Journal of Applied Econometrics

WILEY

Journal of APPLIED ECONOMETRICS

RESEARCH ARTICLE OPEN ACCESS

Modelling Volatility Cycles: The MF2-GARCH Model

Christian Conrad[1,2,3,4] | Robert F. Engle[5]

[1]Alfred-Weber-Institute, Heidelberg University, Heidelberg, Germany | [2]HEiKA - Heidelberg Karlsruhe Strategic Partnership, Heidelberg University, Karlsruhe Institute of Technology, Karlsruhe, Germany | [3]KOF Swiss Economic Institute, Zurich, Switzerland | [4]ZEW Mannheim, Mannheim, Germany | [5]Stern School of Business, New York University, New York, USA

Correspondence: Christian Conrad (christian.conrad@awi.uni-heidelberg.de)

Received: 6 March 2023 | **Revised:** 22 October 2024 | **Accepted:** 15 December 2024

Funding: This study was funded by the German Federal Ministry of Education and Research (BMBF) and the Baden-Württemberg Ministry of Science as part of Germany's Excellence Strategy (ExU 10.2.31).

Keywords: long- and short-term volatility | long-term forecasting | mixed frequency data | volatility component models | volatility forecasting

ABSTRACT

We propose a novel multiplicative factor multi-frequency GARCH (MF2-GARCH) model, which exploits the empirical fact that the daily standardized forecast errors of one-component GARCH models are predictable by a moving average of past standardized forecast errors. In contrast to other multiplicative component GARCH models, the MF2-GARCH features stationary returns, and long-term volatility forecasts are mean-reverting. When applied to the S&P 500, the new component model significantly outperforms the one-component GJR-GARCH, the GARCH-MIDAS-RV, and the log-HAR model in long-term out-of-sample forecasting. We illustrate the MF2-GARCH's scalability by applying the new model to more than 2100 individual stocks in the Volatility Lab at NYU Stern.

1 | Introduction

There is strong empirical evidence that the conditional variance of stock returns consists of several components. Early evidence for volatility components was provided in, for example, Ding and Granger 1996 and Engle and Lee 1999. More recent evidence can be found in Christoffersen et al. 2008, Kim and Nelson 2013, Dorion 2016, and Conrad and Kleen 2020, among others. While the GARCH models of Ding and Granger 1996 and Engle and Lee 1999 have additive volatility components, more recent GARCH-type models decompose the conditional variance into multiplicative short- and long-term components. For example, in the Spline-GARCH model of Engle and Rangel 2008 and the multiplicative time-varying GARCH (MTV-GARCH) of Amado and Teräsvirta 2013 and Amado and Teräsvirta 2017, the long-term volatility component is a deterministic function of calendar time. In contrast, in the GARCH-MIDAS of Engle, Ghysels, and Sohn 2013, the long-term component depends either on a rolling window realized variance (henceforth GARCH-MIDAS-RV, see also Wang and Ghysels 2015) or on low-frequency macroeconomic or financial variables (henceforth GARCH-MIDAS-X, see also Asgharian, Hou, and Javed 2013, and Conrad and Loch 2015). Those multiplicative volatility models are based on the idea that returns follow a stationary GARCH process once divided by the long-term volatility component. However, there is no consensus yet on the most suitable approach for modeling the long-term component.

We propose a novel specification for the long-term volatility component in multiplicative GARCH models. The specification is motivated by a new empirical fact that we document for the volatility forecast errors of one-component GARCH models: While the daily standardized forecast errors are essentially unpredictable based on past *daily* standardized forecast errors, a *rolling window moving average* of the past daily standardized forecast errors does have predictive power. This is

because one-component models tend to underpredict or overpredict volatility for *extended* time periods. While overprediction typically happens during economic expansions, underprediction materializes during economic recessions and other crises periods.

The new model combines a short-term GJR-GARCH (see Glosten, Jagannathan, and Runkle 1993) component with a long-term component specified as a multiplicative error model (MEM) for the past forecast errors of the GARCH component. That is, the long-term component exploits the predictability in the averaged standardized forecast errors of the short-term component. Intuitively, the long-term component scales the GARCH component's volatility forecast up/down if the short-term component's forecasts have underestimated/overestimated volatility in the recent past, that is, the model is learning from past forecast errors. Because the long-term component can either evolve at the same frequency as the short-term component or at a lower frequency, the new specification belongs to the class of mixed frequency data sampling (MIDAS) models pioneered by Ghysels, Santa-Clara, and Valkanov 2004. We refer to the proposed specification as **M**ultiplicative **F**actor **M**ulti-**F**requency GARCH. As the "**MF**" appears twice, we abbreviate the model as MF2-GARCH.

The properties of the MF2-GARCH model clearly distinguish it from previous specifications. First, while in other multiplicative models, there is typically no feedback from the short-term to the long-term component (e.g., in the Spline-GARCH), the MF2-GARCH explicitly specifies the long-term component as a function of the short-term component's past forecast errors. Second, because the actual economic drivers of long-term volatility are unknown and may vary over time, it is challenging to correctly specify the long-term component in the GARCH-MIDAS-X in real time. Our specification avoids this problem and is based on a simple MEM equation for the long-term component. Interestingly, the MF2-GARCH can be rewritten as a GARCH-MIDAS-X with an explanatory variable "generated within the model." Third, because the MF2-GARCH is dynamically complete, that is, it fully specifies the dynamics of the conditional variance, it is straightforward to construct multistep ahead volatility forecasts.

We obtain the following theoretical results for the MF2-GARCH: First, we derive the unconditional variance of the daily returns. While the unconditional variance is time-varying in the Spline-GARCH and infinite in the GARCH-MIDAS-RV of Wang and Ghysels 2015, returns are covariance stationary in the MF2-GARCH. Due to the feedback between the short- and long-term components, the unconditional variance depends not only on the model parameters but also on the fourth moment of the innovation. Second, we obtain the news impact curve (NIC). The NIC illustrates that the responsiveness to news changes with the level of volatility which is another essential feature that distinguishes the MF2-GARCH from other models in the GARCH family. Specifically, conditional volatility is more responsive to news during low volatility periods than during high volatility periods. Third, we derive expressions for multistep ahead forecasts of conditional volatility. Our results show that the forecasts are much more flexible than forecasts from the nested GARCH model. The forecasts reflect the current stance of the conditional variance and the prevailing volatility regime. In the short term, the conditional volatility forecast will approach the forecast of the long-term component before it converges to the unconditional volatility in the long run. Forecasts from the MF2-GARCH also differ from forecasts of standard Spline-GARCH or GARCH-MIDAS models. The forecasts of the latter models are typically assumed to converge to the current level of the long-term component. Thus, the forecasts from these models do not feature mean reversion in the long run. Forth, we discuss the quasi-maximum likelihood estimation of the MF2-GARCH and provide a Monte Carlo simulation showing that standard asymptotic results lead to valid inference. Finally, we provide an analysis of the degree of misspecification of the nested one-component GJR-GARCH and the GARCH-MIDAS-RV when the true data-generating process is an MF2-GARCH. As discussed in Patton 2020, in the presence of estimation error and depending on the employed loss function, a parsimoniously misspecified model might dominate the true but more complex model in terms of forecast performance. We show by simulations that for reasonable parameter values of the MF2-GARCH, the degree of misspecification of the GJR-GARCH and the GARCH-MIDAS-RV is so severe that the MF2-GARCH outperforms both models when evaluated by the squared error (SE) and the QLIKE loss.

Our empirical results strongly support the MF2-GARCH. First, we estimate the MF2-GARCH for the S&P 500 and 2142 US and international equities in the Volatility Laboratory (V-Lab) at NYU Stern.[1] Our in-sample results show that the MF2-GARCH is clearly preferred to the nested one-component GJR-GARCH, the GARCH-MIDAS-RV and the Spline-GARCH. For the S&P 500, we show that the MF2-GARCH's estimated long-term component is closely related to news about the macroeconomy and monetary policy, particularly news about inflation and interest rates. Thus, we provide further evidence for the close link between economic conditions and long-term volatility (see, e.g., Engle, Ghysels, and Sohn 2013, and Conrad and Loch 2015). We also illustrate that the MF2-GARCH's volatility forecasts, which feature cyclical behavior, are much more flexible than the forecasts of the competitor models. In contrast to the (overly) smooth long-term component of the Spline-GARCH, the MF2-GARCH's long-term component adjusts in response to short-lived periods of market turmoil. Nevertheless, compared to the one-component GJR-GARCH, the MF2-GARCH's forecasts avoid overestimating volatility after a short-lived surge in volatility due to the low persistence in its short-term component.

While most of the literature on volatility forecasting focuses on short-term (e.g., 1-day ahead) prediction horizons, Christoffersen and Diebold 2000, Engle 2009b, and Ghysels et al. 2019, among others, have highlighted that in many areas of finance long-term risk forecasts are the relevant inputs. This leads to the question of how far ahead into the future we can forecast volatility. Hence, in evaluating the out-of-sample forecast performance of the MF2-GARCH, our focus is on medium- and long-term forecast horizons of up to 8 months. We test whether the MF2-GARCH, which is designed to capture volatility cycles, leads to better long-term predictions than the nested GJR-GARCH, the GARCH-MIDAS-RV, the Spline-GARCH, and Corsi and Reno 2012's log-HAR with leverage. For the S&P 500, it turns out that the MF2-GARCH outperforms all competitor models when the forecast horizon is beyond 2 months. The MF2-GARCH's forecast performance is particularly strong during periods of high volatility, where it dominates the competitor models at all forecast horizons. For a cross-section of 20

equities, out-of-sample results from the V-Lab confirm that the MF2-GARCH strongly outperforms the competitor models.

The paper is organized as follows. In Section 2, we show that the volatility forecast errors of the GJR-GARCH are predictable. We introduce the MF2-GARCH and discuss its properties in Section 3. The empirical results are presented in Section 4 and Section 5 concludes. Further model details, proofs, and additional tables and figures can be found in the Supporting Information.

2 | A New Empirical Fact of Volatility Forecast Errors

In this section, we provide evidence for a new empirical fact of volatility forecast errors: Rolling window moving averages of the standardized forecast errors of one-component GARCH models behave counter-cyclically and have predictive power for future standardized forecast errors.

We denote the log-return on day t by r_t. The conditional heteroskedasticity in daily stock returns is commonly modeled as a GARCH process. Daily stock returns are written as

$$r_t = \sqrt{\tilde{h}_t}\zeta_t, \tag{1}$$

where $\tilde{h}_t$ denotes the conditional variance and the ζ_t are assumed to be $i.i.d.$ with mean and variance equal to zero and one, respectively. For illustration, we estimate a GJR-GARCH(1,1) specification for a long time series of daily S&P 500 log-returns covering January 1971 to June 2023.[2] We obtain the following result:

$$\tilde{h}_t = \underset{(0.003)}{0.018} + (\underset{(0.006)}{0.022} + \underset{(0.014)}{0.116}\,\mathbf{1}_{\{r_{t-1}<0\}})\, r_{t-1}^2 + \underset{(0.008)}{0.905}\,\tilde{h}_{t-1} \tag{2}$$

where the numbers in parentheses are Bollerslev–Wooldridge robust standard errors and $\mathbf{1}_{\{r_{t-1}<0\}}$ equals one if $r_{t-1} < 0$, and zero else. As expected, the conditional variance is highly persistent and there is strong evidence for asymmetry.

Several test statistics have been proposed to check a GARCH specification's adequacy. For example, Engle and Ng 1993 and Halunga and Orme 2009 propose Lagrange Multiplier (LM) tests for the null hypothesis that a (GJR-)GARCH(1,1) is correctly specified.

An alternative approach is to check whether Equation (1) is misspecified in the sense that $\zeta_t = \sqrt{\tau_t}Z_t$, where the Z_t are $i.i.d.$ and τ_t represents an omitted multiplicative long-term volatility component. The long-term component evolves either at the same frequency as the daily returns or at a lower (e.g., monthly or quarterly) frequency. The daily returns, r_t, can be either stationary or nonstationary. For example, in the Spline-GARCH model τ_t evolves at the daily frequency and—because the long-term component is a deterministic function of time—the daily returns have a time-varying unconditional second moment. In either case, the scaled returns, $r_t/\sqrt{\tau_t}$, are assumed to follow a stationary GARCH process. LM tests for an omitted τ_t component have been proposed in Lundbergh and Teräsvirta 2002 and Amado and Teräsvirta 2017 for daily long-term components. The LM test of Conrad and Schienle 2020 allows for explanatory variables in the long-term component and either a daily or lower frequency τ_t.

