University of Pretoria
Department of Economics Working Paper Series

Machine Learning and the Forecastability of Cross-Sectional Realized Variance: The Role of Realized Moments

Vasilios Plakandaras
Democritus University of Thrace
Matteo Bonato
University of Johannesburg
Rangan Gupta
University of Pretoria
Oguzhan Cepni
Copenhagen Business School, University of Edinburgh Business School and Ostim Technical University
Working Paper: 2025-18
April 2025

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Machine Learning and the Forecastability of Cross-Sectional Realized Variance: The Role of Realized Moments

Vasilios Plakandaras[*], Matteo Bonato[**], Rangan Gupta[***] and Oguzhan Cepni[****]

Abstract

This paper forecasts monthly cross-sectional realized variance (RV) for U.S. equities across 49 industries and all 50 states. We exploit information in both own-market and cross-market (oil) realized moments—semi-variance, leverage, skewness, kurtosis, and upside and downside tail risk—as predictors. To accommodate cross-sectional dependence, we compare standard econometric panel models with machine-learning approaches and introduce a new machine-learning technique tailored specifically to panel data. Using observations from April 1994 through April 2023, the panel-dedicated machine-learning model consistently outperforms all other methods, while oil-related moments add little incremental predictive power beyond own-market moments. Short-horizon forecasts successfully capture immediate shocks, whereas longer-horizon forecasts reflect broader structural economic changes. These results carry important implications for portfolio allocation and risk management.

Keywords: Cross-sectional realized variance; Realized moments; Machine learning; Forecasting

JEL Codes: C33; C53; G10; G17

1. Introduction

Given that the volatility of stock market returns is a key input for portfolio and hedging decisions, appropriate modelling and accurate forecasts of the same are critical for the effectiveness of portfolio and risk management strategies, as well as the pricing of derivative securities (Poon and Granger, 2003; Rapach et al., 2008). In this regard, recent studies (see, for example, Angelidis et al. (2015), Demirer et al. (2019), Lu and Ma (2023), Niu et al. (2023, 2024), Cepni et al. (forthcoming))[1] highlight the predictive role of industry- and state-level volatility information for aggregate stock-market variability of the United States (U.S.), in line with gradual diffusion of information hypothesis of Hong et al. (2007).[2] These studies tend to focus on the concept of daily or monthly Realized Variance (*RV*), i.e., the sum of intraday or

[*] Department of Economics, Democritus University of Thrace, Komotini, Greece. Email: vplakand@econ.duth.gr.
[**] Corresponding author. Department of Economics and Econometrics, University of Johannesburg, Auckland Park, South Africa; IPAG Business School, 184 Boulevard Saint-Germain, 75006 Paris, France. Email: matteobonato@gmail.com.
[***] Department of Economics, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa. Email: rangan.gupta@up.ac.za.
[****] Ostim Technical University, Ankara, Turkiye; University of Edinburgh Business School, Centre for Business, Climate Change, and Sustainability; Department of Economics, Copenhagen Business School, Denmark. Email: oce.eco@cbs.dk.

[1] In this regard, Tsuji (2012) and Demirer et al. (2024), show that industry-level stock returns can also carry substantial predictive content for the overall volatility of the overall US equity market. Considering the well-established "leverage-effect" (Black, 1976), it would also imply a possible indirect effect on the aggregate stock market volatility, with industry returns impacting their corresponding sectoral volatilities.

[2] Hong et al. (2007) argues that public information is often partially reflected in asset prices as market participants who focus on trading the broad market indexes experience delays in receiving industry-specific information that is already available to professional industry investors. As a result, the authors argue the presence of a predictive relationship between industry returns and the aggregate stock market as the information contained in industry returns spreads gradually throughout the entire market. Though the original hypothesis is for the first-moment, the argument can be extended to the second-moment.

daily squared returns over a day or a month (Andersen and Bollerslev, 1998), thus yielding an observable and unconditional (unlike Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and Stochastic Volatility (SV) models)[3] metric of volatility for both predictors and the predictand. Given the important role of industry- and state-level *RV*s for predicting the future path of the overall U.S. stock market *RV*,[4] a question of paramount importance, academically and from an investment perspective, is: What factor drive cross-sectional *RV*?

