VIX_t) as the covariates of the model.

The third extension is to take into account the asymmetry in volatility dynamics. Asymmetry in volatility dynamics refers to the tendency for stock prices to be more volatile on the day after a decline than on the day after an increase. Giot and Laurent (2004) account for this tendency by utilizing a lagged stock return which takes zero when the return is positive (i.e., min $(r_{t-1}, 0)$). Specifically, they estimate an ARFIMA model with that variable and find that its coefficient is statistically and significantly negative. Furthermore, Omori et al. (2021) estimate a HAR (3) model with a similar variable and draw the same conclusion as Giot and Laurent (2004). Therefore, we also add a similar variable (Equation (8)) to the model.

$$NR_t = \min\ (100 \times (\log(P_t) - \log(P_{t-1})), 0) \tag{8}$$

In summary, the HARX model which we will estimate in the next section is as follows.

$$\begin{aligned}\sqrt{DV_{t+hd}^{(h)}} = \alpha &+ \beta^{(d)}\sqrt{DV_t^{(d)}} + \beta^{(w)}\sqrt{DV_t^{(w)}} + \beta^{(bw)}\sqrt{DV_t^{(bw)}} + \beta^{(m)}\sqrt{DV_t^{(m)}} \\ &+ \beta^{(q)}\sqrt{DV_t^{(q)}} + \gamma^{(d)}\sqrt{MDV_t^{(d)}} + \gamma^{(w)}\sqrt{MDV_t^{(w)}} + \gamma^{(bw)}\sqrt{MDV_t^{(bw)}} \\ &+ \gamma^{(m)}\sqrt{MDV_t^{(m)}} + \gamma^{(q)}\sqrt{MDV_t^{(q)}} + \phi VIX_t + \psi NR_t + \omega_{t+hd}\end{aligned} \tag{9}$$

3 Estimation

3.1 Data

All data are obtained from Interactive Brokers using IB-Matlab unless otherwise noted. The full sample covers the period from October 3, 2005, when the VIX index is first available through IB-Matlab, to the day before the estimation was made (i.e., December 23, 2022).

The stocks we analyze were selected from among the 20 major U.S. companies with highly liquid options markets examined by Hizmeri et al. (2022). Taking into account the balance of industries, we finally chose the 12 companies listed in Table 1. The remainder of this section reports the estimation results for three of these companies, Amazon.com, Inc. (AMZN), International Business Machines Corporation (IBM), and The Coca-Cola Co. (KO).

3.2 Estimation Results

Table 2 presents the parameter estimates for the HAR and HARX models. While the dependent variable is basically the stock return volatility up to one month ahead ($DV^{(22)}_{t+22d}$) for each stock, we also report the estimation result for the model whose dependent variable is the AMZN's stock return volatility up to one week ahead ($DV^{(5)}_{t+5d}$), in order to check the estimates for a shorter time horizon. In addition, the table presents the estimates of the following AR-type model as a benchmark case.

$$\sqrt{DV^{(h)}_{t+hd}} = \alpha + \beta\sqrt{DV^{(h)}_{t}} + \omega_{t+hd} \tag{10}$$

Although the model is not strictly an AR model because the independent variable is the h days lag of the dependent variable, we call it as AR (1) model hereafter.

First, we check the estimation results of the HAR models. We find that the coefficients on the volatility components are estimated to be significantly positive for all time horizons. In addition, the adjusted R-squared and the information criteria of the HAR models improve from those of the AR (1) models. This indicates that information on various time horizons is useful in predicting stock return volatility, as argued by Corsi (2009).

Next, the estimation results of the HARX models show that the coefficient on the newly added quarterly volatility component ($\sqrt{DV^{(q)}_{t}}$) is significantly positive in most cases. The coefficients on the bi-weekly volatility components ($\sqrt{DV^{(bw)}_{t}}$) are less significant while they are estimated to be positive.

The coefficients on the market volatility components are hardly significant with the exception of the quarterly components ($\sqrt{MDV^{(q)}_{t}}$). It may be because most of their useful information is also contained in the VIX index, and they have no additional explanatory power after controlling for the VIX index. On the other hand, the coefficients on the quarterly volatility components ($\sqrt{MDV_t^{(q)}}$) are, interestingly, significantly negative in some cases. The reason for this is not clear, but it is possible that the significantly negative coefficients capture the market cyclicality over longer time horizons such as quarterly frequency.

The coefficients on the VIX index (VIX_t) are positive and highly significant. This implies that market participants' expectations about future market volatility, as represented by the VIX index, are also very useful in predicting individual stock return volatility. However, it is interesting that, as mentioned above, the long-term volatility components of individual stocks ($\sqrt{DV_t^{(q)}}$) and market ($\sqrt{MDV_t^{(q)}}$) are useful in forecasting stock return volatility even after controlling for the information included in the VIX index.

Finally, the coefficient on the variable capturing the volatility dynamics asymmetry (NR_t) is significantly negative in all estimations. This means that there is an asymmetry in volatility dynamics also in our model and sample period, i.e., the stock return volatility tends to be higher after the negative return than after the positive return.