The tests of Lundbergh and Teräsvirta 2002 and Conrad and Schienle 2020 exploit that the squared standardized errors, $\zeta_t^2 = r_t^2/\tilde{h}_t$, are $i.i.d.$ under the null hypothesis of a constant long-term component. The Conrad and Schienle 2020 LM test checks whether ζ_t^2 is predictable by x_{t-1}, x_{t-2}, …, x_{t-K}, where x_t is a predictor variable that can be exogenous or "generated within the model."[3] Under the null hypothesis, the LM test is χ^2 distributed with K degrees of freedom.

We propose to use the m-days rolling window average of the squared standardized errors,

$$\tilde{V}_{t-1}^{(m)} = \frac{1}{m}\sum_{j=1}^{m}\frac{r_{t-j}^2}{\tilde{h}_{t-j}}, \tag{3}$$

as the predictor variable. Under the null hypothesis, r_t^2 is a conditionally unbiased proxy of the true but unobservable conditional variance and $\tilde{h}_t$ is a one-step-ahead forecast for the same quantity. If the GARCH model is correctly specified, the standardized volatility forecast errors, $r_t^2/\tilde{h}_t$, have an expected value of one and a variance of two if ζ_t is Gaussian. Hence, we think of $\tilde{V}_t^{(m)}$ as a measure of the *local bias* of the GARCH conditional variance. For $m = 1$, we obtain $\tilde{V}_{t-1}^{(1)} = \zeta_{t-1}^2$, which is the predictor variable in the "ARCH nested in GARCH" test of Lundbergh and Teräsvirta 2002.

Figure 1 shows $\tilde{V}_t^{(m)}$ based on the conditional variances from the GJR-GARCH in Equation (2) for $m \in \{1, 15, 25, 45\}$. In the upper right and both lower panels, it is visible that, as expected, $\tilde{V}_t^{(m)}$ fluctuates around the value of one. However, as the two lower panels show, there are extended periods during which the one-component GARCH model underestimates or overestimates volatility. That is, for $m = 25$ and $m = 45$, the evolution of $\tilde{V}_{t-1}^{(m)}$ is in line with a local bias of the GJR-GARCH conditional variance. The local bias appears to be counter-cyclical: The one-component GARCH model tends to overestimate volatility during expansions and to underestimate it during recessions. For $m = 1$, $\tilde{V}_t^{(m)}$ is too noisy to reveal this bias. There are also some spikes in $\tilde{V}_t^{(m)}$. These spikes occur due to extraordinary events with unexpectedly high volatility. For example, the two largest spikes are due to the stock market crashes on October 19, 1987 ("Black Monday") and October 13, 1989 ("Mini-Crash"). The spike in March 2020 is due to the emergence of the Covid-19 pandemic and the spike on September 14, 2022 due to the release of higher-than-expected inflation numbers.

In Panel B of Table 1, we formally test whether $\tilde{V}_{t-1}^{(m)}$ has predictive power for ζ_t^2 using the Conrad and Schienle 2020 LM test with $K = 1$. For $m \in \{1, 5, 15\}$, the null hypothesis of a constant long-term component is not rejected. When the true long-term component smoothly varies over time, this is to be expected because for small m, $\tilde{V}_{t-1}^{(m)}$ is too noisy to have explanatory power for ζ_t^2. In contrast, for $m \in \{25, 35, 45, 55\}$, we strongly reject the null hypothesis. Thus, the LM test provides evidence for an omitted long-term component and suggests that $\tilde{V}_{t-1}^{(m)}$ is suitable for modeling the dynamics of the long-term component when m is appropriately chosen.

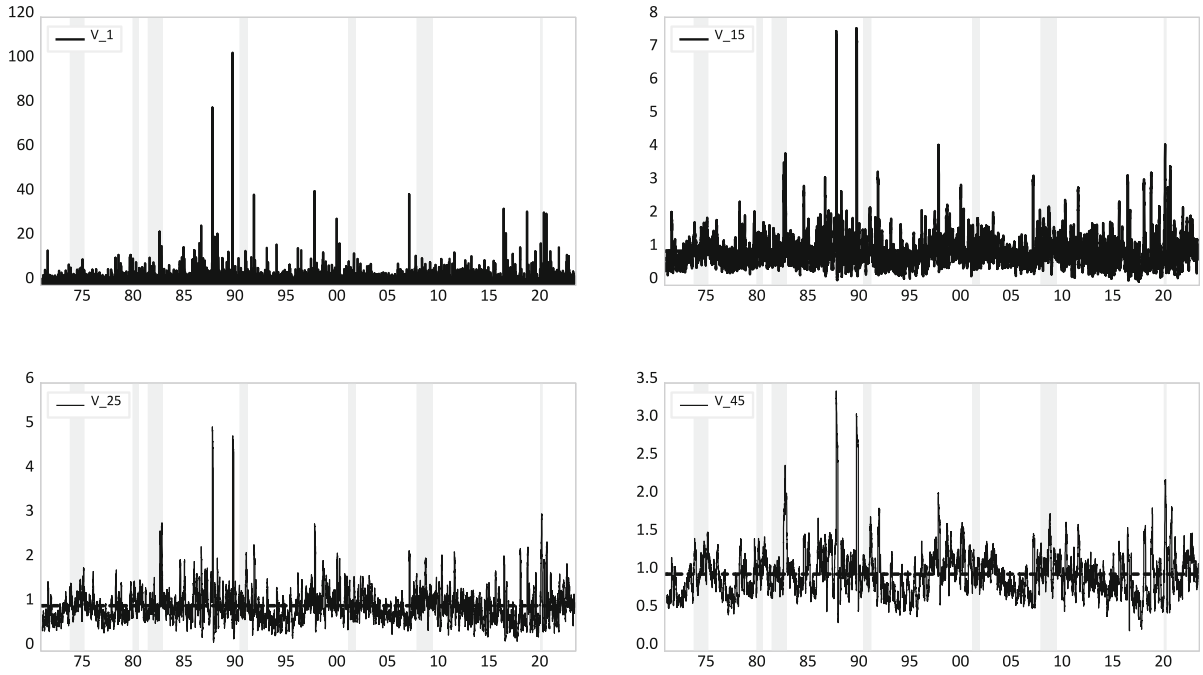

FIGURE 1 | A GJR-GARCH(1,1) is estimated for daily S&P 500 return data for the January 1971 to June 2023 period. The figure shows $\widetilde{V}_t^{(m)}$ for $m = 1$ (upper left panel), $m = 15$ (upper right panel), $m = 25$ (lower left panel), and $m = 45$ (lower right panel). Gray shaded areas represent NBER recession periods.

TABLE 1 | Summary statistics S&P 500 and LM test.

	Mean	SD	Skewness	Kurtosis	Min	Max	AC(1)
				Panel A: Summary statistics			
r_t	0.03	1.09	−1.00	27.11	−22.93	10.71	−0.02
RV_t	0.99	2.94	18.01	486.49	0.02	101.29	0.63
				Panel B: LM test: Explanatory variable $\widetilde{V}_t^{(m)}$			
m	1	5	15	25	35	45	55
p-value	0.930	0.960	0.450	0.030	0.001	0.001	0.010

Note: Panel A shows summary statistics for the daily returns, r_t, and the daily realized variances, RV_t, of the S&P 500. The columns present the mean, the standard deviation (sd), skewness, kurtosis, the minimum (min) and maximum (max) as well as the first-order autocorrelation coefficient (AC(1)). Daily returns for the S&P 500 cover the period January 1971 to June 2023. Realized variances are for the period January 2010 to June 2023. Panel B shows the results of the Conrad and Schienle 2020 LM test for an omitted long-term component under the null hypothesis of a one-component GJR-GARCH. We set $K = 1$. The table shows the p-values of the test for different choices of m.

Importantly, the behavior of the standardized volatility forecast errors is not specific to the one-component GJR-GARCH. We also estimated EGARCH (Nelson 1991), FIGARCH (Baillie, Bollerslev, and Mikkelsen 1996), and Realized GARCH (Hansen, Huang, and Shek 2012) models and obtained very similar results. For example, the correlation between the $\tilde{V}_{t-1}^{(45)}$ of the GJR-GARCH and the $\tilde{V}_{t-1}^{(45)}$ of the EGARCH, FIGARCH, and the Realized GARCH is 0.92, 0.82, and 0.77, respectively. Figure A.1 in the Supporting Information plots $\tilde{V}_{t-1}^{(45)}$ for all four models and confirms that there is strong co-movement. Furthermore, our findings do not only hold for the S&P 500 but also for other international stock indices. For illustration, Figure A.2 in the Supporting Information replicates Figure 1 for the FTSE 100.[4]

In summary, the evidence suggests that one-component GARCH models are misspecified and that the misspecification is detectable when using suitable moving averages of past standardized forecast errors to predict the current standardized forecast error.

3 | The MF2-GARCH Model

This section introduces the MF2-GARCH model. In the main specification, the short- and the long-term components evolve at a daily frequency. For this specification, we derive the unconditional variance of returns, the NIC and multistep ahead forecasts. In Section 3.2, we suggest several directions in which the MF2-GARCH can be extended. In particular, we introduce a parametrization that allows for multiple frequencies, that is, the short-term component evolves at the daily frequency, while the long-term component varies at a lower frequency. Further details

on the MF2-GARCH are provided in the Supporting Information: Section A.1 discusses quasi-maximum likelihood estimation and provides a Monte Carlo simulation. The degree of misspecification of the nested one-component GJR-GARCH and the GARCH-MIDAS-RV when the true model is an MF2-GARCH is analyzed in Section A.2 in the Supporting Information. A comparison of the MF2-GARCH with other component models is provided in Section A.3 in the Supporting Information.

In general and as motivated in Section 2, daily log-returns are defined as $r_t = \sigma_t Z_t = \sqrt{h_t \tau_t} Z_t$. We denote the information set on day t by $\mathcal{F}_t$. σ_t^2 denotes the conditional variance and the short- and long-term volatility components are given by h_t and τ_t. We make the following assumption about the innovations Z_t.

Assumption 1. Let Z_t be i.i.d. The density of Z_t is symmetric with $\mathbf{E}[Z_t] = 0$ and $\mathbf{E}[Z_t^2] = 1$. Further, Z_t^2 has a nondegenerate distribution and $\kappa = \mathbf{E}[Z_t^4] < \infty$.

The assumption that the density of Z_t is symmetric is commonly made for GJR-GARCH models because it allows for a straightforward computation of the multistep ahead conditional variance forecast (see, e.g., Zivot 2009). Also, Ling and McAleer 2002 make this assumption when deriving conditions for the stationarity and the existence of the fourth moment of the GJR-GARCH. Importantly, as shown in Alexander, Lazar, and Stanescu (2021), the symmetry of the density of Z_t does not preclude that the s-step ahead aggregated returns exhibit skewness. The assumption that $\kappa = \mathbf{E}[Z_t^4] < \infty$ is a necessary condition for ensuring the finiteness of the unconditional variance of the returns and for the existence of the variance of the score of the likelihood function.

3.1 | Daily Short- and Long-Term Components

We specify the short-term volatility component as a unit variance GJR-GARCH(1,1)

$$h_t = (1 - \phi) + \left(\alpha + \gamma \mathbf{1}_{\{r_{t-1}<0\}}\right) \frac{r_{t-1}^2}{\tau_{t-1}} + \beta h_{t-1}, \tag{4}$$

where $\phi = \alpha + \gamma/2 + \beta$. Note that the driving variable in Equation (4) is r_{t-1}^2/τ_{t-1}. This distinguishes h_t from the daily conditional variance, $\tilde{h}_t$, in Equation (2). We make the following assumption about the parameters of the short-term component:

Assumption 2. The parameters of the short-term GJR-GARCH component satisfy the conditions $\alpha > 0$, $\alpha + \gamma > 0$, $\beta > 0$ and $\phi = \alpha + \gamma/2 + \beta < 1$.

If Assumptions 1 and 2 hold, then $r_t/\sqrt{\tau_t} = \sqrt{h_t} Z_t$ follows a covariance stationary GJR-GARCH(1,1) with $\mathbf{E}[h_t Z_t^2] = \mathbf{E}[h_t] = 1$. If the GJR-GARCH(1,1) fully captures the conditional heteroskedasticity, then τ_t is equal to a constant and the multiplicative model reduces to a one-component GJR-GARCH for the daily returns.

