

## RESEARCH ARTICLE

# Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS models

Christian Conrad | Onno Kleen

Department of Economics, Heidelberg University, Heidelberg, Germany

**Correspondence**

Onno Kleen, Department of Economics, Heidelberg University, Bergheimer Strasse 58, 69115 Heidelberg, Germany.  
 Email: onno.kleen@awi.uni-heidelberg.de

**Summary**

We examine the properties and forecast performance of multiplicative volatility specifications that belong to the class of generalized autoregressive conditional heteroskedasticity–mixed-data sampling (GARCH-MIDAS) models suggested in Engle, Ghysels, and Sohn (*Review of Economics and Statistics*, 2013, 95, 776–797). In those models volatility is decomposed into a short-term GARCH component and a long-term component that is driven by an explanatory variable. We derive the kurtosis of returns, the autocorrelation function of squared returns, and the  $R^2$  of a Mincer–Zarnowitz regression and evaluate the QMLE and forecast performance of these models in a Monte Carlo simulation. For S&P 500 data, we compare the forecast performance of GARCH-MIDAS models with a wide range of competitor models such as HAR (heterogeneous autoregression), realized GARCH, HEAVY (high-frequency-based volatility) and Markov-switching GARCH. Our results show that the GARCH-MIDAS based on housing starts as an explanatory variable significantly outperforms all competitor models at forecast horizons of 2 and 3 months ahead.

## 1 | INTRODUCTION

The idea of modeling volatility as consisting of multiple components has a long tradition in financial econometrics (see, e.g., Ding & Granger, 1996; Engle & Lee, 1999). Early models typically featured additive volatility components and did not allow for explanatory variables in the conditional variance. More recently, the focus has shifted to multiplicative component models (see, e.g., Amado & Teräsvirta, 2013, 2017; Engle, Ghysels, & Sohn, 2013; Engle & Rangel, 2008; Han & Kristensen, 2015). In particular, the class of generalized autoregressive conditional heteroskedasticity–mixed-data sampling (GARCH-MIDAS) models proposed in Engle et al. (2013) has been proven to be useful for analyzing the link between financial volatility and the macroeconomic environment (see Asgharian, Hou, & Javed, 2013; Conrad & Loch, 2015; Dorion, 2016). In GARCH-MIDAS, a unit-variance GARCH component fluctuates around a time-varying long-term component that is a function of (macroeconomic or financial) explanatory variables. By allowing for a mixed-frequency setting, this approach bridges the gap between daily stock returns and low-frequency (e.g., monthly, quarterly) explanatory variables. For further applications of GARCH-MIDAS-type models see, for example, Conrad, Loch, and Rittler (2014), Opschoor, van Dijk, and van der Wel (2014), Dominicy and Vander Elst (2015), Lindblad (2017), Amendola, Candila, and Scognamiglio (2017), Pan, Wang, Wu, and Yin (2017), Conrad, Custovic, and Ghysels (2018), and Borup and Jakobsen (2019). For a recent survey on multiplicative component models see Amado, Silvennoinen, and Teräsvirta (2019). Throughout this paper, the GARCH-MIDAS model will be our leading example for a multiplicative component GARCH (M-GARCH) model. However, we will also discuss how the class of M-GARCH models nests other specifications such as the Markov-switching GARCH (MS-GARCH) of Haas, Mittnik, and Paoletta (2004), the spline-GARCH of Engle and Rangel (2008), and the multiplicative time-varying GARCH (MTV-GARCH) of Amado and Teräsvirta (2008).

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. Journal of Applied Econometrics published by John Wiley & Sons Ltd

Our contribution to this recent strand of literature is twofold. In the first part of this paper, we analyze several statistical properties of the GARCH-MIDAS model that have not received much attention so far. In the second part of the paper, we compare the out-of-sample (OOS) forecast performance of GARCH-MIDAS with the performance of various competitor models such as the heterogeneous autoregression (HAR) of Corsi (2009), the realized GARCH of Hansen, Huang, and Shek (2012), the high-frequency-based volatility (HEAVY) of Shephard and Sheppard (2010), and the MS-GARCH.

Our main theoretical findings can be summarized as follows. In the GARCH-MIDAS model, the kurtosis of the returns is always bigger than the kurtosis of the returns in the nested GJR-GARCH (see Glosten, Jagannathan, & Runkle, 1993) component. If the long-term component is sufficiently persistent, the autocorrelation function (ACF) of the squared returns as well as the ACF of the conditional variances is more persistent than the corresponding ACFs in the nested GJR-GARCH. Both findings suggest a multiplicative component structure in the volatility of stock returns as a potential explanation for the common failure of simple one-component GARCH models to adequately capture the stylized facts of returns and realized variances. It should also be noted that our results are remarkably similar to findings in Han (2015) on GARCH-X models, even though Han considers models with an additive explanatory variable in the conditional variance and focuses on the asymptotic limit of the sample kurtosis and the sample ACF. Further, we derive an upper bound for the population  $R^2$  in the  $k$ -step-ahead Mincer and Zarnowitz (1969) regression (henceforth MZ regression) of the squared return on the volatility forecast. We show that the population  $R^2$  decreases monotonically in the forecast horizon but increases monotonically in the variability of the long-term component. The latter feature leads to the unpleasant property that the goodness-of-fit is particularly high in situations in which the squared error loss is also high. Clearly, this finding questions the usefulness of the MZ  $R^2$  for comparing forecast accuracy across volatility regimes. In this context, we derive an explicit expression for the one-step-ahead  $R^2$  of the GARCH-MIDAS specification and obtain the results from Andersen and Bollerslev (1998) for the simple GARCH(1, 1) as a special case.

Empirically, we first evaluate the quasi-maximum likelihood estimator (QMLE) of GARCH-MIDAS models by means of a Monte Carlo simulation. We show that the QMLE is unbiased and that the asymptotic standard errors based on Wang and Ghysels (2015) are valid in the presence of exogenous explanatory variables. Further, we show that measurement error in the explanatory variable or a misspecification of the lag structure has only minor effects. We also confirm our theoretical result that the  $R^2$  of a MZ regression is highest in regimes with high volatility, although in those regimes forecast performance is the worst. Following the arguments put forth in Patton and Sheppard (2009) and Patton (2011), we use the QLIKE to evaluate the OOS forecast performance of the GARCH-MIDAS model against the MS-GARCH and the nested GARCH. We find that the correctly specified and, in most settings, even the misspecified GARCH-MIDAS models beat the competitor models.

Finally, we apply the GARCH-MIDAS model to a long time series of S&P 500 returns combined with data on US macroeconomic and financial conditions. We consider GARCH-MIDAS models with one or two explanatory variables and, for the OOS forecast evaluation, estimate all models on a rolling window using the appropriate real-time vintage data. Because macroeconomic time series are revised substantially after the first release, we avoid a “look-ahead-bias” by using real-time data. In the OOS forecast evaluation, we compare the GARCH-MIDAS with eight competitor models: Among those competitor models are the realized GARCH, the HEAVY, the MS-GARCH, and HAR models with and without leverage. We evaluate all models jointly by constructing model confidence sets (MCS) as introduced in Hansen, Lunde, and Nason (2011). For forecast horizons of 2 weeks and 1 month, the MCS consists of the realized GARCH, the HAR, and GARCH-MIDAS models with the CBOE Volatility Index (VIX) (or the VIX combined with another explanatory variable). That is, at these forecast horizons the GARCH-MIDAS is on a par with those models but beats the HEAVY as well as MS-GARCH models. At longer forecast horizons of 2 and 3 months ahead, only GARCH-MIDAS models are included in the MCS. At both horizons the GARCH-MIDAS based on housing starts achieves the lowest QLIKE. This finding is remarkable because our OOS period begins in 2010 and hence does not include the financial crisis and the collapse of the housing bubble.