In terms of forecasting accuracy, the HARX models show sizable improvements compared to the HAR models, which is evidenced by the adjusted R-squared. The information criteria also suggest that the "extra inputs" of the HARX model are useful for volatility forecast.

4 Out-of-sample forecasts

In this section, we compare the out-of-sample forecasting performance of each model. Specifically, we estimate the models in a rolling window of 2,500 days and make out-of-sample volatility forecasts. We then calculate the Mean Squared Error (MSE) and Mean Absolute Error (MAE) from the deviation between the predicted and actual volatility. In addition, we compute the R-squared of the Mincer-Zarnowitz regressions (Equation (11)), where we take the actual volatility as the dependent variable and the predicted volatility as the independent variable.

$$\sqrt{DV_{t+hd}^{(h)}} = b_0 + b_1 \widehat{\sqrt{DV_{t+hd}^{(h)}}} + u_{t+h} \tag{11}$$

Table 3 presents the out-of-sample predictive performance for the one-week and one-month ahead volatility forecasts. We find that, in most cases, the goodness of out-of-sample forecast is in the order of $HARX > HAR > AR$ for any performance measures.

Focusing on the HARX model, for example, the MAEs for the one-week ahead forecasts of AMZN, IBM, and KO are 0.558, 0.393, and 0.304, respectively. This indicates that the absolute forecast errors of the HARX model's one-week ahead forecast (on a daily basis) are

0.558%pt, 0.393%pt, and 0.304%pt on average. On an annualized basis, they are about 8.9%pt, 6.2%pt, and 4.8%pt, respectively, indicating that it is not easy to predict stock return volatility even with our full model. However, as the annualized absolute forecast errors of the AR (1) model are 9.8%pt, 6.7%pt, and 5.3%pt, respectively, the HARX model succeeds in improving the forecast accuracy by about 0.5%pt-0.9%pt compared to the baseline model.

5 Automated algo-trading

5.1 Trading algorithm

Our trading strategy is to compare the predicted volatility (PV) from the HARX model with the implied volatility (IV) calculated from the option prices and to sell short the ATM straddle if the IV exceeds the PV. The investment (short sale) weight of each option is determined in proportion to the size of the difference between IV and PV. In the rest of this subsection, we explain the MATLAB code described in the Appendix.

5.1.1 auto_algo.m

When you run this file, the *auto_order* function (see below) is executed, and the short sales orders are submitted to TWS by *IBMatlab* function. Prior to running this file, you need to manually specify the arguments of *auto_order*.

5.1.2 auto_order.m

This function calculates the IV and PV of each option specified by its arguments and submits a short sale order of ATM straddle to TWS if the IV exceeds the PV. The arguments of this function are as follows:

- *capital*: Total amount of money to be shorted (in dollars).
- *settle_date*: Date on which the transaction is to take place.
- *rf*: Risk-free rate.
- *expiry_list*: List of options maturity dates.
- *symbol_list*: List of symbols.
- *dividend_list*: List of dividend yields corresponding to the symbols in the *symbol_list* (the order of elements must be the same as the *symbol_list*).

The *auto_order* function first calculates the IV, PV, ATM strike price, and ATM straddle price for each option identified by the combination of a maturity date and a symbol in *expiry_list* and *symbol_list* using the *volComparison* function (see below). Then, options whose IV is lower than the PV are deleted. For the remaining options, the function determines the quantity to be sold in proportion to the size of the difference between the IV and PV. Finally, short sale orders for the ATM straddles are executed using the *IBMatlab* function. The limit prices of the orders are set to the most recent prices of the ATM straddles obtained by the *volComparison* function.

5.1.3 volComparison.m

This function calculates the IV, PV, ATM strike price, and ATM straddle price for a given option. Its arguments are as follows:

- *settle_date*: Date on which the transaction is to take place.
- *expiry*: Option maturity date.
- *rf*: Risk-free rate.
- *symbol*: Symbol of the stock.
- *dividend*: Dividend yield of the stock.

The ATM strike price is calculated by rounding the most recent close price of the stock (one-minute data) to the whole number. The price of the ATM straddle is also the most recent close price of the straddle in the one-minute price data obtained by *IBMatlab*.

The IV is calculated as the average of the IVs of ATM call and put options which are computed by the *getImpVol* function (see below). Finally, the PV is the predicted volatility of a given stock from the *settle_date* to the *expiry* which is calculated using the *HARX* function (see below).

5.1.4 getImpVol.m

This function uses the Barone-Adesi and Whaley model (Barone-Adesi and Whaley [1987]) to calculate the IV of a given American option. The eight arguments of this function are as follows:

- *symbol*: Symbol of the stock.
- *right*: Right of the option (call: "C", put: "P").
- *Strike*: Strike price of the option.
- *AssetPrice*: Spot price of the stock.
- *settle_date*: Date on which the transaction is to take place.
- *expiry*: Option maturity date.
- *dividend*: Dividend yield of the stock.
- *rf*: Risk-free rate.