Following Engle 2009a, we refer to $r_t/\sqrt{h_t}$ as "deGARCHed returns" and define $V_t = r_t^2/h_t = \tau_t Z_t^2$ as the squared deGARCHed returns. If the GARCH component fully captures the conditional heteroskedasticity, then by Assumption 1 the V_t are *i.i.d.* However, if τ_t is time-varying and persistent, then V_t is autocorrelated. Because V_t is a nonnegative variable, we specify the long-term component as a MEM equation for the conditional expectation of V_t:

$$\tau_t = \lambda_0 + \lambda_1 V_{t-1}^{(m)} + \lambda_2 \tau_{t-1}, \tag{5}$$

where

$$V_{t-1}^{(m)} = \frac{1}{m}\sum_{j=1}^{m} V_{t-j} = \frac{1}{m}\sum_{j=1}^{m} \frac{r_{t-j}^2}{h_{t-j}}. \tag{6}$$

Recall that we can think of the squared deGARCHed returns as standardized volatility forecast errors. Hence, $V_{t-1}^{(m)}$ is a *rolling window* measure of the local bias of the short-term component's conditional variance over the previous m days. Note that Equation (5) can be written is a MEM$(1, m)$ with the restriction that the m "ARCH" coefficients are given by λ_1/m. The MEM$(1, m)$ is covariance stationary if the sum of the ARCH and GARCH coefficients is less than one. By construction, this is satisfied if $\lambda_1 + \lambda_2 < 1$.

Assumption 3. The parameters of the long-term component satisfy the conditions $\lambda_0 > 0$, $\lambda_1 > 0$, $\lambda_2 > 0$ and $\lambda_1 + \lambda_2 < 1$.

Under Assumptions 1 and 3, it holds that $V_t = \tau_t Z_t^2$ is a covariance stationary MEM with $\mathbf{E}[V_t|\mathcal{F}_{t-1}] = \tau_t$ and $\mathbf{E}[V_t] = \lambda_0/(1 - \lambda_1 - \lambda_2)$. We refer to the parametrization given by Equations (4) and (5) as MF2-GARCH-rw-m, where "rw-m" stands for *rolling window* of length m.

3.1.1 | Unconditional Variance of Daily Returns

Next, we derive the unconditional variance of the daily returns. First, note that the unconditional mean and variance are given by $\mathbf{E}[r_t] = 0$ and $\mathbf{Var}[r_t] = \mathbf{E}[r_t^2] = \mathbf{E}[\tau_t h_t]$. The following theorem provides an expression for $\mathbf{Var}[r_t]$.

Theorem 1. *Let Assumptions 1–3 be satisfied. If the MF2-GARCH-rw-m process, $(r_t)_{t\in\mathbb{Z}}$, is covariance stationary, then*

$$\Gamma_m = \left(\lambda_1 \frac{1}{m}\phi_\kappa + \lambda_2 \phi\right) + \lambda_1 \phi_\kappa \frac{1}{m}\sum_{j=2}^{m} \phi^{j-1} < 1 \tag{7}$$

with $\phi_\kappa = (\alpha + \gamma/2)\kappa + \beta$. The unconditional variance of the daily returns is given by

$$\mathbf{Var}[r_t] = \left(\lambda_0 + \frac{\lambda_0}{1 - \lambda_1 - \lambda_2}(1 - \phi)(\lambda_1 + \lambda_2) + \Delta_m\right)/(1 - \Gamma_m), \tag{8}$$

where

$$\Delta_m = (1 - \phi)\lambda_1 \phi \frac{\lambda_0}{1 - \lambda_1 - \lambda_2}\left(\frac{m-1}{m} + \frac{1}{m}\sum_{j=2}^{m}\sum_{k=1}^{j-2} \phi^k\right).$$

In the proof of Theorem 1, we show that $\Gamma_m < 1$ is satisfied if the returns are covariance stationary. For example, for $m = 1$, the condition reduces to $\Gamma_1 = \lambda_1[(\alpha + \gamma/2)\kappa + \beta] + \lambda_2[\alpha + \gamma/2 + \beta] < 1$. This illustrates that even if the conditions which ensure that $r_t/\sqrt{\tau_t}$ and $V_t = r_t^2/h_t$ are individually covariance

stationary (i.e., Assumptions 1–3) are satisfied, the condition in Equation (7) may be violated if κ is sufficiently large.

In general, the variance of the daily returns will depend on all the model parameters and the innovation's fourth moment, κ. The fact that the unconditional variance depends on κ distinguishes the MF2-GARCH-rw-m from standard GARCH models and is due to the correlation between τ_t and h_t. The unconditional variance of the returns increases in κ and decreases as m increases.[5] Theorem 1 reveals that the MF2-GARCH is fundamentally different from the Spline-GARCH and the GARCH-MIDAS-RV. In the Spline-GARCH, the unconditional variance of the returns is time-varying, and $\mathbf{Var}[r_t] = \infty$ in the GARCH-MIDAS-RV (see Wang and Ghysels 2015).

When imposing the restrictions $m = 1$ and $\gamma = 0$, we obtain a model that can be considered a "multiplicative version" of the additive component model of Engle and Lee 1999. For $m = 1$, both components are symmetric, and we impose the restriction $\alpha + \beta < \lambda_1 + \lambda_2 < 1$ to ensure that τ_t is the long-run component. In the following corollary, we derive the condition for the covariance stationarity of the daily returns when $m = 1$ and $\gamma = 0$ and state their unconditional variance.

Corollary 1. *Let Assumptions 1–3 be satisfied, $m = 1$, and $\gamma = 0$. The necessary and sufficient condition for the covariance stationary of the MF2-GARCH-rw-1 process is $\Gamma_1 = \lambda_1(\alpha\kappa + \beta) + \lambda_2(\alpha + \beta) < 1$. The unconditional variance of the returns is given by*

$$\mathbf{Var}[r_t] = \frac{\lambda_0 + \frac{\lambda_0}{1-\lambda_1-\lambda_2}(1-\alpha-\beta)(\lambda_1+\lambda_2)}{1-[\lambda_1(\alpha\kappa+\beta)+\lambda_2(\alpha+\beta)]}. \tag{9}$$

If the short-term component is constant (i.e., $\alpha = \beta = 0$), the expression in Equation (9) reduces to $\mathbf{Var}[r_t] = \mathbf{E}[\tau_t] = \lambda_0/(1 - \lambda_1 - \lambda_2)$. If the long-term component is constant (i.e., $\lambda_1 = \lambda_2 = 0$), the expression in Equation (9) reduces to $\mathbf{Var}[r_t] = \mathbf{E}[\lambda_0 h_t] = \lambda_0$. In the latter case, the MF2-GARCH-rw-1 reduces to a GARCH(1,1).

As expected, empirically we find that $V_{t-1}^{(1)} = r_{t-1}^2/h_{t-1}$ is too noisy to serve as a proxy of the local bias. In Section 4.2.1, we show that the optimal choice of m is around 63 for the S&P 500, which corresponds to a quarterly moving average.

3.1.2 | News Impact Curve

Following Engle and Ng 1993, we use the NIC to illustrate how the conditional volatility is updated in response to new information. For a GJR-GARCH(1,1) with conditional variance $\tilde{h}_t = \alpha_0 + (\alpha + \gamma\mathbf{1}_{\{r_{t-1}<0\}})r_{t-1}^2 + \beta\tilde{h}_{t-1}$, the NIC is defined as

$$NIC_{t+1}^{GJR} = \tilde{h}_{t+1}(r_t|\tilde{h}_t) = A_t^{GJR} + (\alpha + \gamma\mathbf{1}_{\{r_t<0\}})r_t^2,$$

where $A_t^{GJR} = \alpha_0 + \beta\tilde{h}_t$. That is, the NIC is a function of today's return and A_t^{GJR} is known conditional on $\mathcal{F}_{t-1}$. If $\gamma > 0$, negative news, $r_t < 0$, have a stronger effect on volatility than positive news. However, the size of the effect does not depend on the current level of volatility.

The NIC of the MF2-GARCH-rw-m consists of three terms:

$$\begin{aligned} NIC_{t+1}^{MF} &= \sigma_{t+1}^2(r_t|\tau_t, h_t) = A_t^{MF} + B_t^{MF} \\ &\quad + \left(\lambda_1\beta\frac{1}{m} + \lambda_2(\alpha + \gamma\mathbf{1}_{\{r_t<0\}})\right)r_t^2, \end{aligned} \tag{10}$$

where

$$\begin{aligned} A_t^{MF} &= \lambda_0(1-\phi) + \lambda_0\beta h_t + \lambda_1(1-\phi)\frac{1}{m}\sum_{j=1}^{m-1}\frac{r_{t-j}^2}{h_{t-j}} \\ &\quad + \lambda_1\beta h_t\frac{1}{m}\sum_{j=1}^{m-1}\frac{r_{t-j}^2}{h_{t-j}} + \lambda_2(1-\phi)\tau_t + \lambda_2\beta\sigma_t^2 \end{aligned}$$

and

$$\begin{aligned} B_t^{MF} &= \lambda_0(\alpha + \gamma\mathbf{1}_{\{r_t<0\}})\frac{r_t^2}{\tau_t} + \lambda_1(1-\phi)\frac{1}{m}\frac{r_t^2}{h_t} \\ &\quad + \lambda_1(\alpha + \gamma\mathbf{1}_{\{r_t<0\}})\frac{1}{m}\frac{r_t^4}{\sigma_t^2} + \lambda_1(\alpha + \gamma\mathbf{1}_{\{r_t<0\}})\frac{r_t^2}{\tau_t}\frac{1}{m}\sum_{j=1}^{m-1}\frac{r_{t-j}^2}{h_{t-j}}. \end{aligned}$$

The intercept, A_t^{MF}, is known conditional on $\mathcal{F}_{t-1}$. The last term depends on r_t^2 and model parameters and, hence, is similar to the term $(\alpha + \gamma\mathbf{1}_{\{r_t<0\}})r_t^2$ in the GJR-GARCH. However, the term B_t^{MF} shows that the marginal effect of r_t^2 also depends on the current conditional variance, σ_t^2, as well as h_t and τ_t individually.

Figure 2 shows the (standardized) NIC of the MF2-GARCH-rw-m with $m = 63$ (left panel) and $m = 21$ (right panel). The NIC is plotted as a function of the return, r_t, and for $\tau_t = 1$, $\tau_t = 1.5$, and $\tau_t = 0.5$. The parameters are chosen as in Figure A.3 in the Supporting Information. The short-term component is fixed at its unconditional expectation (i.e., $h_t = 1$). Note that parameters of the long-term component are chosen such that $\mathbf{E}[\tau_t] = 1$. That is, the solid black NIC represents a situation in which both the short- and long-term component are at their unconditional expectation. Due to the asymmetry term in the short-term component, the impact of negative returns is stronger than the impact of positive returns. The other two lines show that the effect of new information becomes stronger/weaker when long-term volatility is below/above its expectation. That is, the conditional volatility is more sensitive to news during a low-volatility period than during a high-volatility period. This is reasonable because a large value of $|r_t|$ is to be expected during turbulent times but less so during tranquil times. In line with the observation that the variance of the returns decreases as m increases, the news impact is slightly weaker for $m = 63$ than for $m = 21$.

3.1.3 | Forecasting Volatility

Because the MF2-GARCH-rw-m model is dynamically complete, we can analytically derive volatility forecasts for any desired horizon. In the following, we assume that a researcher has observed returns up to day t. Based on the information set $\mathcal{F}_t$, she intends to compute a volatility forecast for day $t + s$. First, recall that the s-step ahead forecast of the short-term component can be computed as $\mathbf{E}[h_{t+s}|\mathcal{F}_t] = 1 + \phi^{s-1}(h_{t+1} - 1)$, $s \geq 2$ (see, e.g., Zivot 2009). The forecasts for the long-term component are slightly more involved and are presented in Appendix A.4 of the

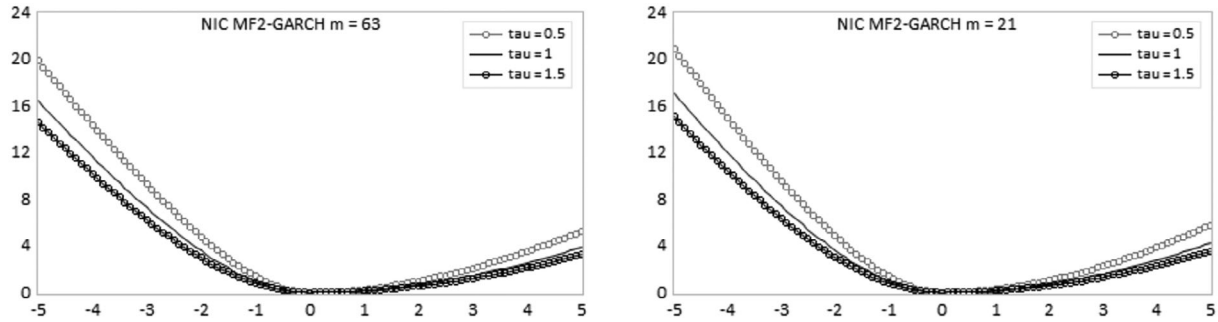

FIGURE 2 | The figure shows the NIC for an MF2-GARCH-rw-m model with $m = 63$ (left panel) and $m = 21$ (right panel) and parameters as in Figure A.3 in the Supporting Information. We fix $h_t = 1$ and assume that the short-term component correctly predicts volatility on days $t-1$ to $t-m-1$, that is, we set $r_{t-j}^2/h_{t-j} = 1$ for $j = 1, \dots, m-1$. The NICs are plotted as a function of the return, r_t, and for $\tau_t \in \{0.5, 1, 1.5\}$. The NICs are standardized such that news impact is zero for $r_t = 0$ and presented as annualized volatilities, that is, we plot $\sqrt{252(\sigma_{t+1}^2(r_t|\tau_t, h_t = 1) - \sigma_{t+1}^2(r_t = 0|\tau_t, h_t = 1))}$.