To facilitate the replication of our results, we provide R packages for downloading real-time data from the ALFRED database of the Federal Reserve Bank of St. Louis (see Kleen, 2017), as well as for estimating GARCH-MIDAS models (see Kleen, 2018).<sup>1</sup>

Our paper is organized as follows: In Section 2, the M-GARCH model and our theoretical results are presented. In Section 3, we perform a simulation study and, in Section 4, we apply the GARCH-MIDAS model to S&P 500 return data. The conclusion follows in Section 5. All proofs are contained in Supporting Information Appendix A of the Online Appendix. Additional material can be found in Appendices B–H.

<sup>1</sup>The packages are available at: <https://cran.r-project.org/package=alfred> and <https://cran.r-project.org/package=mfGARCH>

## 2 | THE MULTIPLICATIVE COMPONENT GARCH MODEL

In this section, the M-GARCH model is introduced and its theoretical properties are derived. In particular, we show that the M-GARCH model inherits certain time series properties that are in line with stylized facts typically observed for financial return data but cannot be captured by simple GARCH models.

### 2.1 | Model specification

We denote daily log-returns by  $r_{i,t}$ , whereby the index  $t = 1, \dots, T$  refers to a certain period (e.g., a week or a month) and the index  $i = 1, \dots, I_t$  to days within that period. For simplicity, we model the returns as  $r_{i,t} = \mu + \varepsilon_{i,t}$ .<sup>2</sup> The M-GARCH model assumes that the scaled (demeaned) returns can be written as

$$\frac{\varepsilon_{i,t}}{\sqrt{\tau_t}} = \sqrt{g_{i,t}} Z_{i,t}, \quad (1)$$

where  $\tau_t$  is specified as a function of a (low-frequency) explanatory variable  $X_t$ ,  $g_{i,t}$  follows a GARCH equation, and  $Z_{i,t}$  is an i.i.d. innovation process with mean zero and variance one. Let  $\mathcal{F}_{i,t}$  denote the information set up to day  $i$  in period  $t$  and define  $\mathcal{F}_t := \mathcal{F}_{I_t,t}$ . If  $\tau_t$  depends on lagged values of  $X_t$  only, then

$$\sigma_{i,t}^2 := g_{i,t} \tau_t \quad (2)$$

is the conditional variance of the daily returns; that is,  $\sigma_{i,t}^2 = \text{var}(\varepsilon_{i,t} | \mathcal{F}_{i-1,t})$ . We refer to  $g_{i,t}$  as the short-term component of volatility and to  $\tau_t$  as the long-term component of volatility. Whereas  $g_{i,t}$  varies daily,  $\tau_t$  is constant across all days within period  $t$  and thus changes at the lower frequency only. The short-term component is intended to describe the well-known day-to-day clustering of volatility and is assumed to follow a mean-reverting unit-variance GJR-GARCH(1,1) process:

$$g_{i,t} = (1 - \alpha - \gamma/2 - \beta) + (\alpha + \gamma \mathbb{1}_{\{\varepsilon_{i-1,t} < 0\}}) \frac{\varepsilon_{i-1,t}^2}{\tau_t} + \beta g_{i-1,t}. \quad (3)$$

**Remark 1.** We use the convention that  $\varepsilon_{0,t} = \varepsilon_{I_{t-1},t-1}$  and  $g_{0,t} = g_{I_{t-1},t-1}$ . Similarly, we can write the long-term component as  $\tau_{i,t} = \tau_t$  for  $i = 1, \dots, n$  and  $\tau_{0,t} = \tau_{I_{t-1},t-1} = \tau_{t-1}$ . That is, for  $I_t > 1$ ,  $\tau_t$  is piecewise constant. If  $I_t = 1$ , then both components vary at the same frequency. In this case we can write  $\varepsilon_{1,t} = \varepsilon_t$ ,  $g_{1,t} = g_t$ ,  $\varepsilon_{0,t} = \varepsilon_{1,t-1} = \varepsilon_{t-1}$ , and  $g_{0,t} = g_{1,t-1} = g_{t-1}$ . Thus we can drop the index  $i$ .

A characteristic of the two-component M-GARCH model defined in Equation (1) is that the scaled returns,  $\varepsilon_{i,t}/\sqrt{\tau_t}$ , are assumed to follow a GARCH process. Hence the forcing variable in Equation (3) is  $\varepsilon_{i-1,t}^2/\tau_t$ . This feature distinguishes the two-component M-GARCH specification from standard GARCH models. In those models it is assumed that  $\tau_t = 1$  and hence the returns themselves follow a GARCH process. Similarly, additive component GARCH models, such as the model of Engle and Lee (1999), assume that  $\tau_t = 1$  and decompose  $g_{i,t}$  into two or more GARCH components (with forcing variable  $\varepsilon_{i-1,t}^2$ ). We make the following assumptions regarding the innovation process  $Z_{i,t}$  and the parameters of the short-term component.

**Assumption 1.** Let  $Z_{i,t}$  be i.i.d. with  $\mathbf{E}[Z_{i,t}] = 0$ ,  $\mathbf{E}[Z_{i,t}^2] = 1$ , and  $1 < \kappa < \infty$ , where  $\kappa = \mathbf{E}[Z_{i,t}^4]$ .

**Assumption 2.** We assume that  $\alpha > 0$ ,  $\alpha + \gamma > 0$ ,  $\beta \geq 0$ , and  $\alpha + \gamma/2 + \beta < 1$ . Moreover, the parameters satisfy the condition  $(\alpha + \gamma/2)^2 \kappa + 2(\alpha + \gamma/2)\beta + \beta^2 < 1$ .

Assumptions 1 and 2 imply that  $\varepsilon_{i,t}/\sqrt{\tau_t} = \sqrt{g_{i,t}} Z_{i,t}$  is a covariance stationary GJR-GARCH(1,1) process. The first- and second-order moments of  $g_{i,t}$  are given by  $\mathbf{E}[g_{i,t}] = 1$ ,

$$\mathbf{E}[g_{i,t}^2] = \frac{1 - (\alpha + \gamma/2 + \beta)^2}{1 - (\alpha + \gamma/2)^2 \kappa - 2(\alpha + \gamma/2)\beta - \beta^2}, \quad (4)$$

<sup>2</sup>It would be straightforward to allow for richer dynamics in the conditional mean. However, for daily return data a constant conditional mean is usually sufficient. For simplicity, in the following we refer to  $\varepsilon_{i,t}$  as the (demeaned) return.

and the fourth moment of  $\sqrt{g_{i,t}}Z_{i,t}$  is finite. The role of the second component,  $\tau_t$ , is to describe smooth movements in the conditional variance. In general, we specify  $\tau_t$  as a measurable, positive-valued function,  $f(\cdot)$ , of the present and  $K \geq 1$  lagged values of an explanatory variable  $X_t$ :

$$\tau_t = f(X_t, X_{t-1}, \dots, X_{t-K}). \quad (5)$$

The appropriate choice of the explanatory variable  $X_t$  and of the function  $f(\cdot)$  is up to the researcher and will depend on the specific application at hand.<sup>3</sup> The explanatory variable can either vary at the daily frequency (i.e.,  $I_t = 1$ ) or at a lower frequency (i.e.,  $I_t > 1$ ). Thus, the choice of  $X_t$  defines the low frequency  $t$ . In GARCH-MIDAS-type models  $\tau_t$  depends on lagged values of  $X_t$  only. By explicitly allowing  $\tau_t$  to depend on  $X_t$  in Equation (5), we ensure that our setting also covers MS-GARCH models (see Section 2.2 for details). We make the following assumption about the explanatory variable  $X_t$  and the function  $f(\cdot)$ :

**Assumption 3.** Let  $f(\cdot) > 0$  be a measurable function and  $X_t$  be a strictly stationary and ergodic time series with  $\mathbf{E}[|X_t|^q] < \infty$ , where  $q$  is sufficiently large to ensure that  $\mathbf{E}[\tau_t^2] < \infty$ .  $X_t$  is independent of  $Z_{i,t-j}$  for all  $t, i$  and  $j$ .