The *impvbybaw* function from Matlab's Financial Instruments Toolbox is used to compute the IV.

5.1.5 HARX.m

This function uses the HARX model described in Section 2 to forecast stock return volatility. The arguments are as follows:

- *symbol*: Symbol of the stock.
- *horizon*: Time horizon over which the volatility forecast is made (in days).

The HARX model is estimated using the last 20 years of daily data obtained by *IBMatlab*. Since the model prediction of stock return volatility is on a daily basis as described in Section 2, it is annualized by being multiplied by $\sqrt{252}$ to make it comparable to the IV.

5.2 P&L from paper trading

We made actual trades on our Interactive Brokers' paper trading accounts by running the algorithm described in the previous subsection. The parameters in *auto_algo.m* were set as follows.

- *capital*: 30000 (USD).
- *settle_date*: "20230103".
- *rf*: 0.045. The upper limit of the FOMC's target range for the Federal Funds Rate.
- *expiry_list*: ["20230120"].
- *symbol_list*: ["AMZN", "ARNC", "BAC", "CAT", "DIS", "EXC", "FCX", "IBM", "KO", "SO", "WFC", "XRX"]. The 12 companies listed in Table 1.
- *dividend_list*: [0, 0, 0.0266, 0.2, 0, 0.0312, 0.0157, 0.0468, 0.0277, 0.0381, 0.0291, 0.069]. We retrieved these values from Yahoo finance.

Using the data up to December 30, 2022, we placed limit sell orders for ATM straddles by running *auto_algo.m* before market open on January 3, 2023. Orders that were not contracted by the end of January 4 were cancelled.

The details of the orders are presented in Table 4. For all of the options, the IV exceeds PV, and the difference between the two is sizably large. This is consistent with the fact that the strategy to long options generally yields large negative returns. In other words, the IVs derived from option prices can be influenced by factors such as investors' risk aversion, and as a result, the IVs are much higher than the volatility forecasts that can be reasonably derived based on past statistical relationships.

Table 5 shows the unrealized gains (losses) as of the market close on January 9 for the seven straddles that were successfully contracted. The five out of the seven options have positive unrealized gains, indicating a certain degree of success of the algorithm. On the other hand, however, the other two options have relatively large unrealized losses, resulting in the negative total unrealized return.

Figure 1 presents the price development of each straddle. According to this chart, the FCX straddle, which suffered the largest losses, shows extremely volatile movements, indicating that its straddle market may not be liquid enough to be suitable as a trading target in our algorithm.

In any case, given the high volatility of option prices, it is too premature to judge the profitability of our algorithm only based on the data up to January 9. Therefore, we will continue monitoring the profitability of the algorithm in the future (P&L from the trade until the maturity [Jan 20] will be added in the final version of the paper).

6 Conclusion

This paper develops an automated trading algorithm utilizing the stock return volatility forecasts of the HARX (Heterogenous Autoregressive with Extra Inputs) model. Our HARX model is built by refining the simple HAR model proposed by Corsi (2009) in the following three aspects: it i) has the higher lag order compared to the HAR model, ii) utilizes information on the volatility of overall stock market, and iii) takes into account asymmetry in volatility dynamics. The forecasting accuracy of the HARX model outperforms the HAR model, both in-sample and out-of-sample, and for any performance measures. Then we detail an automated stock options trading algorithm incorporating the HARX model's predictions using IB-Matlab. Finally, we ran the algorithm after the beginning of 2023 and carried out a preliminary assessment of its profitability.

To conclude this paper, we discuss the future research issues that we have found through the analysis of this paper. First of all, for the sake of more accurate profitability assessment of our algorithm, it is necessary to operate it over a longer period of time as we mentioned in the previous section. Second, although we used the squared value of daily returns instead of realized volatility due to our data constraints, it would be ideal to refine the model by using realized volatility calculated from intraday data. The variance of the squared value of daily returns is considered to be larger than that of realized volatility especially in shorter time horizons. It makes our short-term volatility forecast very difficult, which is evidenced by the difference in the R-squared of the one-week-ahead and the one-month-ahead volatility forecasts presented in Table 2. Third, we should take into account the important economic events, such as the earnings reports and the FOMC meetings, which can have a significant impact on stock prices. Fortunately, whether the model considers these events seems not to have had a large impact on the assessment result since they were not scheduled during the assessment period in Section 5. However, taking such information into account in the model is essential for the practical use of the algorithm in the future.

References

Allen, D. E. (2020), "Stochastic Volatility and GARCH: Do Squared End-of-Day Returns Provide Similar Information?" *Journal of Risk and Financial Management*, vol. 13(9), pp. 1-25.

Andersen, T. G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," International Economic Review, vol. 39(4), pp. 885-905.

Barone-Adesi, G. and R. E. Whaley (1987), "Efficient Analytic Approximation of American Option Values," *The Journal of Finance*, vol. 42(2), pp. 301–320.

Blair, B. J., S.-H. Poon, and S. J. Taylor (2001), "Forecasting S&P 100 Volatility: the Incremental Information Content of Implied Volatilities and High-Frequency Index Returns," *Journal of Econometrics*, vol. 105, pp. 5-26.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, vol. 31, pp. 307-327.