Supporting Information. Using these results, $\mathbf{E}[\sigma_{t+s}^2|\mathcal{F}_t]$ can be computed as follows:

Theorem 2. *Let Assumptions 1–3 and the constraint in Equation (7) be satisfied. Then, in the MF2-GARCH-rw-m the forecast of the conditional variance on day $t+s$, $s \geq 1$, can be computed as follows: First, $\mathbf{E}[\sigma_{t+1}^2|\mathcal{F}_t] = h_{t+1}\tau_{t+1}$. Second, for $s = 2, \dots, m$ the forecasts can be recursively calculated as*

$$\begin{aligned}\mathbf{E}[\sigma_{t+s}^2|\mathcal{F}_t] &= (1-\phi)\mathbf{E}[\tau_{t+s}|\mathcal{F}_t] + \lambda_0\phi\mathbf{E}[h_{t+s-1}|\mathcal{F}_t] \\ &+ \left(\lambda_1\frac{1}{m}\phi_\kappa + \lambda_2\phi\right)\mathbf{E}[\sigma_{t+s-1}^2|\mathcal{F}_t] \\ &+ \lambda_1\phi\mathbf{E}[h_{t+s-1}|\mathcal{F}_t]\frac{1}{m}\sum_{j=s}^{m}\frac{r_{t+s-j}^2}{h_{t+s-j}} \\ &+ (1-\phi)\lambda_1\phi\frac{1}{m}\sum_{j=2}^{s-1}\mathbf{E}[\tau_{t+s-j}|\mathcal{F}_t]\left(1+\sum_{k=1}^{j-2}\phi^k\right) \\ &+ \lambda_1\phi_\kappa\phi\frac{1}{m}\sum_{j=2}^{s-1}\phi^{j-2}\mathbf{E}[\sigma_{t+s-j}^2|\mathcal{F}_t].\end{aligned} \tag{11}$$

Third, for $s > m$, the following recursion applies:

$$\begin{aligned}\mathbf{E}[\sigma_{t+s}^2|\mathcal{F}_t] &= (1-\phi)\mathbf{E}[\tau_{t+s}|\mathcal{F}_t] + \lambda_0\phi\mathbf{E}[h_{t+s-1}|\mathcal{F}_t] \\ &+ \left(\lambda_1\frac{1}{m}\phi_\kappa + \lambda_2\phi\right)\mathbf{E}[\sigma_{t+s-1}^2|\mathcal{F}_t] \\ &+ (1-\phi)\lambda_1\phi\frac{1}{m}\sum_{j=2}^{m}\mathbf{E}[\tau_{t+s-j}|\mathcal{F}_t]\left(1+\sum_{k=1}^{j-2}\phi^k\right) \\ &+ \lambda_1\phi_\kappa\phi\frac{1}{m}\sum_{j=2}^{m}\phi^{j-2}\mathbf{E}[\sigma_{t+s-j}^2|\mathcal{F}_t].\end{aligned} \tag{12}$$

We will illustrate the behavior of the volatility forecasts in Section 4.2.2.

3.2 | Extensions of the MF2-GARCH-rw

3.2.1 | Modifications of the Daily Long-Term Component

MF2-GARCH with beta-weights: In Equation (6), we assume that $V_{t-1}^{(m)}$ is based on the average of the last m standardized forecast errors. Instead of imposing equal weights, we can take a weighted average of the form

$$V_{t-1}^{(m)} = \sum_{j=1}^{m} w_j(\omega)V_{t-j} = \sum_{j=1}^{m} w_j(\omega)\frac{r_{t-j}^2}{h_{t-j}}. \tag{13}$$

Following a common choice in the MIDAS literature (see Ghysels, Santa-Clara, and Valkanov 2006, and Ghysels, Sinko, and Valkanov 2007), we parsimoniously model the weights $w_j(\omega)$ according to a restricted beta-weighting scheme: $w_j(\omega) = ((1-j/(m+1))^{\omega-1})/(\sum_{k=1}^{m}(1-j/(m+1))^{\omega-1})$. By construction, the weights sum to one. In addition, we impose the constraint that the weights are nonincreasing ($\omega \geq 1$). For $\omega > 1$, the weights decline from the first lag. For $\omega = 1$, the weights are given by $1/m$, and hence, we obtain the model with rolling window $V_t^{(m)}$. We will refer to this parametrization as MF2-GARCH-bw-m, where "bw-m" stands for *beta-weights* of length m.

Realized volatility MEM: When m is small, the average standardized forecast error of the short-term component, $V_t^{(m)} = 1/m\sum_{j=1}^{m} r_{t-j}^2/h_{t-j}$, can be a noisy proxy for the local bias. Instead, if we observe daily realized variances, RV_t, we can base τ_t on the realized measure: $V_t^{(m,RV)} = 1/m\sum_{j=1}^{m} RV_{t-j}/h_{t-j}$. Again, we can apply the MEM specification from Equation (5). However, without further assumptions, this specification is no longer dynamically complete. We leave this specification for future research.

3.2.2 | Low-Frequency Long-Term Component

In the MF2-GARCH-rw-m, τ_t varies at the daily frequency. Instead, we can specify the MF2-GARCH so that the long-term component varies at a lower frequency. For this specification, we introduce a notation that allows for mixed frequencies. We distinguish between a low-frequency period t and a high-frequency period i. The high-frequency period i represents days while t

might represent a monthly, quarterly, or semiannual frequency. We assume that there are n days within each period t, that is, $i = 1, \dots, n$, and that we observe $t = 1, \dots, T$ low-frequency periods. Using this notation, we denote the log-return on day i of period t by $r_{i,t}$ (where we use the convention that $r_{0,t} = r_{n,t-1}$). Similarly, we denote the information set on day i in period t by $\mathcal{F}_{i,t}$ and define $\mathcal{F}_t := \mathcal{F}_{n,t}$. Note that the new notation reduces to the notation with a daily long-term component when $n = 1$. In the following, we assume that Assumption 1 holds for the innovations $Z_{i,t}$.

Using the notation for multiple frequencies, we write the short-term volatility component as

$$h_{i,t} = (1 - \phi) + (\alpha + \gamma \mathbf{1}_{\{r_{i-1,t}<0\}}) \frac{r^2_{i-1,t}}{\tau_t} + \beta h_{i-1,t} \tag{14}$$

for $i = 2, \dots, n$ and with $h_{0,t} = h_{n,t-1}$. Thus, for $i = 1$, we obtain $h_{1,t} = (1 - \phi) + (\alpha + \gamma \mathbf{1}_{\{r_{n,t-1}<0\}}) r^2_{n,t-1}/\tau_{t-1} + \beta h_{n,t-1}$. As before, we assume that Assumption 2 holds. We close the model by defining a MEM specification at the lower frequency. By averaging the squared deGARCHed returns within low-frequency period t, we obtain

$$V_t = \frac{1}{n} \sum_{i=1}^{n} \frac{r^2_{i,t}}{h_{i,t}} = \tau_t \frac{1}{n} \sum_{i=1}^{n} Z^2_{i,t} = \tau_t \overline{Z}_t, \tag{15}$$

where $\overline{Z}_t = n^{-1} \sum_{i=1}^{n} Z^2_{i,t}$ with $\mathbf{E}[\overline{Z}_t] = 1$ and $\mathbf{Var}[\overline{Z}_t] = (\kappa - 1)/n$. Again, it follows from Assumption 1 that the $\overline{Z}_t$ are $i.i.d.$ This suggests the specification

$$\tau_t = \lambda_0 + \lambda_1 V_{t-1} + \lambda_2 \tau_{t-1}. \tag{16}$$

Note that the low-frequency component τ_t still measures volatility in daily units. We will refer to this parametrization of the long-term component as MF2-GARCH-lf-n, where "lf" stands for *low-frequency* and n refers to the choice of the low-frequency period.[6] For example, when setting $n = 21$ or $n = 63$, the long-term component varies at the monthly or quarterly frequency. If Assumptions 1 and 3 hold, then $V_t = \tau_t \overline{Z}_t$ is a covariance stationary MEM(1,1) with $\mathbf{E}[V_t|\mathcal{F}_{t-1}] = \tau_t$ and $\mathbf{E}[V_t] = \lambda_0/(1 - \lambda_1 - \lambda_2)$. A drawback of the low-frequency updating of the long-term component is that it introduces a discontinuity into the daily conditional variances. Thus, the daily returns are no longer covariance stationary.

4 | Empirical Application

4.1 | Stock Market Data

We use daily return data for the S&P 500 starting in January 1971 and ending in June 2023. Using the notation for $n = 1$, daily log returns are computed as $r_t = 100(\log(P_t) - \log(P_{t-1}))$, where P_t is the close price on day t. We employ realized variances based on intraday data provided by Tick Data to evaluate the forecast performance. Daily realized variances, RV_t, are defined as the sum of the squared five-minute intraday log-returns on day t plus the squared overnight log-return (see Bollerslev et al. 2018). We compute realized variances for the period January 2000 to June 2023. Table 1 shows summary statistics for the daily returns and realized variances. The annualized daily returns have a sample mean of 7.38%. The mean of the annualized daily realized volatility is 15.81% during the period January 2010 to June 2023, which is the period that is used for the out-of-sample forecast evaluation. In addition, in Section 4.2.3 we use return data from the V-Lab for 2142 US and international equities.

4.2 | MF2-GARCH in Action

4.2.1 | Application to S&P 500

We first apply the MF2-GARCH model to the daily log returns of the S&P 500. We mainly focus on models with a daily long-term component and estimate the MF2-GARCH-rw-m and the MF2-GARCH-bw-m for the entire sample period from January 1971 to June 2023.

Choice of m: We first estimate both models for values of m up to 160 and determine the optimal m as the one that minimizes the BIC.[7] For both models, the upper left panel of Figure 3 shows the BIC as a function of m. The lowest value of the BIC materializes for $m = 63$ for both models. For this value of m, the MF2-GARCH-rw-m is clearly preferred relative to the MF2-GARCH-bw-m. To investigate whether this pattern holds more generally, we reestimate both models for three subsamples: January 1971 - December 2009 (upper right panel), January 1980 - December 2009 (lower left panel), and January 1980 to June 2023 (lower right panel). In all three panels, the optimal choice of m is around 63 and the MF2-GARCH-rw-m is the preferred model. Overall, the subsample analysis shows that the optimal choice of m is very stable for the S&P 500 and that equal weights are preferred to beta-weights.

Parameter estimates (sample period January 1971 to June 2023): The first two rows of Table 2 (labeled as "τ_t const.") show the parameter estimates of the nested one-component GJR-GARCH. The parameter estimates of α, γ, and β take typical values and indicate strong persistence in the GARCH component ($\alpha + \gamma/2 + \beta = 0.982$). Next, the table displays the parameter estimates of the MF2-GARCH-rw-m model. First, for the optimal window length, that is, $m = 63$, the estimates of the parameters in the long-term component, λ_1 and λ_2, are both significant. As expected, the persistence in the long-term component (0.982) is much stronger than the persistence in the short-term component (0.924). Also, due to introducing a time-varying long-term component, the short-term GJR-component is much less persistent than the one-component GJR-GARCH. The BIC clearly favors the two-component MF2-GARCH-rw-63 over the one-component model. To illustrate the consequences of choosing m too small and too large, we present parameter estimates for a monthly ($m = 21$) and semiannual ($m = 126$) averaging of past forecast errors. For $m = 21$, the persistence in the long-term component increases to 0.995. Presumably, this is because increasing λ_2 smoothes the long-term component and, thereby, counteracts the effect of decreasing m. For $m = 126$, the estimates of the long-term component's parameters are almost the same as for $m = 63$, but the standard errors increase considerably.