Note that Assumption 3 implies that  $\tau_t$  is strictly stationary (see Billingsley, 1995, p. 495), covariance stationary, and independent of the ‘GARCH part’ (i.e.  $g_{i,t-j}Z_{i,t-j}^2$ ) of the model. In empirical applications the function  $f(\cdot) > 0$  is often chosen as being linear in the lagged  $X_t$ :

$$\tau_t = m + \pi_1 X_{t-1} + \dots + \pi_K X_{t-K}. \quad (6)$$

The linear specification requires  $m > 0$  and  $\pi_l \geq 0$ , for  $l = 1, \dots, K$ , and is feasible only if  $X_t$  is a nonnegative variable. If  $X_t$  can take positive as well as negative values, it is natural to opt for an exponential specification:

$$\tau_t = \exp(m + \pi_1 X_{t-1} + \dots + \pi_K X_{t-K}). \quad (7)$$

The assumption that  $X_t$  is independent of  $Z_{i,t-j}$  for all  $t, i$ , and  $j$  might appear to be rather strong. However, without imposing any restrictions on the functional form of  $f(\cdot)$ , it greatly simplifies the analysis when discussing the statistical properties of M-GARCH models in Section 2.3. From an empirical perspective, we believe that it is reasonable to assume that a low-frequency explanatory variable  $X_t$ —such as monthly industrial production growth—is (close to being) independent of the daily innovations  $Z_{i,t-j}$ . For daily explanatory variables (e.g., measures of realized volatility) the independence assumption might appear to be restrictive. However, even if there is a dependence between the innovation to the daily returns and the daily explanatory variable, the dependence between  $\tau_t$  and  $Z_{i,t-j}$  is likely to be negligible. This is because  $\tau_t$  is a rather smooth function that is obtained as a weighted average of many lags of the daily  $X_t$ . Indeed, in Section 3 and Supporting Information Appendix D we illustrate in simulations that a mild violation of the independence assumption does not affect our main results.

It should also be noted that the same independence assumption has been previously made in related literature on M-GARCH models (see Han & Kristensen, 2015). Nevertheless, it clearly imposes a limitation that should be overcome in future work. Two examples in this direction are the estimation of GARCH-MIDAS models employing lagged values of realized variances (Wang & Ghysels, 2015) and testing for an omitted long-term component in one-component GARCH models (Conrad & Schienle, 2018).

Assumptions 1, 2, and 3 imply that the  $\varepsilon_{i,t}$  have mean zero, are uncorrelated, and have an unconditional variance given by  $\text{var}(\varepsilon_{i,t}) = \mathbf{E}[\tau_t]$ . Moreover, the unconditional variance of the squared returns is well defined:  $\text{var}(\varepsilon_{i,t}^2) = \kappa \mathbf{E}[\tau_t^2] \mathbf{E}[g_{i,t}^2] - \mathbf{E}[\tau_t]^2$ . If the long-term component is constant and chosen as  $\tau_t = \omega / (1 - \alpha - \gamma / 2 - \beta)$ , our model reduces to the GJR-GARCH with intercept  $\omega$ .

A measure that is often used to quantify the relative importance of the long-term component is the following variance ratio (see Engle et al., 2013):

$$\text{VR} = \text{var}(\log(\tau_t)) / \text{var}(\log(\tau_t g_t)), \quad (8)$$

where  $g_t = \sum_{i=1}^{I_t} g_{i,t}$ . The ratio measures how much of the total variation in the (log) conditional variance can be explained by the variation in the (log) long-term component.

<sup>3</sup>While we focus on multiplicative GARCH models, Han and Park (2014) and Han (2015) analyze the properties of a GARCH-X specification with an explanatory variable that enters additively into the conditional variance equation. See also Francq and Thieu (2019).

## 2.2 | Nested and related specifications

We first discuss two models that are directly nested in the M-GARCH setting. The two models are the GARCH-MIDAS of Engle et al. (2013) and (a restricted version of) the MS-GARCH model of Haas et al. (2004). Closely related are the Spline-GARCH of Engle and Rangel (2008) and the MTV-GARCH of Amado and Teräsvirta (2008). For further models that have a multiplicative component structure see Amado et al. (2019).

### 2.2.1 | GARCH-MIDAS

In the GARCH-MIDAS the long-term component is defined as in Equation (6) or (7), whereby the weights  $\pi_l$  are parsimoniously specified via a weighting scheme. The most common choice of long-term component is based on the exponential specification with  $\pi_l = \theta \cdot \varphi_l(w_1, w_2)$ . Here, the parameter  $\theta$  determines the sign of the effect of the lagged  $X_t$  on the long-term component and the weights  $\varphi_l(w_1, w_2) \geq 0$  are parametrized via the Beta weighting scheme

$$\varphi_l(w_1, w_2) = \frac{[l/(K+1)]^{w_1-1} \cdot [1-l/(K+1)]^{w_2-1}}{\sum_{j=1}^K [j/(K+1)]^{w_1-1} \cdot [1-j/(K+1)]^{w_2-1}}. \quad (9)$$

By construction, the weights sum to one; that is,  $\sum_{l=1}^K \varphi_l(w_1, w_2) = 1$ . It directly follows that  $E[\tau_{t+1}|F_t] = \tau_{t+1}$ . Engle et al. (2013) use monthly industrial production growth and monthly inflation as explanatory variables, whereas Conrad and Loch (2015) employ quarterly macroeconomic variables such as gross domestic product (GDP) growth. For further applications of this model see Asgharian et al. (2013), Opschoor et al. (2014), and Dorion (2016). Wang and Ghysels (2015) consider the special case that  $f(\cdot)$  is linear,  $I_t = 1$  and  $X_t = \sum_{j=0}^{J-1} \varepsilon_{t-j}^2$ . That is,  $X_t$  is the realized variance based on the last  $J$  daily returns. Note that for this specification  $X_t$  and  $Z_t$  are dependent and hence Assumption 3 would be violated.

### 2.2.2 | MS-GARCH

In MS-GARCH the returns are given by  $\varepsilon_t = \tilde{\sigma}_{X_t,t} Z_t$ , where  $\{X_t\}$  is a Markov chain with finite state space  $S = \{1, 2, \dots, s\}$  and transition matrix  $\mathbf{P}$  with typical element  $p_{i,j} = P(X_t = j | X_{t-1} = i)$ . A restricted version of the MS-GARCH model of Haas et al. (2004) is nested in our setting with  $I_t = 1$ . This is best illustrated in the case of  $s = 2$ : We assume that the conditional variances in the regimes differ in the intercepts but have the same ARCH and GARCH parameters; for example,  $\tilde{\sigma}_{k,t}^2 = \omega_k + \alpha \varepsilon_{t-1}^2 + \beta \tilde{\sigma}_{k,t-1}^2$ ,  $k \in S$ . Defining  $\tau_t = [(2 - X_t)\omega_1 + (X_t - 1)\omega_2]/(1 - \alpha - \beta)$ , we can rewrite the returns as  $\varepsilon_t = \sqrt{\tilde{\sigma}_{X_t,t}^2} Z_t = \sqrt{g_t \tau_t} Z_t$ , where  $g_t = (1 - \alpha - \beta) + (\alpha Z_{t-1}^2 + \beta)g_{t-1}$ . Thus the conditional variance has a multiplicative structure. In the following, we will refer to this model as MS-GARCH with time-varying intercept (MS-GARCH-TVI). Stationarity conditions for MS-GARCH models can be found in Haas et al. (2004).

### 2.2.3 | Spline-GARCH and multiplicative time-varying (MTV) GARCH

In both models it is assumed that  $I_t = 1$ . The spline-GARCH model specifies the long-term component as a spline function and chooses  $X_t = t$ . Similarly, in the MTV-GARCH  $f(\cdot)$  is specified in terms of logistic transition functions and  $X_t = t/T$  is the rescaled time. Thus in both models the long-term component is a deterministic function of time and hence Assumption 3 is violated.