Corsi, F. (2009), "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, vol. 7(2), pp. 174–196.

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, vol. 50, pp. 987-1007.

Fernandes, M., M. C. Medeiros, and M. Scharth (2014), "Modeling and Predicting the CBOE Market Volatility Index," *Journal of Banking & Finance*, vol. 40, pp. 1-10.

Giot, P. and S. Laurent (2004), "Modeling Daily Value-at-Risk Using Realized Volatility and ARCH type Models," *Journal of Empirical Finance*, vol. 11, pp. 379-398.

Hizmeri, R. M. Izzeldin, I. Nolte, and V. Pappas (2022) "A Generalized Heterogeneous Autoregressive Model Using Market Information," *Quantitative Finance*, vol. 22(8), pp. 1513-1534.

Martens, M. and J. Zein (2004), "Predicting Financial Volatility: High-Frequency Time-Series Forecasts Vis-a-Vis Implied Volatility," *Journal of Futures Markets*, vol. 24(11), pp. 1005-1028.

Omori, Y., M. Takahashi, and T. Watanabe (2021), "Forecasting Daily Volatility of Stock Price Index Using Daily Returns and Realized Volatility," *Econometrics and Statistics*, available online (in press).

Table 1 Stocks analyzed in the paper

Company	Symbol
Amazon.com, Inc.	AMZN
Arconic Inc.	ARNC
Bank of America Corporation	BAC
Caterpillar Inc	CAT
The Walt Disney Company	DIS
Exelon Corporation	EXC
Freeport-McMoRan Inc.	FCX
International Business Machines Corporation	IBM
The Coca-Cola Co.	KO
Southern Co.	SO
Wells Fargo & Company	WFC
Xerox Corporation	XRX

Table 2 Estimation results

Stock: Amazon.com, Inc. (AMZN)

Dependent Variable: $DV^{(22)}_{t+22d}$

	AR	HAR	HARX
Constant	0.616*** (0.11)	0.612*** (0.107)	0.235** (0.109)
AR coefficient	0.626*** (0.08)		
$\sqrt{DV_t^{(d)}}$		0.013*** (0.004)	-0.006 (0.005)
$\sqrt{DV_t^{(w)}}$		0.125*** (0.045)	0.031** (0.015)
$\sqrt{DV_t^{(bw)}}$			0.108*** (0.045)
$\sqrt{DV_t^{(m)}}$		0.491*** (0.047)	-0.006 (0.113)
$\sqrt{DV_t^{(q)}}$			0.553*** (0.152)
$\sqrt{MDV_t^{(d)}}$			0.000 (0.008)
$\sqrt{MDV_t^{(w)}}$			0.101 (0.072)
$\sqrt{MDV_t^{(bw)}}$			-0.048 (0.099)
$\sqrt{MDV_t^{(m)}}$			0.123 (0.168)
$\sqrt{MDV_t^{(q)}}$			-0.588*** (0.228)
VIX_t			0.030*** (0.007)
NR_t			-0.033*** (0.012)
Adjusted R^2	0.39	0.405	0.494
AIC	818.5	807.7	733.1
BIC	819.7	810.3	741.3
Observations	4,278	4,278	4,234

We provide the point estimates for the coefficients as well as their standard errors computed with Newey–West correction for serial correlation within parentheses. ***, **, and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 2 *Cont'd*

Stock: Amazon.com, Inc. (AMZN)

Dependent Variable: $DV_{t+5d}^{(5)}$

	AR	HAR	HARX
Constant	0.878***	0.452***	0.026
	(0.113)	(0.102)	(0.099)
AR coefficient	0.468***		
	(0.079)		
$\sqrt{DV_t^{(d)}}$		0.018**	-0.006
		(0.008)	(0.011)
$\sqrt{DV_t^{(w)}}$		0.139***	0.017
		(0.045)	(0.035)
$\sqrt{DV_t^{(bw)}}$			0.101
			(0.081)
$\sqrt{DV_t^{(m)}}$		0.568***	0.118
		(0.056)	(0.11)
$\sqrt{DV_t^{(q)}}$			0.476***
			(0.102)
$\sqrt{MDV_t^{(d)}}$			-0.028
			(0.021)
$\sqrt{MDV_t^{(w)}}$			0.152
			(0.121)
$\sqrt{MDV_t^{(bw)}}$			0.003
			(0.177)
$\sqrt{MDV_t^{(m)}}$			-0.057
			(0.142)
$\sqrt{MDV_t^{(q)}}$			-0.604***
			(0.149)
VIX_t			0.043***
			(0.007)
NR_t			-0.051***
			(0.014)
Adjusted R^2	0.219	0.307	0.373
AIC	1,176.0	1,119.0	1,067.7
BIC	1,177.2	1,121.5	1,075.9
Observations	4,312	4,295	4,251

We provide the point estimates for the coefficients as well as their standard errors computed with Newey–West correction for serial correlation within parentheses. ***, **, and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 2 *Cont'd*