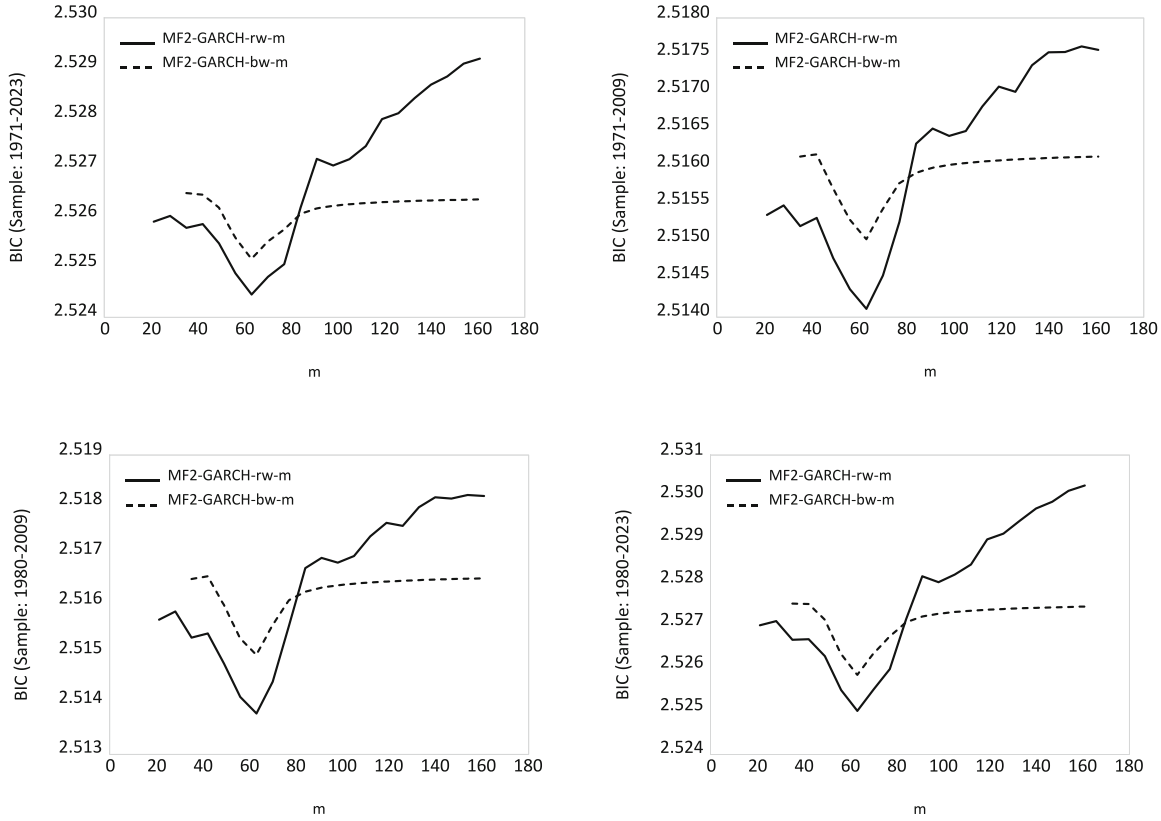

FIGURE 3 | MF2-GARCH-rw-*m* and MF2-GARCH-bw-*m* models are estimated using daily S&P 500 return data for the sample periods January 1971 to June 2023 (upper left), January 1971 to December 2009 (upper right), January 1980 to December 2009 (lower left), and January 1980 to June 2023 (lower right). The figure plots the BIC as a function of *m*.

TABLE 2 | Estimation results MF2-GARCH models: S&P 500.

τ_t	α	γ	β	λ_0	λ_1	λ_2	ω	LLF	BIC
				GJR-GARCH					
Constant	0.023*** (0.006)	0.115*** (0.021)	0.901*** (0.016)	1.015*** (0.136)				−16753.53	2.53433
				MF2-GARCH-rw-*m*					
$m = 63$	0.003 (0.007)	0.162*** (0.021)	0.840*** (0.018)	0.018* (0.009)	0.112** (0.054)	0.870*** (0.062)		−16678.61	**2.52445**
$m = 21$	0.002 (0.007)	0.174*** (0.021)	0.814*** (0.023)	0.005*** (0.002)	0.034*** (0.011)	0.961*** (0.012)		−16688.31	2.52591
$m = 126$	0.012* (0.007)	0.145*** (0.022)	0.864*** (0.018)	0.019 (0.016)	0.101 (0.094)	0.879*** (0.109)		−16702.77	2.52809
				MF2-GARCH-bw-*m*					
$m = 63$	0.003 (0.018)	0.162*** (0.055)	0.840*** (0.103)	0.018 (0.045)	0.112 (0.247)	0.870*** (0.294)	1.000 (3.505)	−16678.61	2.52516
$m = 126$	0.001 (0.008)	0.169*** (0.022)	0.824*** (0.027)	0.008** (0.003)	0.058*** (0.021)	0.933*** (0.023)	4.900*** (1.416)	−16686.25	2.52632
				MF2-GARCH-lf-*n*					
$n = 63$	0.016*** (0.006)	0.145*** (0.021)	0.864*** (0.016)	0.153*** (0.045)	0.745*** (0.130)	0.098 (0.137)		−16699.87	2.52766
$n = 21$	0.008 (0.007)	0.169*** (0.023)	0.825*** (0.026)	0.062*** (0.020)	0.390*** (0.101)	0.547*** (0.108)		−16702.62	2.52807
				GARCH-MIDAS-RV					
$m = 63$	0.005 (0.006)	0.177*** (0.021)	0.808*** (0.018)	0.313*** (0.044)	0.628*** (0.057)		73.671*** (7.047)	−16692.25	2.52651
$m = 21$	0.010 (0.006)	0.180*** (0.023)	0.813*** (0.023)	0.323*** (0.055)	0.608*** (0.072)		4.933*** (1.738)	−16694.05	2.52678

Note: The table reports estimation results for GJR-GARCH, MF2-GARCH-rw-*m*, MF2-GARCH-bw-*m*, MF2-GARCH-lf-*n*, and GARCH-MIDAS-RV (with $K = 252$) models. For the GARCH-MIDAS-RV, λ_0 corresponds to c (see Section A.3 in the Supporting Information). All models are estimated using daily return data for the period January 1971 to June 2023. The numbers in parentheses are Bollerslev–Wooldridge robust standard errors. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. LLF is the value of the maximized log-likelihood function. BIC is the Bayesian information criterion.

For the MF2-GARCH-bw-m specification, we focus on the optimal choice of $m = 63$. For this value of m, the estimates of the short- and long-term component's parameters are close to those of the MF2-GARCH-rw-63 model. The estimate of ω is equal to one. That is, the optimal beta-weighting scheme for $m = 63$ is flat. As suggested by the BIC comparison, this confirms that the MF2-GARCH-rw-63 is the preferred model. Thus, for the S&P 500, allowing for a flexible weighting scheme, only increases the standard errors of the remaining parameters. For the MF2-GARCH-lf-n models with a low-frequency long-term component, we display parameter estimates for a quarterly ($n = 63$) and a monthly ($n = 21$) long-term component. When the long-term component varies at the quarterly frequency, the estimate of λ_2 is insignificant. That is, when n is large, a MEM(0,1) is sufficient to capture the movements in the conditional expectation of V_t. In contrast, when τ_t evolves at a monthly ($n = 21$) frequency, λ_2 becomes significant and a MEM(1,1) is preferred. In terms of the BIC, the MF2-GARCH-rw-63 model with a daily long-term component is clearly preferred to models with a low-frequency long-term component. Finally, we estimate the GARCH-MIDAS-RV model as in equation (31) of Section A.3 in the Supporting Information with quarterly ($m = 63$) or monthly ($m = 21$) rolling window realized variance and a beta-weighting scheme with $K = 252$. In this model, λ_1 measures the effect of RV_t^{rw} on long-term volatility. As expected, this effect is estimated to be positive. Again, the BIC favors the MF2-GARCH-rw-63 over the GARCH-MIDAS-RV models.

Figure 4 shows the annualized conditional volatility, $\sqrt{h_t \tau_t}$, and the long-term volatility, $\sqrt{\tau_t}$, as implied by the MF2-GARCH-rw-63. As expected, the behavior of $\sqrt{\tau_t}$ mirrors that of the averaged standardized forecast errors from the one-component GJR-GARCH in Figure 1. That is, the figure confirms that there are prolonged cycles in long-term volatility.[8] Specifically, long-term volatility is high during the Great Recession and the Covid-19 pandemic but also after the crash on October 19, 1987. Using Baker et al. 2021 equity market volatility (EMV) trackers, we show in Appendix A.5 of the Supporting Information that the MF2-GARCH's long-term volatility is driven by various types of economic news (such as macroeconomic and monetary policy news), which a GARCH-MIDAS-X model with a single macroeconomic or financial explanatory variable is unlikely to capture fully.

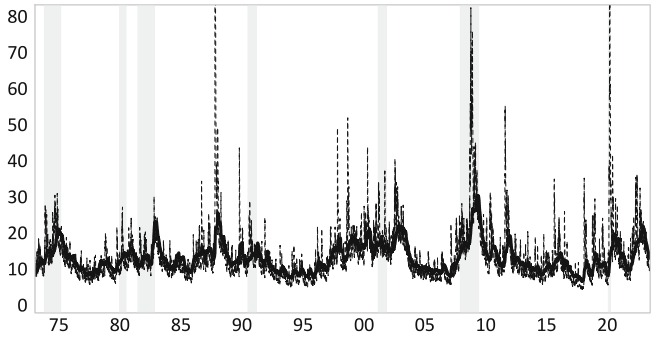

FIGURE 4 | The figure shows the estimated conditional volatility, $\sqrt{h_t \tau_t}$, (dashed line) and long-term volatility, $\sqrt{\tau_t}$, (solid line) from the MF2-GARCH-rw-63 model for the daily S&P 500 returns presented in Table 2. All quantities are annualized. Gray shaded areas represent NBER recession periods.

4.2.2 | Implications for Volatility Forecasting

In this section, we illustrate the behavior of the volatility forecasts of the MF2-GARCH-rw-m. In particular, we hint at the differences between the MF2-GARCH's forecasts and those of the nested GJR-GARCH. We rely on the parameter estimates for the MF2-GARCH-rw-63 and the GJR-GARCH for the S&P 500 in Table 2. The left panels of Figure 5 show the conditional volatility (solid black line) and long-component (solid gray line) of the MF2-GARCH for two subsamples. The horizontal line represents the unconditional volatility. We compute volatility forecasts from day $t = 50$ onwards. In the upper/lower panel, $t = 50$ corresponds to August 10, 2011, and September 14, 2016, respectively. The dashed black lines correspond to the forecasts from the MF2-GARCH. In the right panels, we plot the volatility forecasts from the GJR-GARCH as dashed black lines, whereby we impose the same unconditional variance as for the MF2-GARCH-rw-63 and the same conditional variance at the forecast origin.[9] We have chosen the forecast origin so that we can illustrate the differences between the forecasts from the two models. In the upper panels, the conditional volatility as well as the long-term component are below the unconditional volatility until there is a jump in volatility up to a level above 50%.[10] As expected, the right panel shows that the forecasts from the GJR-GARCH overestimate volatility after the surge in volatility. This can be explained by the high persistence in the one-component GARCH model. In contrast, the forecasts from the MF2-GARCH (left panel) are much more in line with the actual behavior of the conditional volatility. This is due to the lower persistence in the short-term component, which dominates the behavior of the MF2-GARCH forecast in the short run. In the lower panels, volatility stayed below the unconditional volatility for an extended period. Thus, the long-term volatility component is below the unconditional volatility at the forecast origin. However, shortly before the forecast origin, the conditional volatility spikes and is still slightly above the level of the unconditional volatility at the forecast origin.[11] The forecasts from the GJR-GARCH directly converge to the unconditional volatility. The MF2-GARCH forecast behaves very differently. In the medium run, the forecast converges towards the forecast of the long-term component (dashed gray line). That is, the forecast decreases *below* the unconditional volatility. Only in the very long run, the MF2-GARCH forecast will converge towards the unconditional volatility. This illustrates that the MF2-GARCH forecast captures the empirical observation that there are persistent cyclical movements of the conditional volatility around the unconditional volatility. In summary, both examples illustrate that forecasts from the MF2-GARCH are more realistic than forecasts from the one-component GJR-GARCH. The MF2-GARCH forecasts reflect both the current level of volatility, $\sqrt{h_{t+1} \tau_{t+1}}$, as well as the prevailing volatility trend, $\sqrt{\tau_{t+1}}$.