Forecasting aggregate stock market realized variance (RV) in the U.S. has generated extensive literature, with studies typically relying on behavioral, financial, and macroeconomic predictors (see, for example, discussions in Salisu et al., 2022; Gupta et al., 2023a; Lu et al., 2023; Souropanis and Vivian, 2023). However, recent research (e.g., Mei et al., 2017; Zhang et al., 2021; Bonato et al., 2022, and references therein) emphasizes the predictive power of realized moments, such as realized semi-variances ("good" and "bad" RVs), realized leverage, realized skewness, realized kurtosis, and realized upside and downside tail risks. These realized moments have been shown to forecast aggregate U.S. stock market RV, as well as sector-specific RV (for example, real estate), equally well or even better than traditional predictors. Bonato et al. (2023, 2024a, b) attribute this superior performance to the fact that price-based factors inherently incorporate information about extreme behaviors in macroeconomic and financial variables, reflecting investor sentiment and disaster risks.

Against this backdrop, our paper analyzes the role of the aforementioned realized moments in forecasting monthly cross-sectional RVs across 49 industries and 50 U.S. states over the period from April 1994 to April 2023. Recognizing the importance of the link between oil shocks, uncertainty, and disaggregated stock volatilities (Salisu et al., 2025; forthcoming), we also incorporate oil market variables into our models—specifically, good and bad RVs, realized leverage, realized skewness, realized kurtosis, and realized upside and downside tail risks of oil returns. Moreover, moments of oil prices are expected to reflect the underlying state of the economy (Gupta et al., 2023b), a key state variable for the stock market (Salisu et al., 2024). To achieve our econometric objectives, we employ panel data regressions using both econometric and machine learning approaches, consistent with the methodology in Gu et al. (2020). The six machine learning approaches adopted in this paper allow us to explore a broad set of high-dimensional models for statistical prediction, using regularization techniques for model selection and overfitting mitigation, based on efficient algorithms that search across a wide array of potential model specifications.

To the best of our knowledge, this is the first paper to utilize multiple panel data-based machine learning approaches to forecast U.S. cross-sectional RV at both the industry and state levels using realized moments. The remainder of the paper is organized as follows: Section 2 discusses the data and the computation of the moments that form our set of predictors. Section 3 outlines the econometric models, while Section 4 presents the forecasting results. Finally, Section 5 concludes the paper.

[3] The reader is referred to studies by Hwang and Satchell (2005) and Byun (2016), for the utilization of such conditional models involving the nexus between cross-sectional and market volatilities.
[4] Note that, some studies (Ang et al., 2006; Detzel et al., 2023) have also related cross-sectional volatility to expected market returns.

2. Data and Realized Moments

As far as our dependent variable is concerned, we use the classical estimator of RV, i.e., the sum of squared daily returns during a month, given by $RV_t = \sum_{i=1}^{M} r_{t,i}^2$ where $r_{t,i}$ denotes monthly volatility on daily i returns over month t, with daily value-weighted returns for the 49 U.S. industries obtained from the Data Library of Professor Kenneth R. French.[5] At the same time, daily state-level stock returns are computed using data for the state-level stock market indexes obtained from Bloomberg terminal. These indexes represent the capitalization-weighted portfolios of equities domiciled in each state. As far as daily oil returns are concerned, we rely on the West Texas Intermediate (WTI) crude oil spot prices, obtained from the U.S. Energy Information Administration[6]. In terms of our predictors involving both the stock and the oil markets, we consider "good" (positive) and "bad" (negative) RVs to capture potential sign asymmetries in the RV process, such that:

$$RVG_t = \sum_{i=1}^{M} r_{t,i}^2 \times I_{[(r_{t,i} > 0]} \tag{1}$$

$$RV\quad_t = \sum_{i=1}^{M} r_{t,i}^2 \times I_{[(r_{t,i} < 0]} \tag{2}$$

where I is the indicator function. Then we include realized leverage ($RLEV$), which is the negative only values of monthly realized returns, such that:

$$RLEV_t = \sum_{i=1}^{M} r_{t,i} \times I_{[(r_{t,i} < 0]} \tag{3}$$

Besides RVG, RVB and $RLEV$, we also include realized skewness, RSK, realized kurtosis, RKU, and realized upside and downside tail risks, TR_u and TR_d, respectively.
Like Amaya et al. (2015), we use RSK to capture the asymmetry of the return's distribution, and RKU accounts for extremes. We compute RSK as:

$$RSK_t = \frac{\sqrt{M}\sum_{i=1}^{M} r_{t,i}^3}{RV_t^{3/2}} \tag{4}$$

and RKU as:

$$RK\quad_t = \frac{M\sum_{i=1}^{M} r_{t,i}^4}{RV_t^2} \tag{5}$$

where the scaling by $(M)^{1/2}$ and M turns the statistics into the corresponding monthly realized skewness and kurtosis values, respectively.
Last, we consider the Hill-tail risk estimator (Hill, 1975), to derive our realized upside and downside tail risks. Let $X_{t,i}$ the set of reordered daily returns on month $t, r_{t,i}$ in such a way that:

$$X_{t,i} \geq X_{t,j} \;\; for\; i < j \tag{6}$$

We compute the (monthly) positive tail risk estimator (TR_U) as:

$$H_t^{up} = \frac{1}{k}\sum_{i=1}^{k} \ln\left(X_{t,i}\right) - \ln(X_{t,k}) \tag{7}$$

and the (monthly) negative tail risk estimator (TR_d) as:

$$H_t^{down} = \frac{1}{k}\sum_{i=n-k}^{n} \ln\left(X_{t,i}\right) - \ln(X_{t,n-k}) \tag{8}$$

where k is the observation denoting the chosen α=5% tail interval.

---

[5] The data is available for download from the following internet page: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
[6] The data can be accessed from: https://www.eia.gov/dnav/pet/hist/RWTCD.htm.

Based on data availability at the time of writing this paper, and to allow for comparison across the industry- and state-level results, our monthly sample period covers April 1994 to April 2023.

3. Methodology

The Support Vector Regression (SVR), proposed by Vapnik (1995), builds on the idea of finding a loss function that minimizes the forecasting error based on a predetermined deviation (error) ε from the actual observation, instead of minimizing the typically used squared error. In detail, for a training dataset $D = [(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots .(\mathbf{x}_\mathrm{n}, y_n)], \mathbf{x}_\mathrm{i} \in \mathbb{R}^n, y_i \in \mathbb{R}, i = 1,2, \dots .n$, where $\mathbf{x}_\mathrm{i}$ is a vector of independent variables and y_i is the dependent one, the linear SVR function takes the form: $y = f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$. Then, the minimization problem is:

$$\min\left(\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{\iota=1}^{n}(\zeta_i + \zeta_i^*)\right) \tag{9}$$

$$\text{subject to} \begin{cases} y_i - (\mathbf{w}\mathbf{x}_\mathrm{i} + b) \leq \varepsilon + \zeta_i \\ (\mathbf{w}\mathbf{x}_\mathrm{i} + b) - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases}$$

where ε defines the width of the tolerance band, and ζ_i, ζ_i^* are slack variables controlled through a penalty parameter C that provide insensitivity to error. Although SVR solves several issues of the typical regression model, it fails to account for the cross-sectional dependence in the case of panel data. To introduce this issue to the SVR model, we embed cross-correlations directly to the kernel matrix (Plakandaras et al., 2024), along with heterogeneous coefficients across cross-sections. Suppose that we have panel observations collected from N entities (industries or states) at T occasions/periods (days) with K features (variables). The expanded matrix x_{it} takes the form $NT \times K$, where $\boldsymbol{y}$ is an NT$\times$ 1 vector of $\boldsymbol{y_i} = [y_{i1}, y_{i2}, \cdots, y_{it}]^T$ column vectors of size $T \times 1$. A typical heterogenous time-varying panel model would be written in the form:

$$y_{it} = d_i + x_{it}^T\beta_{it} + u_{it} \tag{11}$$

where d_i are individual specific effects, β_{it} are K-dimensional vectors of time-varying heterogenous coefficients over i, and model errors u_{it} are stationary over time t, but may be cross-sectionally dependent. We extend model (11) by introducing the $NT\times 1$ time-varying parameter vector $\boldsymbol{c_i} = [1, c_{i1}, c_{i2}, \cdots, c_{iT-}\ ]^T$ of column vectors of size $T \times 1$. So, for each entity i we formulate our dataset as:

$$x_{i1} + c_{i1}x_{i2} + \cdots + c_{iT-}\ x_{it} = x_{it}^T\boldsymbol{c_i} \tag{12}$$

$$y_{i1} + c_{i1}y_{i2} + \cdots + c_{iT-}\ y_{it} = \boldsymbol{y_{it}^T}\boldsymbol{c_i} \tag{13}$$

where the unknown parameter vector $\boldsymbol{c_i}$ considers observational dependence within sections (entities) and evolves with time. The introduction of parameter c both to x_{it} and y_{it} ensures that the SVR model will observe heterogeneous cross-sectional dependencies. In matrix form, considering a kernel function $\boldsymbol{K_m}$, the optimization process yields:

$$\min_{\boldsymbol{a_m}} := \frac{1}{2}\boldsymbol{a}_m \begin{bmatrix} \boldsymbol{K_m} & -\boldsymbol{K_m} \\ -\boldsymbol{K_m} & \boldsymbol{K_m} \end{bmatrix} \boldsymbol{a_m} + \boldsymbol{y_m^T}\boldsymbol{a_m} \tag{14}$$