Stock: International Business Machines Corporation (IBM)

Dependent Variable: $DV_{t+22d}^{(22)}$

	AR	HAR	HARX
Constant	0.408***	0.401***	0.231***
	(0.066)	(0.063)	(0.063)
AR coefficient	0.496***		
	(0.078)		
$\sqrt{DV_t^{(d)}}$		0.02***	-0.013
		(0.005)	(0.008)
$\sqrt{DV_t^{(w)}}$		0.195	0.012
		(0.067)	(0.024)
$\sqrt{DV_t^{(bw)}}$			0.063
			(0.055)
$\sqrt{DV_t^{(m)}}$		0.388***	0.047
		(0.072)	(0.181)
$\sqrt{DV_t^{(q)}}$			0.308
			(0.203)
$\sqrt{MDV_t^{(d)}}$			0.01
			(0.01)
$\sqrt{MDV_t^{(w)}}$			0.118***
			(0.046)
$\sqrt{MDV_t^{(bw)}}$			0.129*
			(0.073)
$\sqrt{MDV_t^{(m)}}$			-0.105
			(0.224)
$\sqrt{MDV_t^{(q)}}$			-0.282*
			(0.171)
VIX_t			0.023***
			(0.005)
NR_t			-0.031***
			(0.012)
Adjusted R^2	0.355	0.388	0.484
AIC	453.9	432.0	358.2
BIC	455.2	434.5	366.5
Observations	4,277	4,277	4,233

We provide the point estimates for the coefficients as well as their standard errors computed with Newey–West correction for serial correlation within parentheses. ***, **, and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 2 *Cont'd*

Stock: The Coca-Cola Co. (KO)

Dependent Variable: $DV^{(22)}_{t+22d}$

	AR	HAR	HARX
Constant	0.369*** (0.036)	0.36*** (0.035)	0.21*** (0.044)
AR coefficient	0.542*** (0.054)		
$\sqrt{DV_t^{(d)}}$		0.029*** (0.006)	0.000 (0.007)
$\sqrt{DV_t^{(w)}}$		0.235*** (0.093)	0.054*** (0.021)
$\sqrt{DV_t^{(bw)}}$			0.094* (0.055)
$\sqrt{DV_t^{(m)}}$		0.289*** (0.076)	-0.247** (0.122)
$\sqrt{DV_t^{(q)}}$			0.378** (0.192)
$\sqrt{MDV_t^{(d)}}$			0.008 (0.007)
$\sqrt{MDV_t^{(w)}}$			0.102*** (0.042)
$\sqrt{MDV_t^{(bw)}}$			0.093 (0.058)
$\sqrt{MDV_t^{(m)}}$			0.061 (0.092)
$\sqrt{MDV_t^{(q)}}$			-0.259 (0.163)
VIX_t			0.018*** (0.004)
NR_t			-0.029*** (0.009)
Adjusted R^2	0.295	0.345	0.484
AIC	357.2	325.5	224.5
BIC	358.4	328.1	232.8
Observations	4,277	4,277	4,233

We provide the point estimates for the coefficients as well as their standard errors computed with Newey–West correction for serial correlation within parentheses. ***, **, and * indicate a significance level of 1%, 5%, and 10%, respectively.

Table 3 Predictive performance of each model

	AMZN			IBM			KO		
	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2
5 DAYS AHEAD									
AR	0.684	0.615	0.146	0.368	0.425	0.161	0.247	0.331	0.215
HAR	0.609	0.572	0.242	0.336	0.405	0.229	0.227	0.315	0.277
HARX	0.572	0.558	0.287	0.286	0.393	0.348	0.198	0.304	0.368
22 DAYS AHEAD									
AR	0.331	0.445	0.243	0.210	0.299	0.157	0.167	0.238	0.137
HAR	0.324	0.440	0.259	0.202	0.294	0.188	0.164	0.237	0.152
HARX	0.338	0.435	0.250	0.193	0.277	0.246	0.157	0.236	0.195

We use a rolling window of 2,500 time-series observations to estimate the different models and then perform out-of-sample forecasting evaluation in the remaining of the series. We gauge forecasting performance by means of the mean squared forecast error (MSE), the mean absolute forecast error (MAE), and the R^2 of the Mincer-Zarnowitz regressions.