## 2.3 | Properties of the M-GARCH

In the following, we derive properties of M-GARCH models for which Assumptions 1, 2, and 3 are satisfied.

### 2.3.1 | Kurtosis and autocorrelation function

Financial returns are often found to be leptokurtic. Hence a desirable feature of a volatility model is that it generates returns with a kurtosis that is similar to the one empirically observed for financial returns. Under Assumptions 1, 2, and 3, the kurtosis of the returns defined in Equation (1) is given by



$$\mathcal{K}^{\text{MG}} = \frac{\mathbf{E}[\varepsilon_{i,t}^4]}{(\mathbf{E}[\varepsilon_{i,t}^2])^2} = \frac{\mathbf{E}[\sigma_{i,t}^4]}{(\mathbf{E}[\sigma_{i,t}^2])^2} \kappa > \kappa.$$

Thus the kurtosis of the M-GARCH process is larger than the kurtosis of the innovation  $Z_{i,t}$ . This is a well-known feature of GARCH-type processes. The following proposition relates the kurtosis  $\mathcal{K}^{\text{MG}}$  of the M-GARCH to the kurtosis  $\mathcal{K}^{\text{GA}}$  of the nested GARCH(1, 1).

**Proposition 1.** *Under Assumptions 1-3, the kurtosis  $\mathcal{K}^{\text{MG}}$  of an M-GARCH process is given by*

$$\mathcal{K}^{\text{MG}} = \frac{\mathbf{E}[\tau_t^2]}{\mathbf{E}[\tau_t]^2} \cdot \mathcal{K}^{\text{GA}} \geq \mathcal{K}^{\text{GA}},$$

where  $\mathcal{K}^{\text{GA}} = \kappa \cdot \mathbf{E}[g_{i,t}^2]$  is the kurtosis of the nested GARCH process and where the equality holds if and only if  $\tau_t$  is constant.

Hence, for nonconstant  $\tau_t$ , the kurtosis  $\mathcal{K}^{\text{MG}}$  is the product of  $\mathcal{K}^{\text{GA}}$  and the ratio  $\mathbf{E}[\tau_t^2]/\mathbf{E}[\tau_t]^2 > 1$ . When  $\tau_t = \omega/(1 - \alpha - \gamma/2 - \beta)$  is constant, Proposition 1 nests the kurtosis of the GJR-GARCH model. Thus, for volatile long-term components the kurtosis of an M-GARCH process can be much larger than the kurtosis of the nested GARCH model.<sup>4</sup> Specifically, Proposition 1 holds for the GARCH-MIDAS and for the MS-GARCH-TVI defined in Section 2.2.

The empirical ACFs of volatility proxies such as squared returns or realized variances are known to be very persistent (see, e.g., Andersen, Bollerslev, Diebold, & Labys, 2003; Ding, Granger, & Engle, 1993). In particular, squared returns are often found to decay more slowly than the exponentially decaying ACF implied by the simple GARCH(1, 1) model. In the literature on GARCH models, this is usually interpreted as either evidence for long memory (see, e.g., Baillie, Bollerslev, & Mikkelsen, 1996), structural breaks (see, e.g., Hillebrand, 2005), or an omitted persistent covariate (see Han & Park, 2014) in the conditional variance.

The following propositions show that the theoretical ACFs of the M-GARCH process have a much slower decay than the ACF of the nested GARCH component if the long-term component is sufficiently persistent. Hence the multiplicative structure provides an alternative explanation for the empirical observation of highly persistent ACFs of squared returns or realized variances. For Propositions 2 and 3, we consider the case that both components are varying at the same frequency; that is, the length of the period  $t$  is one day ( $I_t = 1$ ).

**Proposition 2.** *If  $I_t = 1$  and Assumptions 1-3 are satisfied, the ACF,  $\rho_k^{\text{MG}}(\varepsilon^2)$ , of the squared returns from an M-GARCH process is given by*

$$\rho_k^{\text{MG}}(\varepsilon^2) = \text{corr}(\varepsilon_t^2, \varepsilon_{t-k}^2) = \rho_k^\tau \frac{\text{var}(\tau_t)}{\text{var}(\varepsilon_t^2)} + \rho_k^{\text{GA}} \frac{\text{var}(g_t Z_t^2)}{\text{var}(\varepsilon_t^2)} (\rho_k^\tau \text{var}(\tau_t) + \mathbf{E}[\tau_t]^2) \quad (10)$$

with  $\rho_k^\tau = \text{corr}(\tau_t, \tau_{t-k})$  and

$$\rho_k^{\text{GA}} = \text{corr}(g_t Z_t^2, g_{t-k} Z_{t-k}^2) = (\alpha + \gamma/2 + \beta)^{k-1} \frac{(\alpha + \gamma/2)[1 - (\alpha + \gamma/2)\beta - \beta^2]}{1 - 2(\alpha + \gamma/2)\beta - \beta^2}$$

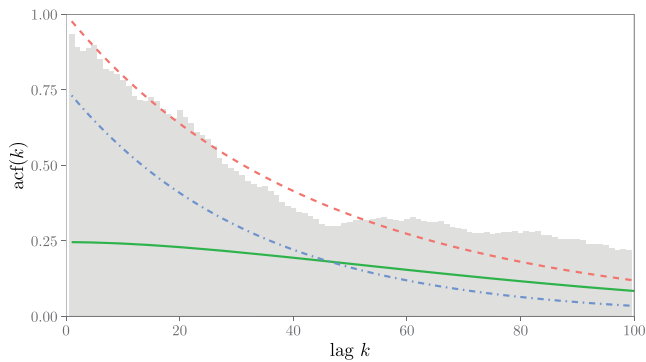
being the ACF of the GJR-GARCH component.<sup>5</sup>

Proposition 2 shows that the ACF of the squared returns is given by the sum of two terms: the first term corresponds to the ACF of the long-term component  $\rho_k^\tau$  times a constant, whereas the second term equals the exponentially decaying ACF of the nested GARCH model  $\rho_k^{\text{GA}}$  times a term that depends again on  $\rho_k^\tau$ . Hence, if  $\tau_t$  is sufficiently persistent,  $\rho_k^{\text{MG}}(\varepsilon^2)$  will essentially behave as  $\rho_k^\tau$  for  $k$  large.<sup>6</sup> For  $\tau_t$  being constant, the first term in Equation (10) is equal to zero and the second

<sup>4</sup>Han (2015) obtains a similar result for the sample kurtosis of the returns from a GARCH-X model with a covariate that can either be stationary or nonstationary.

<sup>5</sup>Note that  $\rho_k^{\text{GA}}$  reduces to the ACF of a (symmetric) GARCH(1, 1) when  $\gamma = 0$  (see Karanasos, 1999).

<sup>6</sup>Again, Han (2015) also obtains a bicomponent structure for the sample ACF of the squared returns from a GARCH-X model with a fractionally integrated covariate. Similarly, Han and Kristensen (2015) show that the empirical ACF in a multiplicative model can display long-memory-type behavior.



**FIGURE 1** Autocorrelation function of the volatility process in a GARCH-MIDAS model. We depict the ACF of the volatility process in a GARCH-MIDAS model (red, dashed) and its components: the first (green, solid) and second term (blue, dot-dashed) in Equation (11). The long-term component is defined as in Equations (7) and (9) with  $m = -0.1$ ,  $\theta = 0.3$ ,  $w_1 = 1$ ,  $w_2 = 5$ , and  $K = 264$ . The explanatory variable is given by  $X_t = \phi X_{t-1} + \xi_t$ ,  $\xi_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\xi^2)$ , where  $\phi = 0.98$  and  $\sigma_\xi^2 = 0.35^2$ . The GARCH(1, 1) parameters are  $\alpha = 0.06$ ,  $\beta = 0.91$ , and  $\gamma = 0$ . Moreover, we set  $\kappa = 3$ . Bars in light gray display the empirical autocorrelation of S&P 500 daily realized variances between 2000:M1 and 2018:M4 as measured by Hansen and Lunde (2014). For details see Section 4 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

term reduces to the ACF of an asymmetric GARCH(1, 1). Also, note that the ratio  $\text{var}(\tau_t)/\text{var}(\varepsilon_t^2)$  is closely related to the variance ratio defined in Equation (8) and measures how much of the variation in the squared returns can be attributed to the variation in the long-term component; that is, it measures the importance of the long-term component.