with

$$\boldsymbol{a_m} = \begin{bmatrix} a^{*T} & c_1 a^{*T} & \cdots & c_{T-1} a^{*T} & a^T & c_1 a^T & \cdots & c_{T-1} a^T \end{bmatrix}^T$$
$$\boldsymbol{y_m} = [\varepsilon e^T - y_1^T \quad -y_2^T \quad \cdots \quad -y_t^T \quad \varepsilon e^T + y_1^T \quad y_2^T \quad \cdots \quad y_t^T\ ]^T$$
$$G = \begin{bmatrix} K(\boldsymbol{x_1}, \boldsymbol{x_1}) & \cdots & K(\boldsymbol{x_1}, \boldsymbol{x_t}) \\ \vdots & \ddots & \vdots \\ K(\boldsymbol{x_T}, \boldsymbol{x_1}) & \cdots & K(\boldsymbol{x_T}, \boldsymbol{x_t}) \end{bmatrix}$$
$$a_i, a_i^* \epsilon [0, C]$$
$$\sum_{i=1}^{n} (a_i - a_i^*) = 0$$
$$\boldsymbol{z}^T \begin{bmatrix} \boldsymbol{a}^* \\ \boldsymbol{a} \end{bmatrix} = 0$$

with no constraint imposed on parameter **c.** So, after determining a_m using gradient descent, we can estimate **c** that depicts the responses of each cross-section per time period. For comparison reasons with the panel dedicated SVR model, we also consider the Least Absolute Shrinkage and Selection Operator (LASSO), the typical SVR, an Artificial Neural Network (ANN), the Random Forest (RF), and XGBoost: a boosted RF model, while the popular Two-Way Fixed Effects (TWFE) regression is used as the benchmark.

4. Forecasting Results

As stated in the previous section, the scope of this paper is to analyze the ability of econometric and machine learning methods to forecast industry- and state-level U.S. stock return RVs. To this end, we train panel regressions to capture the cross-sectional dependence in the time series. Specifically, we use the standard econometric forecasting workhorse—the two-way fixed effects (TWFE) model—as our benchmark and implement rolling-window regressions with a window size of 24 months and a sliding step of 1 month to account for non-stationarity and structural breaks in the series. All machine learning models are trained using a 5-fold cross-validation scheme to avoid overfitting and to optimize model parameter selection. Additionally, we consider alternative independent variables—namely, the realized moments—as potential regressors to better understand the data-generating process driving the evolution of RV. In Table 1, we report the average out-of-sample Root Mean Square Errors (RMSEs) for forecasting horizons of 1, 3, 6, 12, and 24 months ahead. Panels A, B, and C correspond to forecasts of industry-level RVs, while Panels D, E, and F correspond to state-level RVs.

[INSERT TABLE 1]

In forecasting industry-level realized variances (RVs), the Panel-SVR model demonstrates clear and consistent dominance, achieving the lowest RMSE across all horizons and predictor sets (Panels A, B, and C). In nearly all cases, Panel-SVR significantly outperforms the benchmark two-way fixed effects (TWFE) model at the 1% level, underscoring its superior predictive capacity for industry portfolio volatility. In Panel A, based on industry moments and lagged volatility, LASSO, SVR, ANN, RF, and XGBoost generally outperform TWFE, particularly at horizons of three months and beyond. When WTI realized volatility is incorporated (Panel B), all machine learning models deliver statistically significant improvements over TWFE at short horizons, with LASSO, SVR, RF, and XGBoost often maintaining their superiority at longer horizons. In Panel C, which includes all WTI moments, TWFE continues to perform poorly at the 1-month horizon, while the machine learning models consistently yield better predictive accuracy. Nevertheless, there is no evidence that incorporating WTI moments enhances the forecasting performance for industry-level RVs beyond what is achieved using their own realized moments.

Similar patterns emerge for state-level RVs (Panels D, E, and F). The Panel-SVR model again achieves the lowest RMSE across all forecast horizons and predictor sets and significantly outperforms TWFE. Machine learning models consistently surpass the TWFE benchmark, and, as with the industry results, WTI moments provide no additional predictive value beyond the states' own realized moments. Overall, in line with the existing literature (Mei et al., 2017; Zhang et al., 2021; Bonato et al., 2022), our findings provide robust evidence that realized moments are powerful predictors of both industry- and state-level RVs, while cross-market (oil-related) moments offer limited incremental information. These results reinforce the view that stock price-based measures already embed critical macroeconomic and financial information relevant for volatility forecasting.