Table 4 Order details

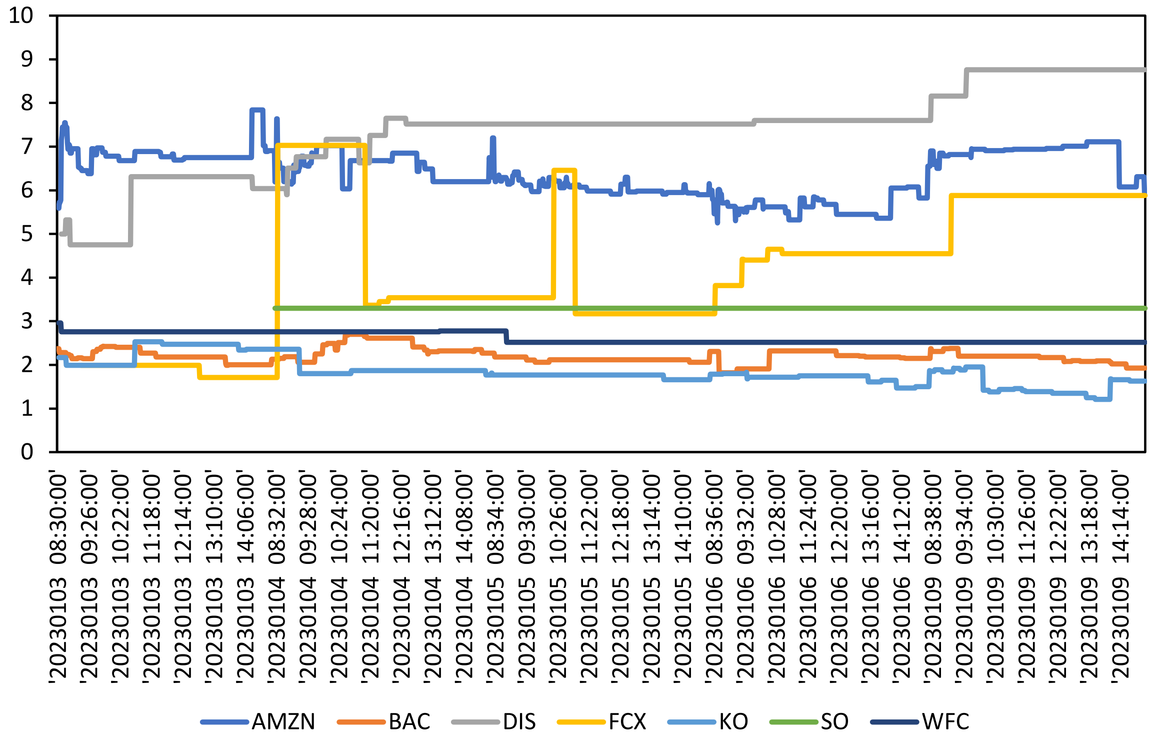
Symbol	Expiry	Strike price	IV	PV	IV - PV	Limit price	Order quantity	Contracted orders
AMZN	"20230120"	84	0.472	0.303	0.170	6.88	4	4
ARNC	"20230120"	21	0.504	0.382	0.122	3	7	0
BAC	"20230120"	33	0.375	0.203	0.172	2.14	15	15
CAT	"20230120"	240	0.315	0.205	0.110	13.03	2	0
DIS	"20230120"	87	0.402	0.219	0.183	6.69	5	5
EXC	"20230120"	43	0.246	0.134	0.113	1.9	11	0
FCX	"20230120"	38	0.515	0.301	0.214	3.2	12	12
IBM	"20230120"	141	0.260	0.147	0.113	5.91	3	0
KO	"20230120"	64	0.175	0.117	0.058	1.9	6	6
SO	"20230120"	71	0.227	0.141	0.086	2.8	6	6
WFC	"20230120"	41	0.381	0.197	0.184	2.71	12	12
XRX	"20230120"	15	0.440	0.301	0.139	1.2	21	0

All of the orders are selling short the straddles. Contracted orders are the number of orders contracted by the market close of Jan 4, 2023.

Table 5 Gains and losses as of Jan 9, 2023

Symbol	Limit price	Close price on Jan 9	Unrealized gain (in USD)	Symbol	Limit price	Close price on Jan 9	Unrealized gain (in USD)
AMZN	6.88	5.93	380	KO	1.9	1.8	60
BAC	2.14	2.06	120	SO	2.8	2.03	462
DIS	6.69	8.52	-915	WFC	2.71	2.7	12
FCX	3.2	5.01	-2,172	TOTAL		-2,053	

Figure 1   Straddle price fluctuations from Jan 3 to Jan 9, 2023



Appendix: MATLAB Code

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% auto_algo.m %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Amout of money you want to sell short in total (in USD)
capital = 30000;

%The date when you want to trade options
settle_date = "20230103";

%Risk free rate
rf = 0.045; %The upper bound of FFR

%Expirations of options you want to trade
expiry_list = ["20230120"];

%Symbols of options you want to trade
symbol_list = ["AMZN", "ARNC", "BAC", "CAT", "DIS", ...
    "EXC", "FCX", "IBM", "KO", "SO", "WFC", "XRX"];

%Dividend yields of options you want to trade
dividend_list = [0,0,0.0266,0.2,0,...
    0.0312,0.0157,0.0468,0.0277,0.0381,0.0291,0.069]; %Obtained from Yahoo
finance