Haas et al. (2004, p. 503) make a similar observation for the MS-GARCH-TVI model that we discussed in Section 2.2. For this model, they show that the autocorrelations of the squared returns decay at a rate of  $\max\{\alpha + \beta, \varpi\}$ , where  $\varpi = p_{1,1} + p_{2,2} - 1$  is the degree of persistence due to the Markov effects.<sup>7</sup> If  $\varpi$  is close to one—that is, if the long-term component is very persistent—the decay rate of this component dominates the decay of the autocorrelation function.

A standard misspecification test for GARCH models is the Ljung–Box statistic applied to the squared deGARCHed residuals,  $\varepsilon_t^2/g_t$ . The result in Proposition 2 may explain why in empirical applications the null hypothesis of this test is often rejected. In the multiplicative model, the ACF of the squared deGARCHed residuals is given by  $\rho_k^\tau \cdot \text{var}(\tau_t)/(\kappa \mathbf{E}[\tau_t^2] - \mathbf{E}[\tau_t]^2)$ , which follows the rate of decay of the long-term component and hence is still persistent. Using similar arguments to those in the proof of Proposition 2, we can derive the ACF of  $\sigma_t^2$ .

**Proposition 3.** *If  $I_t = 1$  and Assumptions 1–3 are satisfied, the ACF,  $\rho_k^{MG}(\sigma^2)$ , of  $\sigma_t^2$  is given by*

$$\rho_k^{MG}(\sigma^2) = \text{corr}(\sigma_t^2, \sigma_{t-k}^2) = \rho_k^\tau \frac{\text{var}(\tau_t)}{\text{var}(\sigma_t^2)} + \rho_k^g \frac{\text{var}(g_t)}{\text{var}(\sigma_t^2)} (\rho_k^\tau \text{var}(\tau_t) + \mathbf{E}[\tau_t]^2) \quad (11)$$

with  $\rho_k^\tau$  as before and  $\rho_k^g = \text{corr}(g_t, g_{t-k}) = (\alpha + \gamma/2 + \beta)^k$  being the ACF of the  $g_t$  component.

Again, Assumption 3 holds for the GARCH-MIDAS and the MS-GARCH-TVI.

The implications of Proposition 3 are depicted in Figure 1. The bars in light gray display the empirical ACF of the daily S&P 500 realized variances for the 2000:M1 to 2018:M4 period.<sup>8</sup> The autocorrelations were estimated using the instrumental variables estimator suggested in Hansen and Lunde (2014). We employ their preferred specification, a two-stage least squares estimator in which lagged realized variances of order 4–10 are used as instrumental variables (see Hansen & Lunde, 2014, p. 82). By choosing appropriate parameter values for a GARCH-MIDAS process, we obtain an ACF of  $\sigma_t^2$  (dashed red line), which behaves very similar to the empirical ACF of the realized volatilities. The figure shows that the second term—that is, the ACF of  $g_t$  (dot-dashed blue line)—determines the decay behavior of  $\rho_k(\sigma^2)^{MG}$  when  $k$  is small, whereas the first term—that is, the ACF of  $\tau_t$  (solid green line)—dominates when  $k$  is large. Finally, it is important to note that although our results on the kurtosis and the ACFs are presented for a GJR-GARCH(1, 1) short-term component, they directly extend to a covariance stationary GJR-GARCH( $p, q$ ) component.

### 2.3.2 | Forecast evaluation with MZ regression

In empirical applications, the coefficient of determination from an MZ regression is often used as a measure of forecast accuracy. In this section, we will argue against using this measure when comparing forecast performance across volatility regimes. We now exclusively focus on the case of a GARCH-MIDAS. We assume that forecasts are produced on the last day  $I_t$  of period  $t$  and denote the  $k$ -step-ahead volatility forecast by  $h_{k,t+1|t}$  with  $k \leq I_{t+1}$ . The optimal forecast from

<sup>7</sup>Haas et al. (2004) consider a symmetric GARCH. Hence the persistence in the GARCH component is  $\alpha + \beta$ .

<sup>8</sup>The underlying data will be described in detail in Section 4.1.

the GARCH-MIDAS is  $h_{k,t+1|t} = \mathbf{E} \left[ \sigma_{k,t+1}^2 | \mathcal{F}_t \right] = \tau_{t+1} g_{k,t+1|t}$ , where  $g_{k,t+1|t} = \mathbf{E} \left[ g_{k,t+1} | \mathcal{F}_t \right] = 1 + (\alpha + \gamma/2 + \beta)^{k-1} (g_{1,t+1|t} - 1)$ . When evaluating the volatility forecast, one has to deal with the problem that the true conditional variance,  $\sigma_{k,t+1}^2$ , is unobservable. Patton (2011) discusses the situation in which the forecast evaluation is based on some conditionally unbiased volatility proxy  $\hat{\sigma}_{k,t+1}^2$  instead. He defines a loss function  $L \left( \sigma_{k,t+1}^2, h_{k,t+1|t} \right)$  as *robust* if the expected loss ranking of two competing forecasts is preserved when replacing  $\sigma_{k,t+1}^2$  by  $\hat{\sigma}_{k,t+1}^2$ . In the MZ regression  $\sigma_{k,t+1}^2$  is often replaced by the conditionally unbiased but noisy proxy  $\hat{\sigma}_{k,t+1}^2 = \varepsilon_{k,t+1}^2$ .<sup>9</sup>

The MZ regression for evaluating the  $k$ -step-ahead volatility forecast is given by

$$\varepsilon_{k,t+1}^2 = \delta_0 + \delta_1 h_{k,t+1|t} + \eta_{k,t+1}. \quad (12)$$

We denote the respective coefficient of determination by  $R_k^2$ . As shown in Hansen and Lunde (2006), the ranking of competing one-step-ahead volatility forecasts based on the  $R_1^2$  of the MZ regression is robust to using the proxy  $\varepsilon_{1,t+1}^2$  instead of the latent conditional variance  $\sigma_{1,t+1}^2$  as the dependent variable. For  $h_{k,t+1|t} = \tau_{t+1} g_{k,t+1|t}$ , the population parameters of the MZ regression are given by  $\delta_0 = 0$  and  $\delta_1 = 1$  and hence the population  $R_k^2$  can be written as

$$R_k^2 = 1 - \frac{\text{var}(\eta_{k,t+1})}{\text{var}(\varepsilon_{k,t+1}^2)} = 1 - \frac{\mathbf{E} \left[ \text{SE} \left( \varepsilon_{k,t+1}^2, h_{k,t+1|t} \right) \right]}{\text{var}(\varepsilon_{k,t}^2)}, \quad (13)$$

where we use that the variance of  $\eta_{k,t+1}$  equals the expected squared error (SE) loss of the forecast evaluated against  $\varepsilon_{k,t+1}^2$ ; that is,  $\mathbf{E} \left[ \text{SE} \left( \varepsilon_{k,t+1}^2, h_{k,t+1|t} \right) \right] = \mathbf{E} \left[ \left( \varepsilon_{k,t+1}^2 - h_{k,t+1|t} \right)^2 \right]$ . Using that  $\mathbf{E} \left[ \varepsilon_{k,t+1}^2 | \mathcal{F}_{k-1,t+1} \right] = \sigma_{k,t+1}^2$ , it follows that

$$\mathbf{E} \left[ \text{SE} \left( \varepsilon_{k,t+1}^2, h_{k,t+1|t} \right) \right] = \mathbf{E} \left[ \text{SE} \left( \sigma_{k,t+1}^2, h_{k,t+1|t} \right) \right] + (\kappa - 1) \mathbf{E} \left[ \sigma_{k,t+1}^4 \right]. \quad (14)$$

That is, the expected SE based on the noisy proxy equals the expected SE based on the latent volatility plus a term that depends on the fourth moment,  $\kappa$ , of  $Z_{i,t}$  and the expected value of the squared conditional variance. Hence using a noisy proxy for forecast evaluation can lead to a substantially higher expected SE than the expected SE based on the latent volatility. Patton, (2011, p. 248) basically makes the same point by arguing that “although the ranking obtained from a robust loss function will be invariant to noise in the proxy, the actual *level* of expected loss obtained using a proxy will be larger than that which would be obtained when using the true conditional variance.”