4.2.3 | Evidence From the V-Lab

We complement the previous findings with evidence from the V-Lab. Overall, we consider 2142 US and international stocks. The estimation period for all stocks ends on April 5, 2024. The beginning of the estimation period is stock-specific and

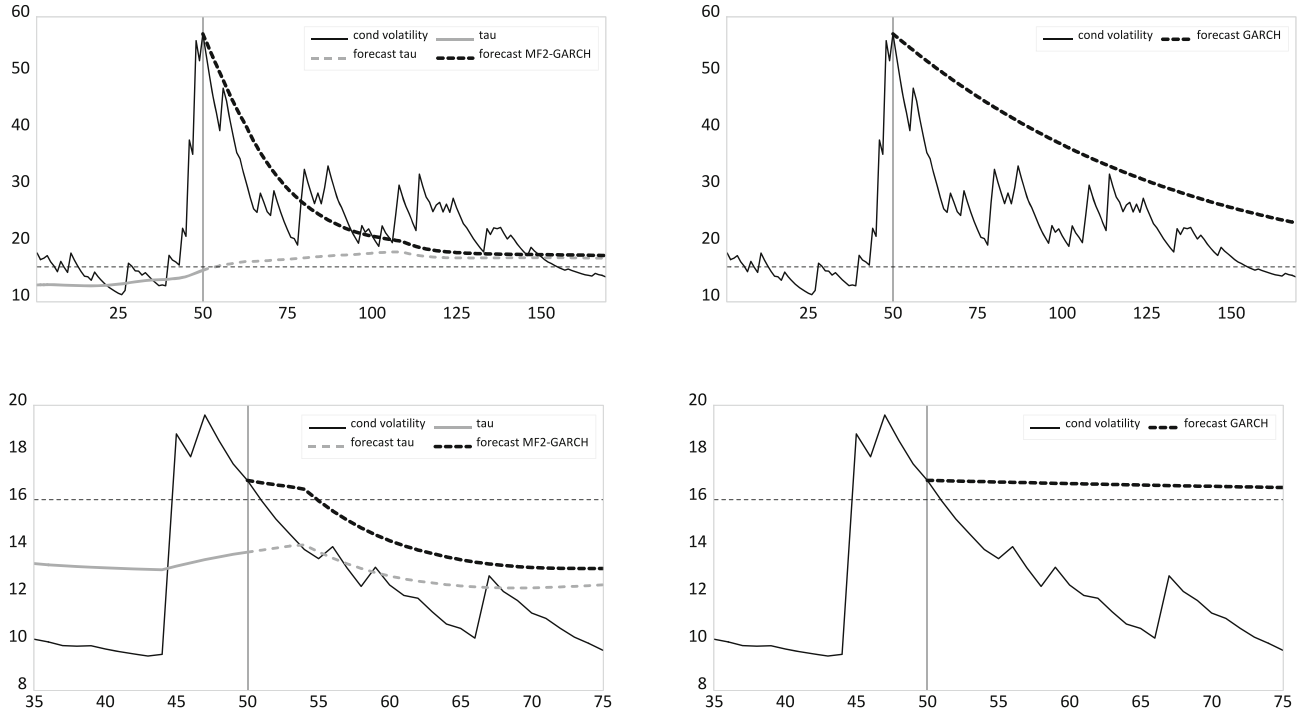

FIGURE 5 | The left panels show the conditional volatility (solid black line) from an MF2-GARCH-rw-63 estimated for S&P 500 returns (see Table 2). From day $t = 50$ (indicated by the black vertical line) onwards, we compute volatility forecasts (dashed black line). In the upper/lower panels $t = 50$ corresponds to August 10, 2011 and September 14, 2016, respectively. The left panels also show the long-term components (gray line) and the forecast of long-term volatility (dashed gray line). The dashed lines in the right panels are the volatility forecast from a GJR-GARCH(1,1). We impose the same conditional volatility at the forecast origin and the same unconditional volatility. All quantities are annualized.

TABLE 3 | Parameter estimates for 2142 stocks in V-Lab: Summary statistics.

	α	γ	β	ϕ	$\lambda_1 + \lambda_2$	m	BIC(MF2)< BIC(•)
GJR-GARCH	0.079	0.032	0.883	0.978	—	—	0.959
MF2-GARCH-rw	0.111	0.043	0.726	0.861	0.993	51.000	—
Spline-GARCH	0.118	—	0.800	0.920	—	—	0.632
ZS-Spline-GARCH	0.117	—	0.804	0.923	—	—	0.658

Note: The table presents median parameter estimates for 2142 US and international stocks. The GJR-GARCH, MF2-GARCH-rw-*m*, Spline-GARCH and ZS-Spline-GARCH models were estimated in V-Lab. The last column, BIC(MF2)< BIC(•), reports the percentage of equities for which the BIC of the MF2-GARCH is smaller than the BIC of the model in the respective row. The estimation period ended on April 5, 2024.

depends on data availability. For each stock, we select the optimal MF2-GARCH-rw-m specification by choosing the value of m that minimizes the BIC. Table 3 presents summary statistics for comparing the MF2-GARCH parameter estimates with those for the nested GJR-GARCH.[12] In addition, we report results for the Spline-GARCH and the Zero Slope Spline-GARCH (ZS-Spline-GARCH). For those models, we only report the estimates of the parameters in the short-term component. For all models, Table 3 shows the median parameter estimates of α, γ and β. In addition, the table displays the median persistence in the short- (ϕ) and long-term ($\lambda_1 + \lambda_2$) component and the median value of m. The table shows that the estimate of β decreases substantially when estimating an MF2-GARCH instead of a GJR-GARCH. Conversely, the median estimates of α and γ slightly increase. Overall, the persistence in the short-term component is strongly reduced when introducing the long-term component. Naturally, the persistence in the smooth long-term component of the MF2-GARCH is considerably higher than in the short-term component. Interestingly, for the Spline-GARCH and the ZS-Spline-GARCH, the reduction of the persistence in the short-term component is less pronounced. Presumably, this is because the long-term components of these models are too smooth. The smooth long-term components capture low-frequency variations (i.e., at the business cycle) in long-term volatility but not increases in volatility due to isolated events such as the crash on October 19, 1987. Across the 2142 stocks, the median number of knots that are selected for both the Spline-GARCH and the ZS-Spline-GARCH is eight (this number is not reported in the table). Thus, the Spline-GARCH models require the estimation of considerably more parameters than the MF2-GARCH. The last column of Table 3 reports the percentage of stocks for which the BIC of the MF2-GARCH-rw is smaller than the BIC of the models in the respective rows. The MF2-GARCH is preferred over the GJR-GARCH for 96% of the stocks. In addition, it is preferred over the two Spline-GARCH

models for over 60% of the stocks. In summary, our findings from the V-Lab confirm the results for the S&P 500 from Table 2 for a large number of stocks.

4.3 | Out-of-Sample Forecasting

Finally, we evaluate the MF2-GARCH's out-of-sample forecast performance. Most previous literature focuses on short forecast horizons, such as one-day ahead. However, as discussed in, for example, Christoffersen and Diebold 2000, Engle 2009b, Ederington and Guan 2010, positions are often held for more extended periods and, hence, long-term risk predictions should also be evaluated. Indeed, the failure of risk management during the financial crisis can be partly attributed to focusing on short-term risks, while neglecting long-term risks (Engle 2009b). Thus, our particular focus is on longer-term volatility predictions. The MF2-GARCH is designed to capture slow movements in the long-run volatility, and we intend to test whether the model helps to improve longer-term volatility forecasts. In addition, our results will provide new evidence for answering the question of how far ahead we can forecast volatility (see also Ghysels et al. 2019).

4.3.1 | S&P 500

We focus on the MF2-GARCH specification with a daily long-term component and rolling window weighting scheme. As competitor models, we consider the nested GJR-GARCH, the GARCH-MIDAS-RV and the log-HAR model with leverage. While the MF2-GARCH, the GJR-GARCH, and the GARCH-MIDAS-RV model are based on daily returns, the log-HAR with leverage makes use of realized variances based on intraday data. Within the respective model classes, the GJR-GARCH (see, for instance, Hansen and Lunde 2005, and Brownlees, Engle, and Bryan 2012) and the log-HAR (see, e.g., Bollerslev et al. 2018, and Conrad and Kleen 2020) are well known for their excellent short-term forecast performance.

The out-of-sample (OOS) period is January 2010 to June 2023. On each day of the OOS period, we compute forecasts of the variance on the next day, over the next week, as well as the s-months ahead *forward variance* with $s \in \{1, \dots, 8\}$. For computing forward variances, we assume that each month has 21 trading days. For example, the MF2-GARCH forecast of 2-month forward variance is $\hat{\sigma}^2_{t+22,t+42|t} = \sum_{j=1}^{21} \hat{\sigma}^2_{t+21+j|t}$.

We reestimate all models on a rolling window of observations and update parameters on a monthly basis. For the MF2-GARCH, we select the optimal m based on the BIC and estimate κ based on the respective in-sample observations. The GARCH-MIDAS-RV models are based on a rolling window RV_t^{rw} with $m \in \{21, 63\}$ and $K = 252$. When forecasting the s-months ahead *forward variance*, we specify the log-HAR model with leverage as

$$\begin{aligned}&\log(RV_{t+(s-1)\cdot 21+1,t+s\cdot 21})\\&= b_0 + b_1 \log RV_t + b_2 \log\left(\frac{RV_{t-4,t}}{5}\right) + b_3 \log\left(\frac{RV_{t-20,t}}{21}\right)\\&+ b_4 r_t \mathbf{1}_{\{r_t<0\}} + b_5 \frac{r_{t-4,t}}{5}\mathbf{1}_{\{r_{t-4,t}<0\}} + b_6 \frac{r_{t-20,t}}{21}\mathbf{1}_{\{r_{t-20,t}<0\}} + \zeta_t^{(s)},\end{aligned} \quad (17)$$

where $RV_{t+1,t+k} = \sum_{j=1}^{k} RV_{t+j}$, $r_{t+1,t+k} = \sum_{j=1}^{k} r_{t+j}$ and $\mathbf{1}_{\{r_{t+1,t+k}<0\}}$ is an indicator function which takes the value one if $r_{t+1,t+k} < 0$, and zero otherwise. For daily and weekly forecast horizons, the dependent variable is $\log(RV_{t+1})$ and $\log(RV_{t+1,t+5})$, respectively. The log-HAR is estimated by OLS and forecasts are based on the assumption that the $\zeta_t^{(s)}$ are normally distributed.[13]

We evaluate the forecasts based on the SE and the QLIKE loss, which are both robust loss functions (see Patton 2011). For the forecast evaluation, we proxy the true latent volatility by the corresponding realized variances. The realization of the s-months forward variance is $RV_{t+(s-1)\cdot 21+1,t+s\cdot 21} = \sum_{j=1}^{21} RV_{t+(s-1)\cdot 21+j}$. To increase the readability of the forecast losses, we compare them to the losses of a simple benchmark. As benchmark, we consider the "historical volatility forecast," which we define as the appropriately scaled mean of the daily realized variances during the 10 years prior to the forecast origin. For each model and forecast horizon, we report the model performance as the ratio of the root mean square error (RMSE) of the respective model and the RMSE of historical volatility. Similarly, we present the respective QLIKE ratio. In addition, we compute a model confidence set (MCS) as proposed in Hansen, Lunde, and Nason 2011. In Appendix A.6 of the Supporting Information, we describe in detail how the MCS is computed.