%automatic trading
auto_order(capital, settle_date, rf, expiry_list, symbol_list, dividend_list);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% auto_order.m %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function ord_tbl = auto_order(capital, settle_date, rf, expiry_list,
symbol_list, dividend_list)

symbol_odr_list = [];
expiry_odr_list = [];
strd_odr_list   = [];
strike_odr_list = [];
impvol_odr_list = [];
predvol_odr_list = [];

for i=1:length(symbol_list)
    symbol = symbol_list(i);
    dividend = dividend_list(i);

    for j=1:length(expiry_list)
        expiry = expiry_list(j);
        try
            [symbol_tmp, expiry_tmp, strd_tmp, strike_tmp, impvol_tmp,
predvol_tmp] = ...
                    (settle_date, expiry, rf, symbol, dividend);
            symbol_odr_list = [symbol_odr_list, symbol_tmp];
            expiry_odr_list = [expiry_odr_list, expiry_tmp];
            strd_odr_list   = [strd_odr_list,   strd_tmp];
            strike_odr_list = [strike_odr_list, strike_tmp];
            impvol_odr_list = [impvol_odr_list, impvol_tmp];
            predvol_odr_list = [predvol_odr_list, predvol_tmp];
        catch
        end
        formatSpec = "Calculation for %s's stock option that expires on %s has
been completed.";
        A1 = symbol;
        A2 = expiry;
        disp(sprintf(formatSpec,A1,A2))
    end
end

ord_tbl = table(symbol_odr_list', expiry_odr_list', strd_odr_list',
strike_odr_list', ...
    impvol_odr_list', predvol_odr_list');
ord_tbl =
renamevars(ord_tbl,1:6,["symbol","expiry","price","strike","ImpVol","PredVol"]);

%Remove data such that ImpVol < PredVol
ord_tbl.dif = ord_tbl.ImpVol - ord_tbl.PredVol;
index = ord_tbl.dif > 0;
ord_tbl = ord_tbl(index,:);
```

```
%Weight for each option is propotional to the difference btw ImpVol &
%PredVol
total_dif = sum(ord_tbl.dif);
ord_tbl.quantity = round(capital*(ord_tbl.dif/total_dif)./(ord_tbl.price*100));

for k = 1:height(ord_tbl)
    symbol = ord_tbl.symbol(k);
    expiry = ord_tbl.expiry(k);
    limitPrice = ord_tbl.price(k);
    strike = ord_tbl.strike(k);
    quantity = ord_tbl.quantity(k);

    IBMatlab('action','sell', 'exchange','CBOE', 'quantity',quantity,...
        'SecType','OPT', 'type','LMT', 'limitPrice',limitPrice, ...
        'symbol',symbol,'expiry',expiry,'right',{'Call','Put'}, ...
        'strike',[strike,strike], 'ComboActions',{'Buy','Buy'});
end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```matlab
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% volComparison.m %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [symbol, expiry, strd_price, Strike, ImpVol, PredVol] =
volComparison(settle_date, expiry, rf, symbol, dividend)

stock_data = IBMatlab('action','history', 'symbol', symbol, 'exchange', 'NYSE');

AssetPrice = stock_data.close(1,width(stock_data.close));
Strike = round(AssetPrice);

%Calculate implied volatility using Barone-Adesi and Whaley option pricing model
ImpVol_call = getImpVol(symbol, "C", Strike, AssetPrice, settle_date, expiry,
dividend, rf);
ImpVol_put  = getImpVol(symbol, "P", Strike, AssetPrice, settle_date, expiry,
dividend, rf);
ImpVol = (ImpVol_call + ImpVol_put)/2;

pause(5);
strd_data = IBMatlab('action','history', 'exchange','CBOE','SecType','OPT', ...
    'symbol',symbol,'expiry',expiry,'right',{'Call','Put'}, ...
    'strike',[Strike,Strike], 'ComboActions',{'Buy','Buy'});

strd_price = [strd_data.close(1,width(strd_data.close))];

yyyy = str2double(extractBetween(expiry,1,4));
mm   = str2double(extractBetween(expiry,5,6));
dd   = str2double(extractBetween(expiry,7,8));

yyyy_s = str2double(extractBetween(settle_date,1,4));
mm_s   = str2double(extractBetween(settle_date,5,6));
dd_s   = str2double(extractBetween(settle_date,7,8));

date_e = datetime(yyyy,mm,dd);
date_s = datetime(yyyy_s,mm_s,dd_s);

%Predict future volatility using HARX model
horizon = length(busdays(date_s,date_e))+1;
PredVol = HARX(symbol, horizon);

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```matlab
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% getImpVol.m %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function ImpVol = getImpVol(symbol, right, Strike, AssetPrice, settle_date,
expiry, dividend, rf)

price_data = IBMatlab('action', 'history', 'symbol', symbol, 'right', right, ...
    'secType','OPT', 'expiry', expiry, 'strike', Strike, 'multiplier',100);

yyyy = str2num(extractBetween(expiry,1,4));
mm   = str2num(extractBetween(expiry,5,6));
dd   = str2num(extractBetween(expiry,7,8));

yyyy_s = str2num(extractBetween(settle_date,1,4));
mm_s   = str2num(extractBetween(settle_date,5,6));
dd_s   = str2num(extractBetween(settle_date,7,8));

Settle = datetime(yyyy_s,mm_s,dd_s);
Maturity = datetime(yyyy,mm,dd);
DivAmount = dividend;
Rate = rf;