Using the insight from Equation (14) that the expected SE loss based on the noisy proxy is at least  $(\kappa - 1) \mathbf{E} \left[ \sigma_{k,t}^4 \right]$ , we obtain the following bound:

$$R_k^2 \leq 1 - \frac{(\kappa - 1) \mathbf{E} \left[ \sigma_{k,t}^4 \right]}{\kappa \mathbf{E} \left[ \sigma_{k,t}^4 \right] - \left( \mathbf{E} \left[ \sigma_{k,t}^2 \right] \right)^2} = \frac{1 - \left( \mathbf{E} \left[ \sigma_{k,t}^2 \right] \right)^2 / \mathbf{E} \left[ \sigma_{k,t}^4 \right]}{\kappa - \left( \mathbf{E} \left[ \sigma_{k,t}^2 \right] \right)^2 / \mathbf{E} \left[ \sigma_{k,t}^4 \right]} < \frac{1}{\kappa}. \quad (15)$$

The upper bound for  $R_k^2$  given by Equation (15) nicely illustrates that a low  $R_k^2$  is not necessarily evidence for model misspecification but can simply be due to using a noisy volatility proxy. This point has been made before by Andersen and Bollerslev (1998), but for the special case of a one-step-ahead forecast from a GARCH(1, 1).<sup>10</sup> Note that the result in Equation (15) does not depend on the two-component structure of the model but is true for any conditionally heteroskedastic process.

Next, we derive an explicit expression for the MZ  $R_k^2$  of the GARCH-MIDAS model.

<sup>9</sup>To illustrate the severeness of the noise, consider an example with  $Z_{k,t+1} \sim \mathcal{N}(0, 1)$ . Then  $\varepsilon_{k,t+1}^2$  will either over- or underestimate the true  $\sigma_{k,t+1}^2$  by more than 50% with a probability of about 74%.

<sup>10</sup>See Andersen, Bollerslev, and Meddahi (2005) for a model-free adjustment procedure for the predictive  $R^2$ .



**Proposition 4.** If  $\varepsilon_{k,t+1}^2$  follows a GARCH-MIDAS process, Assumptions 1-3 are satisfied, and  $h_{k,t+1|t} = \tau_{t+1}g_{k,t+1|t}$ , then the population  $R_k^2$  of the MZ regression is given by

$$R_k^2 = \frac{\text{var}(h_{k,t+1|t})}{\text{var}(\varepsilon_{k,t+1}^2)} = \frac{\mathbf{E}[g_{k,t+1|t}^2] \mathbf{E}[\tau_{t+1}^2] - \mathbf{E}[\tau_{t+1}]^2}{\mathbf{E}[g_{k,t+1}^2] \mathbf{E}[\tau_{t+1}^2] \kappa - \mathbf{E}[\tau_{t+1}]^2} \quad (16)$$

with  $\mathbf{E}[g_{k,t+1}^2]$  as in Equation (4) and

$$\mathbf{E}[g_{k,t+1|t}^2] = 1 + (\alpha + \gamma/2 + \beta)^{2(k-1)} (\mathbf{E}[g_{1,t+1}^2] - 1). \quad (17)$$

We obtain the following two properties:

1.  $R_k^2$  decreases monotonically with increasing forecast horizon  $k$  and, in the limit, converges<sup>11</sup> to  $R_\infty^2 = \text{var}(\tau_{t+1})/\text{var}(\varepsilon_{k,t+1}^2)$ .
2.  $R_k^2$  increases monotonically in  $\mathbf{E}[\tau_{t+1}^2]$ .

The first property rests on the insight that the forecast of the GARCH component converges to one (as  $k \rightarrow \infty$ ) and hence the MZ regression reduces to a regression of  $\varepsilon_{k,t+1}^2$  on a constant and  $\tau_{t+1}$ . Thus  $R_\infty^2$  can be interpreted as the fraction of the total variation in daily returns that can be attributed to the variation in the long-term component. Note that  $R_\infty^2$  corresponds to the weight that is attached to the ACF of  $\tau_t$  in the first term in Equation (10).

Second, the result that  $R_k^2$  increases when  $\tau_{t+1}$  gets more volatile implies that for the very same model the  $R_k^2$  will be higher in high-volatility regimes (i.e., when the squared error loss is high) than in low-volatility regimes (i.e., when the squared error loss is low). This can be misleading when calculating  $R_k^2$  for different regimes. The intuition is best illustrated when looking at one-step-ahead forecasts. Equations (13) and (14) imply

$$R_1^2 = 1 - \frac{\mathbf{E}[\text{SE}(\varepsilon_{1,t+1}^2, h_{1,t+1|t})]}{\text{var}(\varepsilon_{1,t+1}^2)} = 1 - \frac{(\kappa - 1)\mathbf{E}[g_{1,t+1}^2] \mathbf{E}[\tau_{t+1}^2]}{\mathbf{E}[g_{1,t+1}^2] \mathbf{E}[\tau_{t+1}^2] \kappa - \mathbf{E}[\tau_{t+1}]^2}. \quad (18)$$

When  $\mathbf{E}[\tau_{t+1}^2]$  is increasing, the unconditional variance of returns rises at a faster rate than the expected squared error and hence the MZ  $R_1^2$  is increasing. We can express  $R_1^2$  directly as a function of the model parameters:

**Lemma 1.** If  $\varepsilon_{k,t+1}^2$  follows a GARCH-MIDAS process, Assumptions 1-3 are satisfied, and  $h_{1,t+1|t} = \tau_{t+1}g_{1,t+1|t}$ , then the population  $R_1^2$  of the MZ regression is given by

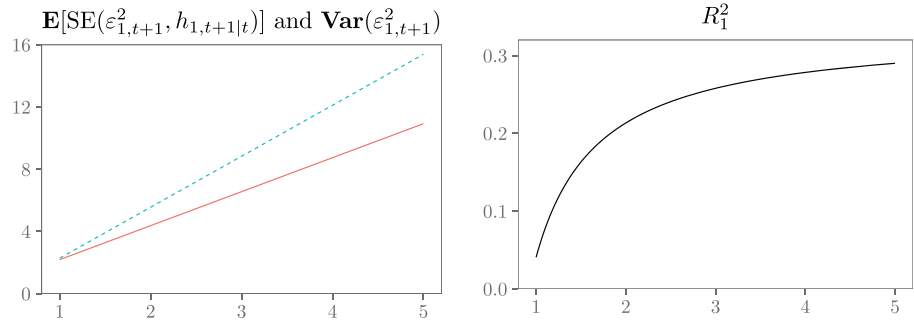
$$R_1^2 = \frac{[1 - (\alpha + \gamma/2 + \beta)^2]\mathbf{E}[\tau_{t+1}^2] - [1 - (\alpha + \gamma/2)^2\kappa - 2(\alpha + \gamma/2)\beta - \beta^2]\mathbf{E}[\tau_{t+1}]^2}{[1 - (\alpha + \gamma/2 + \beta)^2]\mathbf{E}[\tau_{t+1}^2] \kappa - [1 - (\alpha + \gamma/2)^2\kappa - 2(\alpha + \gamma/2)\beta - \beta^2]\mathbf{E}[\tau_{t+1}]^2}. \quad (19)$$

For  $\tau_{t+1}$  being constant and  $\gamma = 0$ , Equation (19) is reduced to the expression in Andersen and Bollerslev (1998, p. 892) for the symmetric GARCH(1, 1); that is,  $R_1^2 = \alpha^2/(1 - 2\alpha\beta - \beta^2)$ .