To better understand the behavior of the Panel-SVR model, Figure 1 depicts RMSE spikes from the rolling-window estimation procedure at the monthly and semi-annual forecasting horizons for U.S. industries and states, across three model specifications: (i) own moments; (ii) own moments plus WTI realized volatility; and (iii) own moments plus all WTI moments.

For the industry-level dataset, the monthly horizon reveals more frequent and varied RMSE spikes across all three model specifications. Notable volatility peaks include: August 2006 (RMSE = 26.0621; dominated by "Own + WTI moments") amid Middle East conflicts and oil price concerns; June 2007 (RMSE = 20.4905; "Own moments") during the early phase of the subprime mortgage crisis; November 2009 (RMSE = 22.1504; "Own moments + WTI realized volatility") in the aftermath of the global financial crisis; and August 2018 (RMSE = 18.3037) during the U.S.-China trade tensions. Other significant spikes occurred in February 1997 (Asian financial crisis onset), November 2002 (post-dot-com recession), September 2005 (Hurricane Katrina), May 2008 (pre-Lehman Brothers collapse), June 2012 (European sovereign debt crisis), October 2015 (China slowdown concerns), April 2016 (Brexit uncertainty), and December 2017 (cryptocurrency market volatility). These results suggest that models incorporating WTI moments are particularly sensitive to oil-related shocks.

At the semi-annual horizon, prediction patterns differ notably from the monthly results. The highest spike occurred in May 1997 (RMSE = 34.5911; "Own + WTI moments") during the peak of the Asian financial crisis. Other major episodes include November 1997 (continuation of the Asian crisis), April 2009 ("Own moments + WTI realized volatility"; financial crisis recovery), May 2012 (European sovereign debt crisis escalation), and April 2018 ("Own moments"; rising trade tensions). Additional notable spikes correspond to August 1996 (pre-Asian crisis pressures), July 1998 (Russian financial crisis), July 2006 (Middle East conflicts), January 2007 (early U.S. housing market concerns), and October 2019 (pre-pandemic uncertainty). At longer forecast horizons, greater variation is observed in which model specification best captures economic disruptions, with oil-related moments becoming particularly relevant during energy-driven episodes, while "Own moments" often dominate during broader financial market stress.

Regarding U.S. industries, starting with the monthly horizon, the graph reveals several major episodes of market disruption. The most significant RMSE spikes correspond to key events: July–November 2002 (post-dot-com bubble and corporate scandals such as Enron); November 2009 (aftermath of the 2008 financial crisis); February 2009 (ongoing financial turmoil and economic uncertainty); June 2012 (European sovereign debt crisis affecting global markets);

December 2017 (cryptocurrency bubble and tax reform uncertainty); and August 2018 (U.S.–China trade tensions and market corrections). The "Own moments + WTI realized volatility" specification consistently exhibits more extreme responses during these periods, suggesting that oil price volatility significantly amplifies market turbulence during crises.

At the semi-annual horizon, volatility spikes are even more pronounced. The most dramatic surge occurred in December 2011 (RMSE = 51.0196), coinciding with the European sovereign debt crisis and lingering effects of the 2008 financial collapse. Other notable spikes include April 2009 (RMSE = 24.2718) during the initial recovery from the financial crisis; May 2012 (RMSE = 24.3778) amid renewed European debt concerns and Greek instability; January 2010 (RMSE = 2.0036) during the early stages of economic recovery; February 2014 (RMSE = 2.6528) following emerging market turbulence; September 2017 (RMSE = 3.4850) driven by hurricane impacts and geopolitical tensions; and March 2020 (RMSE = 2.2515) at the onset of the COVID-19 pandemic. The six-month forecast horizon captures volatility patterns that differ markedly from the one-month horizon, with the "Own + WTI moments" model specification often displaying stronger responses during major disruptions.

[INSERT FIGURE 1]

A notable observation across all figures is that forecasting errors are quickly captured by the models, indicating that structural breaks are incorporated rapidly as the rolling window advances.

5. Conclusion

We forecast the monthly cross-sectional realized variance (RV) of 49 industries and 50 U.S. states over the period from April 1994 to April 2023, using information contained in both own and cross-market (oil) realized moments—namely, realized semi-variances, realized leverage, realized skewness, realized kurtosis, and realized upside and downside tail risks. To this end, we employ dedicated panel regression machine learning methods, which consistently outperform traditional econometric and machine learning approaches in terms of forecasting accuracy. Our results show that oil market moments do not improve forecasting performance beyond own-market moments, and that short- and long-term forecasts are governed by different dynamics, which forecasting models are able to capture rapidly. These findings have important implications for investors, given that cross-sectional RV is known to predict aggregate RV as well as expected market returns.