RateSpec = intenvset('ValuationDate', Settle, 'StartDates', Settle,...
    'EndDates', Maturity, 'Rates', Rate, 'Compounding', -1, 'Basis', 1);

StockSpec = stockspec(NaN, AssetPrice, {'continuous'}, DivAmount);

if right == "C"
    right2 = "call";
else
    right2 = "put";
end

OptSpec = {right2};
OptionPrice = [price_data.close(1,width(price_data.close))];

ImpVol = impvbybaw(RateSpec, StockSpec, Settle, Maturity, OptSpec,...
    Strike, OptionPrice);

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% HARX.m %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function vol = HARX(symbol, horizon)

%download daily data from IB
pause(5);
data = IBMatlab('action','history', 'symbol', symbol, ...
    'barSize','1 day', 'useRTH',0, ...
    'DurationValue',20, 'DurationUnits','Y','exchange', 'NYSE');

market_data = IBMatlab('action','history', 'symbol', 'SPX', ...
    'secType', 'IND', 'localSymbol', 'SPX', 'exchange', 'CBOE', ...
    'currency', 'USD','barSize','1 day', 'useRTH',0, ...
    'DurationValue',20, 'DurationUnits','Y');

VIX_data = IBMatlab('action','history', 'symbol', 'VIX', ...
    'secType', 'IND', 'localSymbol', 'VIX', 'exchange', 'CBOE', ...
    'currency', 'USD','barSize','1 day', 'useRTH',0, ...
    'DurationValue',20, 'DurationUnits','Y');

%%create table
%individual stock
close = data.close';
dt = data.dateTime';
tbl = table(dt, close);
%spx
market_close = market_data.close';
dt = market_data.dateTime';
market_tbl = table(dt, market_close);
%vix
VIX_close = VIX_data.close';
dt = VIX_data.dateTime';
VIX_tbl = table(dt, VIX_close);
%innerjoin
merge_tbl = innerjoin(tbl, market_tbl,'Keys','dt');
merge_tbl = innerjoin(merge_tbl, VIX_tbl,'Keys','dt');

tbl = table(merge_tbl.dt, merge_tbl.close);
market_tbl = table(merge_tbl.dt, merge_tbl.market_close);
VIX_tbl = table(merge_tbl.dt, merge_tbl.VIX_close);

tbl = renamevars(tbl,1:2,["dt","close"]);
market_tbl = renamevars(market_tbl,1:2,["dt","close"]);
VIX_tbl = renamevars(VIX_tbl,1:2,["dt","close"]);

%%create variables
%individual stock
tbl.lclose = log(tbl.close);
tbl.return = cat(1,NaN, diff(tbl.lclose));
```

```
tbl.return_neg = min(tbl.return, 0).*100;
tbl.return2 = sqrt(tbl.return.^2).*100;%daily
tbl.return2_w = movmean(tbl.return2, [4 0]);%weekly
tbl.return2_bw = movmean(tbl.return2, [9 0]);%bi-weekly
tbl.return2_m = movmean(tbl.return2, [21 0]);%monthly
tbl.return2_q = movmean(tbl.return2, [65 0]);%quarterly
tbl.return2_y = movmean(tbl.return2, [horizon-1 0]);%object
%spx
market_tbl.lclose = log(market_tbl.close);
market_tbl.return = cat(1,NaN, diff(market_tbl.lclose));
market_tbl.return2 = sqrt(market_tbl.return.^2).*100;%daily
market_tbl.return2_w = movmean(market_tbl.return2, [4 0]);%weekly
market_tbl.return2_bw = movmean(market_tbl.return2, [9 0]);%bi-weekly
market_tbl.return2_m = movmean(market_tbl.return2, [21 0]);%monthly
market_tbl.return2_q = movmean(market_tbl.return2, [65 0]);%quarterly

%regression
y = tbl.return2_y;
XLag =
lagmatrix([tbl.return2,tbl.return2_w,tbl.return2_bw,tbl.return2_m,tbl.return2_q,
...

market_tbl.return2,market_tbl.return2_w,market_tbl.return2_bw,market_tbl.return2
_m, ...
    market_tbl.return2_q,VIX_tbl.close,tbl.return_neg],horizon);
X = [ones(size(y)), XLag];
b = regress(y,X);

%prediction
T = length(tbl.return2_y);
pred_value = b(1) + b(2)*tbl.return2(T) + b(3)*tbl.return2_w(T) + ...
    b(4)*tbl.return2_bw(T) + b(5)*tbl.return2_m(T) + b(6)*tbl.return2_q(T) + ...
    b(7)*market_tbl.return2(T) + b(8)*market_tbl.return2_w(T) + ...
    b(9)*market_tbl.return2_bw(T) + b(10)*market_tbl.return2_m(T) + ...
    b(11)*market_tbl.return2_q(T) + b(12)*VIX_tbl.close(T) +
b(13)*tbl.return_neg(T);
vol = (pred_value/100)*(252)^0.5; %prediction of annualized volatility (%) over
the horizon

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```