The effect of an increase in  $\mathbf{E}[\tau_{t+1}^2]$  on  $\mathbf{E}[\text{SE}(\varepsilon_{1,t+1}^2, h_{1,t+1|t})]$ ,  $\text{var}(\varepsilon_{1,t+1}^2)$  and  $R_1^2$  is illustrated in Figure 2. We set  $\mathbf{E}[\tau_{t+1}] = 1$ ,  $\alpha = 0.05$ ,  $\beta = 0.92$ ,  $\gamma = 0$ , and  $\kappa = 3$ . As expected, the left-hand panel shows that the expected squared error increases when we move from a low-volatility regime (say  $\mathbf{E}[\tau_{t+1}^2] = 2$ ) to a high-volatility regime (say  $\mathbf{E}[\tau_{t+1}^2] = 5$ ). However, it also shows that the variance of the returns is increasing even faster (as evident from the larger slope coefficient). The right-hand panel of Figure 2 shows that this translates into an increase of  $R_1^2$ . That is, although the expected squared error increases, the “forecast accuracy” as measured by  $R_1^2$  increases as well. In this regard, the  $R^2$  of an MZ regression should be interpreted as a measure of *relative* forecast accuracy; that is, forecast accuracy is measured

<sup>11</sup>Although by assumption  $k \leq I_t$  in our setting, we can think of, for example, a semiannual period and daily volatility forecasts. In this case  $k$  can be at most 132 ( $= 6 \cdot 22$ ). For such a large  $k$  and under reasonable assumptions on the GARCH parameters, we have  $\mathbf{E}[g_{132,t+1|t}^2] \approx 1$ .

**FIGURE 2**  $\mathbf{E} \left[ \text{SE} \left( \varepsilon_{1,t+1}^2, h_{1,t+1|t} \right) \right]$ ,  $\text{var} \left( \varepsilon_{1,t+1}^2 \right)$ , and  $\text{MZ } R_1^2$  as a function of  $\mathbf{E} \left[ \tau_{t+1}^2 \right]$ . The left-hand panel shows  $\mathbf{E} \left[ \text{SE} \left( \varepsilon_{1,t+1}^2, h_{1,t+1|t} \right) \right]$  (red, solid) and  $\text{var} \left( \varepsilon_{1,t+1}^2 \right)$  (blue, dashed) as a function of  $\mathbf{E} \left[ \tau_{t+1}^2 \right]$  (see Equation (18)). The right-hand panel depicts the corresponding population Mincer-Zarnowitz  $R_1^2$  as a function of  $\mathbf{E} \left[ \tau_{t+1}^2 \right]$ . We set  $\mathbf{E}[\tau_{t+1}] = 1$ ,  $\alpha = 0.05$ ,  $\beta = 0.92$ ,  $\gamma = 0$ , and  $\kappa = 3$  [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



relative to the unconditional variance of the process. In contrast, the squared error loss is a measure of *absolute* forecast accuracy. Note that for rather moderate values of  $\mathbf{E} \left[ \tau_{t+1}^2 \right]$  the coefficient of determination is already close to its upper bound of  $1/3$ .

Although the previous results are derived under the assumption that squared daily returns are used as the volatility proxy, it is true that the main insights still hold when using a better volatility proxy. For example, consider the hypothetical case of observing  $\sigma_{k,t+1}^2$  ex post. Then, for  $k \rightarrow \infty$ , we obtain  $R_\infty^2 = \text{var}(\tau_{t+1})/\text{var}(\sigma_{k,t+1}^2) < 1$ . Hence  $R_\infty^2$  would still vary across volatility regimes and increase in the variance of the long-term component. In the simulation in Section 3, we will consider the case in which the realized variance is used as a proxy for  $\sigma_{k,t+1}^2$ .

Finally, we consider cumulative volatility forecasts. The MZ regression for evaluating the cumulative  $k$ -day-ahead volatility forecast is given by

$$\widetilde{\text{RV}}_{1:k,t+1} = \tilde{\delta}_0 + \tilde{\delta}_1 h_{1:k,t+1|t} + \eta_{1:k,t+1},$$

where the latent variance is proxied by the realized variance  $\widetilde{\text{RV}}_{1:k,t+1} = \sum_{i=1}^k \varepsilon_{i,t+1}^2$  (purely based on daily return data) and  $h_{1:k,t+1|t} = \sum_{j=1}^k h_{j,t+1|t}$ . The corresponding  $R_{1:k}^2$  is given by

$$R_{1:k}^2 = \frac{\text{var}(h_{1:k,t+1|t})}{\text{var}(\widetilde{\text{RV}}_{1:k,t+1})} = \frac{\mathbf{E} \left[ \tau_{t+1}^2 \right] \mathbf{E} \left[ \left( \sum_{i=1}^k g_{i,t+1|t} \right)^2 \right] - k^2 \mathbf{E}[\tau_{t+1}]^2}{\mathbf{E} \left[ \tau_{t+1}^2 \right] \mathbf{E} \left[ \left( \sum_{i=1}^k g_{i,t} Z_{i,t}^2 \right)^2 \right] - k^2 \mathbf{E}[\tau_{t+1}]^2}. \quad (20)$$

As before, one can show that  $R_{1:k}^2$  increases monotonically in  $\mathbf{E} \left[ \tau_{t+1}^2 \right]$ .

## 2.4 | Forecasting long-term volatility

In the empirical application and in the simulation in Section 3 we also consider forecasting volatility for horizons that are beyond one low-frequency period. The optimal forecast  $h_{k,t+s|t}$  with  $s > 1$  is then given by  $\mathbf{E}[\tau_{t+s}|\mathcal{F}_t]\mathbf{E}[g_{k,t+s}|\mathcal{F}_t]$ . It is straightforward to obtain  $g_{k,t+s|t} = \mathbf{E}[g_{k,t+s}|\mathcal{F}_t] = 1 + (\alpha + \gamma/2 + \beta)^{(I_{t+1} + \dots + I_{t+s-1} + k-1)}(g_{1,t+1|t} - 1)$ . Because we do not explicitly model the dynamics of  $X_t$ , we are unable to obtain  $\mathbf{E}[\tau_{t+s}|\mathcal{F}_t]$ . Instead, based on the information set  $\mathcal{F}_t$ , we forecast  $\tau_{t+s}$  by  $\tau_{t+1}$ . Holding the long-term component constant when forecasting is reasonable if  $\tau_t$  changes smoothly and the forecast horizon is not “too large.” Otherwise, one may use predictions of  $X_t$ —for example, survey or time series forecasts—for calculating predictions of  $\tau_t$  (see Conrad & Loch, 2015).

### 3 | SIMULATION

In this section, we mainly focus on M-GARCH models from the GARCH-MIDAS class. Since asymptotic theory for the QMLE is available only for the special case of a GARCH-MIDAS with realized volatility as the explanatory variable (see Wang & Ghysels, 2015), we first evaluate the finite-sample performance of the QMLE in a Monte Carlo simulation. Second, we compare the QMLE of the correctly specified model with the QMLE of misspecified models. We consider misspecification in terms of (i) lag length  $K$ , (ii) the explanatory variable being measured with noise, (iii) both, or (iv) omitting the long component completely. Finally, within the Monte Carlo simulation we evaluate the OOS forecast performance and provide empirical support for the theoretical results in Section 2.3.2. For each model specification, we perform 2,000 Monte Carlo replications.