We first evaluate the forecasts during the January 2010 to December 2019 period, which ends before the Covid-19 pandemic. Panels A and B of Table 4 present results for the RMSE and the QLIKE, respectively. For each forecast horizon, the bold numbers indicate the model that achieves the lowest RMSE or QLIKE ratio. Blue shaded cells indicate which models are included in the MCS. In Panel A, the MF2-GARCH achieves the lowest RMSE for all forecast horizons. The entry of 0.685 at the one month horizon implies that the RMSE of the MF2-GARCH is 31.5% lower than the RMSE of the historical volatility forecast. Even at the longest forecast horizon of 8 months, the RMSE is almost 20% lower. The MF2-GARCH is included in the MCS as the single model for all forecast horizons longer than one week. In Panel B, the log-HAR is the preferred model for forecast horizons up to 2 months.[14] Beyond that horizon, the MF2-GARCH produces the best forecasts. Overall, the MF2-GARCH is successful in forecasting volatility far into the future. We believe that the MF2-GARCH's outperformance at long forecast horizons is due to the specification of the long-term component, which is designed to capture the underlying volatility cycle. In the medium term, the MF2-GARCH's forecasts are strongly influenced by the dynamic structure of the long-term component but eventually converge to the unconditional variance. On the other hand, the GARCH-MIDAS-RV's long-term component is a proxy of the level of volatility in the recent past. Imposing that the forecast converges to this level does not pay off.

The full OOS period ends in June 2023 and covers the Covid-19 pandemic. The beginning of the Covid-19 pandemic in March 2020 is characterized by days with extremely high realized variances. For example, on March 12 and 16, 2020, realized variances took values corresponding to an annualized daily volatility of 122% and 160%. This creates a short period of instability. In particular, the monthly realized forward variances are strongly shifted upwards once an extreme realized variance enters the

TABLE 4 | Out-of-sample forecasting: S&P 500 – pre–Covid-19.

Forecast horizon	Day	Week	1 m	2 m	3 m	4 m	5 m	6 m	7 m	8 m
					Panel A: Relative RMSE					
Historical forecast	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GJR-GARCH	0.788	0.763	0.798	0.864	0.901	0.910	0.899	0.883	0.874	0.866
MF2-GARCH-rw	**0.744**	**0.699**	**0.685**	**0.790**	**0.834**	**0.846**	**0.832**	**0.820**	**0.815**	**0.811**
log-HAR lev.	0.766	0.739	0.725	0.829	0.900	0.926	0.913	0.910	0.899	0.888
GARCH-MIDAS $m = 21$	0.770	0.738	0.715	0.861	0.924	0.958	0.958	0.950	0.938	0.927
GARCH-MIDAS $m = 63$	0.791	0.770	0.746	0.886	0.954	0.992	0.997	0.988	0.978	0.974
					Panel B: Relative QLIKE					
Historical forecast	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GJR-GARCH	0.647	0.640	0.723	0.847	0.901	0.922	0.920	0.918	0.919	0.918
MF2-GARCH-rw	0.615	0.613	0.708	0.827	**0.884**	**0.894**	**0.875**	**0.866**	**0.861**	**0.859**
log-HAR lev.	**0.544**	**0.552**	**0.677**	**0.822**	0.889	0.924	0.903	0.903	0.902	0.893
GARCH-MIDAS $m = 21$	0.636	0.641	0.734	0.857	0.909	0.918	0.907	0.907	0.899	0.885
GARCH-MIDAS $m = 63$	0.638	0.645	0.745	0.876	0.918	0.926	0.910	0.907	0.910	0.908

Note: The out-of-sample period is January 2010 to December 2019. Using a rolling window scheme, all models are re-estimated on a monthly basis (i.e., every 21 days). For both GARCH-MIDAS-RV models, we set $K = 252$. The first in-sample period ends in December 2009. The forecast horizons are one day, one week, and from 1 month (1m) to 8 months (8m). Numbers reported are the RMSE (Panel A) or QLIKE (Panel B) losses for each model and forecast horizon relative to the respective RMSE (Panel A) or QLIKE (Panel B) of historical volatility. Bold numbers indicate the model with the lowest RMSE or QLIKE. Blue shaded cells indicate that the respective model is included in the 85% model confidence set.

measure and stay at elevated levels until the extreme realized variance drops out of the 21-days window. Although this period of extreme volatility is unpredictable for all models, it has a strong influence on the Diebold and Mariano 1995 tests underlying the MCS. As shown in Iacone, Rossini, and Viselli 2024, Diebold and Mariano 1995 tests become uninformative due to low power when short-lived periods of instability are included in the OOS period.[15] Following van Dijk and Franses 2003, we avoid this problem by considering a weighted average loss differential, where we attach zero weight to unpredictable volatility outliers. We define those outliers as a daily realized variance that is above the 99th percentile of the OOS daily realized variances' empirical distribution. Because almost all of these outliers materialize at the beginning of the Covid-19 pandemic, our approach is essentially the same as omitting this short period of instability from the evaluation sample.

Table 5 shows the results for the full OOS period. The results are similar to those in Table 4. Independently of the loss function, the MF2-GARCH's forecasts dominate for forecast horizons of 3 or more months. At shorter horizons, the log-HAR performs well, in particular according to the QLIKE loss. At the shortest forecast horizon of one day, the log-HAR clearly benefits from the information in the realized volatility. Overall, our previous results are confirmed. Figure A.5 in the Supporting Information visualizes the forecast performance of the MF2-GARCH, the log-HAR, and the benchmark model for the s-months forward variances. The figure shows the cumulated losses (left panels: SE, right panels: QLIKE) for forecast horizons of one month (upper panels), 4 months (middle panels) and 8 months (lower panels). In line with the results from Table 5, the MF2-GARCH outperforms the log-HAR in all panels but the upper right one. Most importantly, the ranking of the two models is relatively stable across the entire OOS period.

To shed more light on the potentially time-varying relative forecast performance of the different models, we present a forecast evaluation that is conditional on being in a high volatility regime on the day a forecast is made. We classify a day as belonging to the high volatility regime if the realized variance of that day is above the 70th percentile of the OOS empirical distribution of realized variances. As Table A.4 in the Supporting Information shows, the evidence in favor of the MF2-GARCH becomes stronger when focusing on forecasts that were made during high volatility regimes.

Finally, we compare the relative forecast performance of the MF2-GARCH and the log-HAR by applying the conditional predictive ability test of Giacomini and White 2006. We use the realized variance, RV_t, that materializes on the day the forecasts are made as a predictor for the loss difference. That is, for each forecast horizon, we run a regression of the loss difference on a constant and the realized variance. The results (not reported) show that a high realized variance today predicts that the MF2-GARCH will outperform the log-HAR at essentially all forecast horizons. To illustrate this effect, Figure A.6 in the Supporting Information shows the estimated conditional mean of the regression and the corresponding 95% confidence interval as a function of RV_t.

4.3.2 | V-Lab

We also evaluated the forecast performance of the MF2-GARCH in the V-Lab. Here, we focus on the GJR-GARCH, the Spline-GARCH and the Zero-Slope Spline-GARCH as competitor models.[16] The forecast performance of the different models is evaluated for 20 (randomly selected) US stocks: Boeing (BA), Berkshire Hills Bancorp (BHLB), BlackRock Inc (BLK), Cardinal

TABLE 5 | Out-of-sample forecasting: S&P 500 – full OOS period.

Forecast horizon	Day	Week	1 m	2 m	3 m	4 m	5 m	6 m	7 m	8 m
Panel A: Relative RMSE										
Historical forecast	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GJR-GARCH	0.990	0.837	0.869	1.618	0.863	0.987	0.912	0.834	0.814	0.798
MF2-GARCH-rw	0.781	**0.641**	**0.602**	0.766	**0.735**	**0.757**	**0.772**	**0.780**	**0.781**	**0.786**
log-HAR lev.	**0.776**	0.662	0.628	**0.765**	0.818	0.841	0.859	0.856	0.848	0.843
GARCH-MIDAS $m = 21$	1.197	0.891	0.938	1.101	1.215	1.293	1.296	1.306	1.333	1.368
GARCH-MIDAS $m = 63$	1.167	0.916	0.903	1.339	1.329	1.411	1.440	1.437	1.492	1.495
Panel B: Relative QLIKE										
Historical forecast	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GJR-GARCH	0.676	0.652	0.702	0.829	0.873	0.892	0.898	0.894	0.891	0.889
MF2-GARCH-rw	0.646	0.613	0.664	0.774	**0.832**	**0.853**	**0.857**	**0.857**	**0.856**	**0.861**
log-HAR lev.	**0.572**	**0.549**	**0.625**	**0.770**	0.840	0.874	0.885	0.888	0.890	0.886
GARCH-MIDAS $m = 21$	0.669	0.650	0.707	0.821	0.878	0.898	0.909	0.911	0.920	0.927
GARCH-MIDAS $m = 63$	0.669	0.653	0.713	0.833	0.879	0.899	0.905	0.906	0.918	0.932

Note: The out-of-sample period is January 2010 to June 2023. See notes of Table 4.

TABLE 6 | Out-of-sample forecasting in V-Lab: 20 US stocks.

	Day	Week	1 m	2 m	3 m	4 m	5 m	6 m	7 m	8 m
Panel A: Ranking according to RMSE										
GJR-GARCH	2.45	2.45	2.70	3.05	2.80	3.10	2.90	2.80	2.75	2.70
MF2-GARCH-rw	**1.90**	**1.80**	**1.85**	**1.60**	**1.70**	**1.60**	**1.60**	**1.55**	**1.50**	**1.50**
Spline-GARCH	3.35	3.55	3.25	3.35	3.50	3.35	3.45	3.50	3.40	3.50
ZS-Spline-GARCH	2.30	2.20	2.20	2.00	2.00	1.95	2.05	2.15	2.35	2.30
Panel B: Ranking according to QLIKE										
GJR-GARCH	2.20	2.10	1.75	**1.90**	2.20	2.45	2.50	2.55	2.65	2.70
MF2-GARCH-rw	**1.40**	**1.35**	**1.70**	**1.90**	**1.70**	**1.65**	**1.70**	**1.70**	**1.75**	**1.80**
Spline-GARCH	3.95	4.00	3.85	3.60	3.60	3.40	3.30	3.25	3.15	3.20
ZS-Spline-GARCH	2.45	2.55	2.70	2.60	2.50	2.50	2.50	2.50	2.45	2.30

Note: The out-of-sample period is January 2010 to June 2023. All models are re-estimated on a daily basis. The first in-sample period ends in December 2009. For each forecast horizon, the numbers represent the average ranking of the respective models across the 20 US stocks according to the RMSE (Panel A) and the QLIKE (Panel B). Bold numbers indicate the model with the lowest ranking.

Health Inc (CAH), Edison International (EIX), Farmer Brothers Co (FARM), Guess (GES), HEICO Corp (HEI), Harley-Davidson Inc (HOG), Houston American Energy (HUSA), Marriott International Inc (MAR), Moody's Corp (MCO), MetLife Inc (MET), Microsoft Corp (MSFT), Novavax (NVAX), Oil-Dri Corp of America (ODC), Pfizer (PFE), Ralph Lauren (RL), Union Pacific (UNP), and Whirlpool Corp (WHR). Because there is no intraday data available in V-Lab, squared daily returns are used to compute realized variances. The out-of-sample period is January 2010 to June 2023. For each forecast horizon and for each stock, based on the RMSE (respectively the QLIKE) the four models are ranked from rank 1 (best) to rank 4 (worst). For each forecast horizon, Table 6 presents the average rank across the 20 stocks. According to both loss functions, the MF2-GARCH performs best for all forecast horizons. Thus, Table 6 provides further evidence for the superior forecast performance of the MF2-GARCH within the class of GARCH models.

5 | Conclusions

We suggest a multiplicative factor multifrequency component GARCH model. The new model is motivated by the observation that a rolling window average of past standardized forecast errors of one-component GARCH models has predictive power for current and future standardized forecast errors. This predictability is due to counter-cyclical movements in financial volatility, which simple GARCH models do not adequately capture. In contrast, the MF2-GARCH explicitly models these volatility cycles. We show that the MF2-GARCH is straightforward to estimate, features stationary returns and a NIC that is more responsive to news during low volatility periods than during high volatility periods. In addition, multistep ahead volatility forecasts can be easily computed. Overall, the properties of the MF2-GARCH clearly distinguish it from other component models. Finally, our empirical results show that the new specification outperforms the nested

GJR-GARCH, the Spline-GARCH, the GARCH-MIDAS-RV and the log-HAR in out-of-sample forecast performance.

In general, the MF2-GARCH model will benefit applications that require long-term forecasts of financial volatility, such as long-run value-at-risk predictions or measurement of systemic risk. For example, Conrad, Schoelkopf, and Tushteva 2024 show that the MF2-GARCH allows to compute "volatility news" separately for the short-term and the long-term components. Assuming a positive relation between expected returns and the conditional variance of returns in a GARCH-in-Mean type model, unexpected returns can be decomposed into cash flow and discount rate news. In this framework, discount rate news is mainly driven by news to the MF2-GARCH's long-term component. This is because only news to the long-term component is persistent enough to generate sizable variation in discount rates. From this, it follows that the MF2-GARCH's long-term volatility component is a strong predictor for the strength of the instantaneous response of the stock market to surprises in macroeconomic announcements (see Conrad, Schoelkopf, and Tushteva 2024).