As part of future research, it would be interesting to extend this analysis by incorporating other industry- and state-specific economic and financial variables, and by applying the framework to other developed and emerging stock markets, as well as to panels of alternative asset markets such as commodities and (crypto-)currencies.

References

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? Journal of Financial Economics, 118(1), 135-167.

Andersen, T.G., and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review, 39(4), 885-905.

Ang, A., Hodrick, R.J., Xing, Y., and Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. The Journal of Finance, 61(1), 259-299.

Angelidis, T., Sakkas, A., and Tessaromatis, N. (2015). Stock market dispersion, the business cycle and expected factor returns? Journal of Banking & Finance, 59(C), 265-279.

Black, F. (1976). Studies of stock price volatility changes. In: Proceedings of the 1976 Meeting of the Business and Economic Statistics Section; American Statistical Association: Washington DC, USA, 177-181.

Bonato, M., Cepni, O. Gupta, R., and Pierdzioch, C. (2022). Forecasting realized volatility of international REITs: The role of realized skewness and realized kurtosis. Journal of Forecasting, 41(2), 303-315.

Bonato, M., Cepni, O. Gupta, R., and Pierdzioch, C. (2023). Business applications and state-level stock market realized volatility: A forecasting experiment. Journal of Forecasting, 43(2), 456-472.

Bonato, M., Cepni, O. Gupta, R., and Pierdzioch, C. (2024a). Financial stress and realized volatility: The case of agricultural commodities. Research in International Business and Finance, 71, 102442.

Bonato, M., Cepni, O. Gupta, R., and Pierdzioch, C. (2024b). Forecasting the realized volatility of agricultural commodity prices: Does sentiment matter?, Journal of Forecasting, 43(6), 2088-2125.

Byun, S.J. (2016). The usefulness of cross-sectional dispersion for forecasting aggregate stock price volatility. Journal of Empirical Finance, 36(C), 162-180.

Cepni, O., Demirer, R., Gupta, R., and Pierdzioch, C. (Forthcoming). Political Geography and Stock Market Volatility: The Role of Political Alignment across Sentiment Regimes. Scottish Journal of Political Economy.

Demirer, R., Gupta, R., Lv, Z., and Wong, W-K. (2019). Equity Return Dispersion and Stock Market Volatility: Evidence from Multivariate Linear and Nonlinear Causality Tests. Sustainability, 11(2), 351.

Demirer, R., Gupta, R., and Pierdzioch, C. (2024). Realized stock-market volatility: Do industry returns have predictive value? In Computational Finance: A Scientific Approach to Financial Analysis, Edited: L.A. Maglaras, S. Das, N. Tripathy, and S. Patnaik, Vol. 254, Chapter 18, Berlin, Germany: Springer, 377-405.

Detzel, A., Duarte, J., Kamara, A., Siegel, S., and Sun, C. (2023). The Cross-Section of Volatility and Expected Returns: Then and Now. Critical Finance Review, 12(1-4), 9-56.

Diebold, F. X. and Mariano, R.S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics, 13(3), 253-263.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. The Review of Financial Studies, 33(5), 2223-2273.

Gupta, R., Ji, Q., Pierdzioch, C., and Plakandaras, V. (2023b). Forecasting the conditional distribution of realized volatility of oil price returns: The role of skewness over 1859 to 2023. Finance Research Letters, 58(Part C), 104501.

Gupta, R., Nel, J., and Pierdzioch, C. (2023a). Investor confidence and forecastability of US stock market realized volatility: Evidence from machine learning. Journal of Behavioral Finance, 24(1), 111-122.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. Annals of Statistics, 3(5), 1163-1173.
Hong, H., Torous, W., and Valkanov, R. (2007). Do industries lead stock markets? Journal of Financial Economics, 83(2), 367-396.
Hwang S., Satchell S.E. (2005). GARCH model with cross-sectional volatility: GARCHX models. Applied Financial Economics, 15(3), 203-216.
Lu, F., and Ma, F. (2023). Cross-sectional uncertainty and stock market volatility: New evidence. Finance Research Letters, 57(C), 104202.
Lu, F., Ma, F., Guo, Q. (2023). Less is more? New evidence from stock market volatility predictability. International Review of Financial Analysis, 89(C), 102819.
Mei, D., Liu, J., Ma, F., and Chen, W. (2017). Forecasting stock market volatility: Do realized skewness and kurtosis help?, Physica A: Statistical Mechanics and Its Applications, 481(C), 153-159.
Plakandaras, V., Pragidis, I. and Karypidis, P. (2024) Deciphering the US metropolitan house price dynamics. Real Estate Economics, 52(2), 434-485.
Poon, S-H., and Granger, C.W.J. (2003). Forecasting volatility in financial markets: A review. Journal of Economic Literature, 41(2), 47-539.
Niu, Z., Demirer, R., Suleman, M.T., Zhang, H. (2023). Cross-sectional return dispersion and stock market volatility: Evidence from high-frequency data. Journal of Forecasting, 42(6), 1309-1328.
Niu, Z., Demirer, R., Suleman, M.T., Zhang, H., and Zhu, X. (2024). Do industries predict stock market volatility? Evidence from machine learning models. Journal of International Financial Markets, Institutions and Money, 90(C), 101903.
Rapach, D.E., Strauss, J.K., and Wohar, M.E. (2008). Forecasting stock return volatility in the presence of structural breaks. In Forecasting in the Presence of Structural Breaks and Model Uncertainty, Edited: D.E. Rapach and M.E. Wohar, Vol. 3 of Frontiers of Economics and Globalization, Bingley, United Kingdom: Emerald, 381-416.
Salisu, A.A., Gupta, R., and Ogbonna, A.E. (2022). A moving average heterogeneous autoregressive model for forecasting the realized volatility of the US stock market: Evidence from over a century of data. International Journal of Finance and Economics, 27(1), 384-400.
Salisu, A.A., Gupta, R., Cepni, O., and Caraiani, P. (2024). Oil shocks and state-level stock market volatility of the United States: a GARCH-MIDAS approach. Review of Quantitative Finance and Accounting, 63(4), 1473-1510.
Salisu, A.A., Liao, W., Gupta, R., and Cepni, O. (2025). Economic Conditions and Predictability of US Stock Returns Volatility: Local Factor Versus National Factor in a GARCH-MIDAS Model. Journal of Forecasting. DOI: https://doi.org/10.1002/for.3251.
Salisu, A.A., Ogbonna, A.E., Gupta, R., and Cepni, O. (Forthcoming). The Role of Energy Market Uncertainties in Forecasting US State-Level Stock Market Volatility: A GARCH-MIDAS Approach. Financial Innovation.
Souropanis, I., and Vivian, A. (2023). Forecasting realized volatility with wavelet decomposition. Journal of Empirical Finance, 74(C), 101432.
Tsuji, C. (2012). Do industries contain predictive information for the Fama-French factors? Quantitative Finance, 12(6), 969-991.
Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer, New York.
Zhang, Z., He, M., Zhang, Y., and Wang, Y. (2021). Realized skewness and the short-term predictability for aggregate stock market volatility. Economic Modelling, 103(C), 105614.