It will also be interesting to employ the long-term component in applications that require low-frequency estimates of volatility, for example, when analyzing the link between financial volatility and financial crises (see, for instance, Danielsson, Valenzuela, and Zer 2018).

Acknowledgments

We are grateful to the co-editor, Eric Ghysels, and two anonymous referees for their comments, which greatly improved our paper. We would like to thank Jörg Breitung, Christian Brownlees, Rob Capellini, Zeno Enders, Jean-David Fermanian, Christian Francq, Christian Gourieroux, Onno Kleen, Robinson Kruse-Becher, Enno Mammen, Anne Opschoor, Lara Schadwinkel, Julius Schölkopf, Timo Teräsvirta, and Jean Michel Zakoïan as well as seminar and conference participants at CREST (January 2020), the ES World Congress (2020), the 13th Annual SoFiE Conference (2021), and the 11th ECB Conference on Forecasting Techniques (2021) for their feedback on earlier versions of the paper. Funding by the German Federal Ministry of Education and Research (BMBF) and the Baden-Württemberg Ministry of Science as part of Germany's Excellence Strategy (ExU 10.2.31) is gratefully acknowledged. Open Access funding enabled and organized by Projekt DEAL.

Data Availability Statement

The authors have nothing to report.

Open Research Badges

This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Data is available at https://doi.org/10.15456/jae.2025013.1232487362.

Endnotes

[1] In V-Lab, the MF2-GARCH is estimated for more than 18,000 assets from different asset classes on a weekly basis. See: https://vlab.stern.nyu.edu/docs/volatility/MF2-GARCH.

[2] The data will be introduced and discussed in more detail in Section 4.1. See also Panel A of Table 1.

[3] For details, see the discussion below Assumption 6 in Conrad and Schienle 2020, as well as Remark 7 in the Supplementary Appendix of their paper.

[4] We obtained similar results for the DAX and the Hang Seng Index (HSI). In addition, we found evidence for considerable co-movement in standardized volatility forecast errors internationally (see also Engle and Campos-Martins 2023).

[5] For a graphical illustration, Figure A.3 in the Supporting Information plots the annualized unconditional volatility as a function of m and for $\kappa \in \{3, 5, 7\}$. The model parameters are chosen as $\alpha = 0.02$, $\gamma = 0.10$, $\beta = 0.8$, $\lambda_0 = 0.01$, $\lambda_1 = 0.05$, and $\lambda_2 = 0.94$.

[6] We treat n as a fixed number that is not "too large." If $n \to \infty$, then $\overline{Z}_t$ converges to one in probability and, hence, an identification issue arises due to the linear dependence of V_t and τ_t.

[7] As the choice of m does not affect the number of parameters, we could also determine the optimal value based on the likelihood function. We prefer the BIC because this allows for a meaningful comparison with the nested one-component GJR-GARCH.

[8] We checked whether there is still predictability in the volatility forecast errors of the MF2-GARCH-rw-63. We found no evidence for autocorrelation in the daily squared standardized residuals, $\hat{Z}_t^2$, when averaged at lower frequencies. In addition, the empirical density of $\hat{Z}_t$ is close to being symmetric.

[9] That is, the s-step ahead forecast is given by: $\mathbf{Var}[r_t] + (\alpha^{GA} + \gamma^{GA}/2 + \beta^{GA})^{s-1}(h_{t+1}\tau_{t+1} - \mathbf{Var}[r_t])$, where α^{GA}, γ^{GA} and β^{GA} are the parameters of the GJR-GARCH and $\mathbf{Var}[r_t]$ is the unconditional variance of the MF2-GARCH.

[10] The increase in volatility was driven by the European sovereign debt crisis and a downgrade of the U.S.'s credit rating by Standard & Poor's.

[11] The spike in volatility was associated with fears that the Federal Reserve might raise interest rates.

[12] We only include stocks for which the MF2-GARCH parameter estimates satisfy Assumptions 2 and 3.

[13] We also considered log-HAR models with (the log of) quarterly and semiannual averages of realized variances as additional explanatory variables. However, those specifications did not lead to an improved forecast performance relative to the baseline model. In addition, we considered the log-HAR without leverage and the pure HAR model. Again, both specifications did not lead to improvements in forecast performance.

[14] As discussed in Patton 2020, the ranking of models implied by the MSE and QLIKE can differ due to model misspecification or parameter estimation error. See also Section A.2 of the Supporting Information.

[15] In line with their result, we find that the MCS includes essentially all models when we do not control for this period of instability. In this setting, the assumption that the loss differences are stationary, which underlies the MCS procedure (see Hansen, Lunde, and Nason 2011, Assumption 2), is likely to be violated.

[16] V-Lab does not produce forecasts for the HAR model or the GARCH-MIDAS.

Bibliography

Alexander, C., E. Lazar, and S. Stanescu. 2021. "Analytic Moments for GJR-GARCH (1,1) Processes." *International Journal of Forecasting* 37: 105–124.

Amado, C., and T. Teräsvirta. 2013. "Modelling Volatility by Variance Decomposition." *Journal of Econometrics* 175: 142–153.

Amado, C., and T. Teräsvirta. 2017. "Specification and Testing of Multiplicative Time-Varying GARCH Models With Applications." *Econometric Reviews* 36: 421–446.

Asgharian, H., A. J. Hou, and F. Javed. 2013. "The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH-MIDAS Approach." *Journal of Forecasting* 32: 600–612.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen. 1996. "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 74: 3–30.

Baker, S. R., N. Bloom, S. J. Davis, and K. Kost. 2021. Policy News and Stock Market Volatility. Available at SSRN: https://doi.org/10.2139/ssrn.3363862.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen. 2018. "Risk Everywhere: Modeling and Managing Volatility." *Review of Financial Studies* 31: 2730–2773.

Brownlees, C., R. F. Engle, and K. Bryan. 2012. "A Practical Guide to Volatility Forecasting Through Calm and Storm." *Journal of Risk* 14: 3–22.

Christoffersen, P. F., and F. X. Diebold. 2000. "How Relevant Is Volatility Forecasting for Financial Risk Management?" *The Review of Economics and Statistics* 82: 12–22.

Christoffersen, P. F., K. Jacobs, C. Ornthanalai, and Y. Wang. 2008. "Option Valuation With Long-Run and Short-Run Volatility Components." *Journal of Financial Economics* 90: 272–297.

Conrad, C., and O. Kleen. 2020. "Two Are Better Than One: Volatility Forecasting Using Multiplicative Component GARCH-MIDAS Models." *Journal of Applied Econometrics* 35: 19–45.

Conrad, C., and K. Loch. 2015. "Anticipating Long-Term Stock Market Volatility." *Journal of Applied Econometrics* 30: 1090–1114.

Conrad, C., and M. Schienle. 2020. "Testing for an Omitted Multiplicative Long-Term Component in GARCH Models." *Journal of Business & Economic Statistics* 38: 229–242.

Conrad, C., J. T. Schoelkopf, and N. Tushteva. 2024. Long-Term Volatility Shapes the Stock Market's Sensitivity to News. Available at SSRN: https://doi.org/10.2139/ssrn.4632733

Corsi, F., and R. Reno. 2012. "Discrete-Time Volatility Forecasting With Persistent Leverage Effect and the Link With Continuous-Time Volatility Modeling." *Journal of Business & Economic Statistics* 30: 368–380.

Danielsson, J., M. Valenzuela, and I. Zer. 2018. "Learning From History: Volatility and Financial Crises." *Review of Financial Studies* 31: 2774–2805.

Diebold, F. X., and R. S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13: 253–263.

Ding, Z., and C. Granger. 1996. "Modeling Volatility Persistence of Speculative Returns: A New Approach." *Journal of Econometrics* 73: 185–215.

Dorion, C. 2016. "Option Valuation With Macro-Finance Variables." *Journal of Financial and Quantitative Analysis* 51: 13591389.

Ederington, L. H., and W. Guan. 2010. "Longer-Term Time-Series Volatility Forecasts." *Journal of Financial and Quantitative Analysis* 45: 1055–1076.

Engle, R., and G. Lee. 1999. "A Permanent and Transitory Component Model of Stock Return Volatility." In *Cointegration Causality and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, edited by R. F. Engle and H. White, 475–497. Oxford: Oxford University Press.

Engle, R. F. 2009a. *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton, N.J.: Princeton University Press.

Engle, R. F. 2009b. "The Risk That Risk Will Change." *Journal of Investment Management* 7: 1–5.

Engle, R. F., and S. Campos-Martins. 2023. "What Are the Events That Shake Our World? Measuring and Hedging Global COVOL." *Journal of Financial Economics* 147: 221–242.

Engle, R. F., E. Ghysels, and B. Sohn. 2013. "Stock Market Volatility and Macroeconomic Fundamentals." *Review of Economics and Statistics* 95: 776–797.

Engle, R. F., and V. K. Ng. 1993. "Measuring and Testing the Impact of News on Volatility." *The Journal of Finance* 48: 1749–1778.

Engle, R. F., and J. G. Rangel. 2008. "The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes." *Review of Financial Studies* 21: 1187–1222.

Ghysels, E., A. Plazzi, R. Valkanov, A. R. Serrano, and A. Dossani. 2019. "Direct Versus Iterated Multi-Period Volatility Forecasts." *Annual Review of Financial Economics* 11: 173–195.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2004. The MIDAS Touch: Mixed Data Sampling Regression Models. Working Paper, UNC and UCLA.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2006. "Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies." *Journal of Econometrics* 131: 59–95.

Ghysels, E., A. Sinko, and R. Valkanov. 2007. "MIDAS Regressions: Further Results and New Directions." *Econometric Reviews* 26: 53–90.

Giacomini, R., and H. White. 2006. "Tests of Conditional Predictive Ability." *Econometrica* 74: 1545–1578.

Glosten, L. R., R. Jagannathan, and D. E. Runkle. 1993. "On the Relation Between the Expected Value and the Volatility of Nominal Excess Return on Stocks." *Journal of Finance* 48: 1779–1801.

Halunga, A. G., and C. D. Orme. 2009. "First-Order Asymptotic Theory for Parametric Misspecification Tests of GARCH Models." *Econometric Theory* 25: 364–410.

Hansen, P. R., Z. Huang, and H. H. Shek. 2012. "Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility." *Journal of Applied Econometrics* 27: 877–906.

Hansen, P. R., and A. Lunde. 2005. "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)." *Journal of Applied Econometrics* 20: 873–889.

Hansen, P. R., A. Lunde, and J. M. Nason. 2011. "The Model Confidence Set." *Econometrica* 79: 453497.

Iacone, F., L. Rossini, and A. Viselli, 2024. Comparing Predictive Ability in Presence of Instability Over a Very Short Time. Papers 2405.11954, arXiv.org. Available at: https://arxiv.org/abs/2405.11954.

Kim, Y., and C. R. Nelson. 2013. "Pricing Stock Market Volatility: Does It Matter Whether the Volatility Is Related to the Business Cycle?" *Journal of Financial Econometrics* 12: 307–328.

Ling, S., and M. McAleer. 2002. "Stationarity and the Existence of Moments of a Family of GARCH Processes." *Journal of Econometrics* 106: 109–117.

Lundbergh, S., and T. Teräsvirta. 2002. "Evaluating GARCH Models." *Journal of Econometrics* 110: 417–435.

Nelson, D. B. 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica* 59: 347370.

Patton, A. J. 2011. "Volatility Forecast Comparison Using Imperfect Volatility Proxies." *Journal of Econometrics* 160: 246–256.

Patton, A. J. 2020. "Comparing Possibly Misspecified Forecasts." *Journal of Business & Economic Statistics* 38: 796–809.

van Dijk, D., and P. H. Franses. 2003. "Selecting a Nonlinear Time Series Model Using Weighted Tests of Equal Forecast Accuracy." *Oxford Bulletin of Economics and Statistics* 65: 727–744.

Wang, F., and E. Ghysels. 2015. "Econometric Analysis of Volatility Component Models." *Econometric Theory* 31: 362–393.

Zivot, E. 2009. "Practical Issues in the Analysis of Univariate GARCH Models." In *Handbook of Financial Time Series*, edited by T. Mikosch, J. P. Krei, R. Davis, and T. Andersen. Berlin, Heidelberg: Springer.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.