Table 1: Average out-of-sample RMSE

Horizon	TWFE	LASSO	SVR	ANN	RF	XGBoost	Panel-SVR
Panel A: All industry return moments, lagged volatility [U.S. industries]							
1	0.587	0.661***	0.461**	0.734***	0.572***	0.658**	0.351***
3	0.716	0.664**	0.554**	0.763**	0.687**	0.779**	0.377***
6	0.777	0.668**	0.610**	0.789**	0.758**	0.847**	0.253***
12	0.821	0.677*	0.667*	0.784**	0.796**	0.885**	0.343***
24	0.830	0.684*	0.665*	0.784*	0.805*	0.885*	0.245***
Panel B: All industry return moments, lagged volatility, WTI realized volatility [U.S. industries]							
1	4.395	0.661***	0.520**	0.791***	0.561***	0.654***	0.702***
3	0.789	0.662**	0.591	0.821	0.647***	0.728***	0.715***
6	0.777	0.664**	0.616**	0.819**	0.714**	0.794*	0.203***
12	0.822	0.667*	0.644*	0.826*	0.739**	0.811**	0.202***
24	0.830	0.672*	0.663	0.828*	0.757	0.826	0.309***
Panel C: All industry return moments, lagged volatility, WTI all moments [U.S. industries]							
1	8.604	0.661***	1.121	0.866***	0.540***	0.583***	0.332***
3	0.716	0.664**	0.639**	0.871***	0.739***	0.814**	0.418***
6	0.777	0.666**	0.678**	0.903*	0.791**	0.870**	0.467***
12	0.821	0.670*	0.703*	0.897**	0.836*	0.910*	0.356***
24	0.830	0.675*	0.951	0.897**	0.845	0.919	0.248***
Panel D: All stock return moments, lagged volatility [U.S. state stocks]							
1	3.198	1.004***	1.012	1.160***	0.998***	1.102***	0.383***
3	6.420	0.995***	0.944***	1.113***	1.032***	1.127***	0.256***
6	4.211	0.973***	0.971***	1.189***	1.132***	1.191***	0.472***
12	4.332	0.978**	1.033***	1.146***	1.214**	1.353**	0.278***
24	13.866	0.934**	1.040	1.061*	1.088*	1.225*	0.315***
Panel E: All stock return moments, lagged volatility, WTI realized volatility [U.S. state stocks]							
1	5.160	0.995***	0.877***	1.139**	0.978***	1.116***	0.468***
3	2.119	0.995***	0.945***	1.127***	1.100***	1.186***	0.310***
6	2.209	0.973***	0.929***	1.109***	1.136***	1.216***	0.395***
12	2.862	0.978**	0.993**	1.101***	1.221***	1.290***	0.313***
24	5.625	0.934**	0.950**	1.062**	1.164	1.278	0.345***
Panel F: All stock return moments, lagged volatility, WTI all moments [U.S. state stocks]							
1	9.636	0.995***	1.615	1.269***	0.990***	1.113***	0.322***
3	69.141	0.995***	1.359*	1.242***	1.046***	1.172***	0.268***
6	15.892	0.988***	1.248***	1.216***	1.073***	1.195***	0.466***
12	13.872	0.985**	1.196***	1.206***	1.131**	1.253**	0.362***
24	23.231	0.976**	1.151**	1.192**	1.147	1.266*	0.363***

Note: *,** and *** denote rejection of equal forecasting ability of the examined model compared to the TWFE model based on the Diebold and Mariano (1995) test in 10%, 5% and 1% level of significance, respectively.

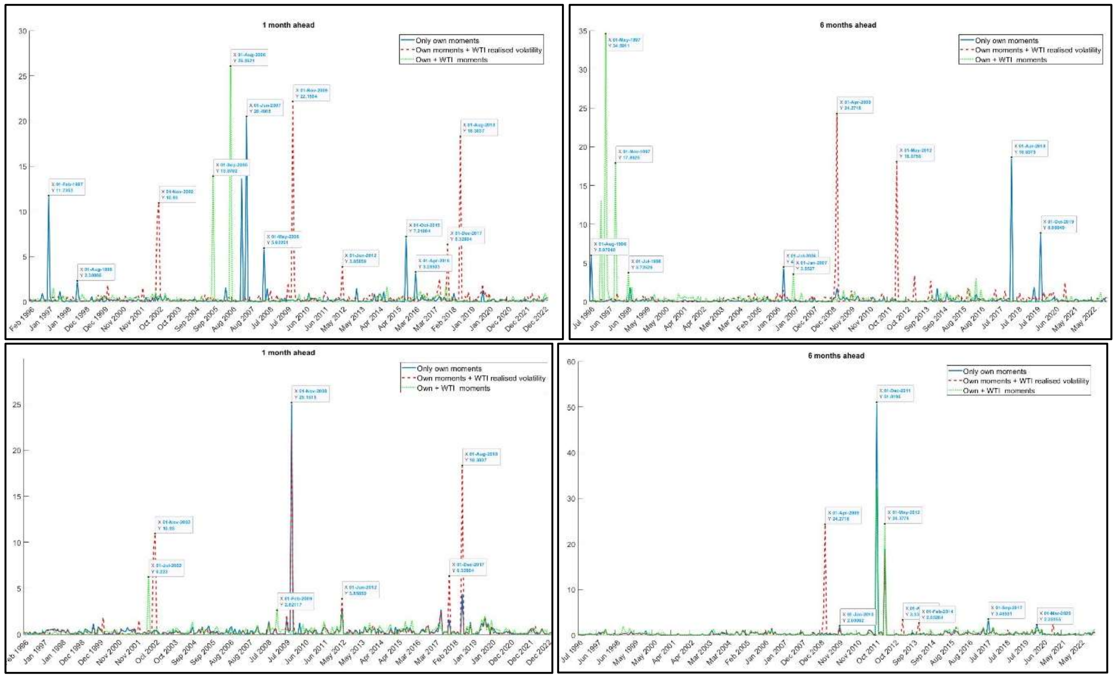

Figure 1: Average RMSEs for U.S. industries at the monthly (top left) and semi-annual horizon, and U.S. states (top right) and (bottom left) and semi-annual (bottom right) horizon, respectively.