#### 3.1 | Data generating process

We simulate an intraday version of the two-component GARCH model as

$$\varepsilon_{n,i,t} = \sqrt{g_{i,t}\tau_t}Z_{n,i,t}/\sqrt{N}, \quad (21)$$

where the index  $n = 1, \dots, N$  now denotes the intraday frequency. The  $Z_{n,i,t}$  are assumed to be i.i.d. and follow either a standard normal or a standardized Student  $t$  distribution with five degrees of freedom. We generate  $N = 48$  intraday returns. Hence, by aggregating returns to a daily frequency,  $\varepsilon_{i,t} = \sum_{n=1}^N \varepsilon_{n,i,t}$ , the model in Equation (21) is consistent with our daily model.<sup>12</sup> Simulating intraday returns allows us to calculate the daily realized variance,  $RV_{i,t} = \sum_{n=1}^N \varepsilon_{n,i,t}^2$ , as a precise measure of the daily variance. Similarly, we obtain the realized variance over the first  $k$  days of month  $t$  as  $RV_{1:k,t} = \sum_{i=1}^k RV_{i,t}$ . We simulate data for a period of 40 years of intradaily returns, from which we construct 10,560 daily return and realized variance observations. The parameters of the GARCH component,  $g_{i,t}$ , are given by  $\alpha = 0.06$ ,  $\beta = 0.91$ , and  $\gamma = 0$ . We consider two alternative specifications of the long-term component:

*Monthly  $\tau_t$ .* The first specification assumes a mixed-frequency setting with  $\tau_t$  fluctuating at a monthly frequency. We assume that each month consists of  $I_t = 22$  days. As in Equation (7), we choose an exponential specification for the long-term component and specify the MIDAS weights according to the Beta weighting scheme in Equation (9) with  $m = 0.1$ ,  $\theta = 0.3$ ,  $w_1 = 1$ ,  $w_2 = 4$ , and  $K = 36$ . The choice of 3 years as MIDAS lag length follows Conrad and Loch (2015). Setting  $w_2 = 4$  implies a monotonically decaying weighting scheme with weights close to zero for lags greater than two-thirds of  $K$ . The explanatory variable  $X_t$  is assumed to follow an AR(1) process,  $X_t = \phi X_{t-1} + \xi_t$ ,  $\xi_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\xi^2)$ , with  $\phi = 0.9$  and  $\sigma_\xi^2 = 0.3^2$ . When averaged over the 2,000 Monte Carlo simulations, these parameter values lead to an empirical VR of 18.60%/18.09% for normally/Student  $t$  distributed innovations (recall that the VR was defined in Equation (8)).

*Daily  $\tau_t$ .* The second specification assumes that both components fluctuate at a daily frequency (i.e.,  $I_t = 1$ ). The parameters of the long-term component are chosen as  $m = -0.1$ ,  $\theta = 0.3$ ,  $w_1 = 1$ ,  $w_2 = 5$ , and  $K = 264$ . Choosing a lag length of roughly 1 year is motivated by our empirical results in Section 4 when estimating a GARCH-MIDAS model using  $RV_{i,t}$  as the explanatory variable. In addition, we choose  $\phi = 0.98$  and  $\sigma_\xi^2 = 0.2^2$ . In the simulations, the former choice leads to an average VR of 32.49%/31.66% for normally/Student  $t$  distributed innovations.

#### 3.2 | Parameter estimates

##### 3.2.1 | Correctly specified models: Bias and asymptotic standard errors

We use the first 20 years of simulated data as the “in-sample” period to obtain QML estimates of the model parameters. Table 1 reports the average bias of the QMLE across the 2,000 Monte Carlo simulations. In panels A/B the innovations  $Z_{n,i,t}$  are normally/Student  $t$  distributed. First, we focus on panel A. In this case the density is correctly specified and the QMLE is the maximum likelihood estimator. Note that for all parameters except  $w_2$  the average bias is close to zero when the conditional variance is correctly specified (i.e., with MIDAS lag length of  $K = 36$  (monthly) and  $K = 264$  (daily) respectively). For  $w_2$  we clearly observe an upward bias.<sup>13</sup> Based on the 2,000 Monte Carlo replications, we also calculate

<sup>12</sup>Alternatively, we simulated the intraday returns using a stochastic volatility model that is consistent with our GARCH-MIDAS setting. The corresponding results, which are very similar to those based on the specification in Equation (21), are presented in Supporting Information Appendix E.

<sup>13</sup>Figure C.1 in the Supporting Information Appendix compares the histogram of the standardized parameter estimates over the 2,000 Monte Carlo replications with a standard normal distribution. The figure shows that for all parameters except  $w_2$  the empirical distribution of the parameter estimates is very well approximated by the normal distribution.

TABLE 1 Monte Carlo parameter estimates

		$\alpha$	$\beta$	$m$	$\theta$	$w_2$	$\kappa - 3$
<b>Panel A: <math>Z_{n,i,t}</math> normally distributed</b>							
Monthly $\tau_t$	GARCH-MIDAS (36)	-0.000 {0.008} (0.009)	-0.004 {0.014} (0.015)	-0.007 {0.071} (0.070)	0.036 {0.145} (0.137)	1.959 {6.494} (12.240)	-0.010
	GARCH-MIDAS (12)	-0.000	-0.003	-0.006	-0.029	-0.470	-0.009
	GARCH-MIDAS (36, $\tilde{X}$ )	0.000	-0.003	-0.006	0.000	0.788	-0.009
	GARCH-MIDAS (12, $\tilde{X}$ )	0.000	-0.002	-0.005	-0.075	-0.869	-0.008
	GARCH	0.000	0.003	0.009	—	—	0.001
Daily $\tau_t$	GARCH-MIDAS (264)	-0.000 {0.008} (0.008)	-0.003 {0.014} (0.014)	-0.003 {0.063} (0.062)	0.010 {0.078} (0.075)	1.030 {5.020} (4.786)	-0.006
	GARCH-MIDAS (66)	-0.000	-0.002	-0.001	-0.053	-3.247	-0.004
	GARCH-MIDAS (264, $\tilde{X}$ )	-0.000	-0.003	-0.002	0.002	0.332	-0.005
	GARCH-MIDAS (66, $\tilde{X}$ )	0.000	-0.002	0.000	-0.066	-3.414	-0.003
	GARCH	0.003	0.003	0.031	—	—	0.020
<b>Panel B: <math>Z_{n,i,t}</math> Student <math>t</math> distributed</b>							
Monthly $\tau_t$	GARCH-MIDAS (36)	-0.000 {0.008} (0.008)	-0.004 {0.014} (0.015)	-0.008 {0.075} (0.071)	0.040 {0.152} (0.141)	1.491 {5.983} (11.033)	0.108
	GARCH-MIDAS (12)	-0.000	-0.003	-0.006	-0.030	-0.589	0.109
	GARCH-MIDAS (36, $\tilde{X}$ )	-0.000	-0.003	-0.006	0.003	0.715	0.110
	GARCH-MIDAS (12, $\tilde{X}$ )	-0.000	-0.002	-0.004	-0.073	-0.797	0.111
	GARCH	-0.000	0.003	0.011	—	—	0.122
Daily $\tau_t$	GARCH-MIDAS (264)	-0.000 {0.008} (0.008)	-0.003 {0.014} (0.014)	-0.002 {0.065} (0.063)	0.012 {0.082} (0.075)	1.136 {5.896} (6.039)	0.112
	GARCH-MIDAS (66)	0.000	-0.002	0.000	-0.052	-2.730	0.114
	GARCH-MIDAS (264, $\tilde{X}$ )	0.000	-0.003	-0.001	0.003	0.341	0.114
	GARCH-MIDAS (66, $\tilde{X}$ )	0.000	-0.002	0.001	-0.064	-3.372	0.116
	GARCH	0.003	0.003	0.034	—	—	0.141

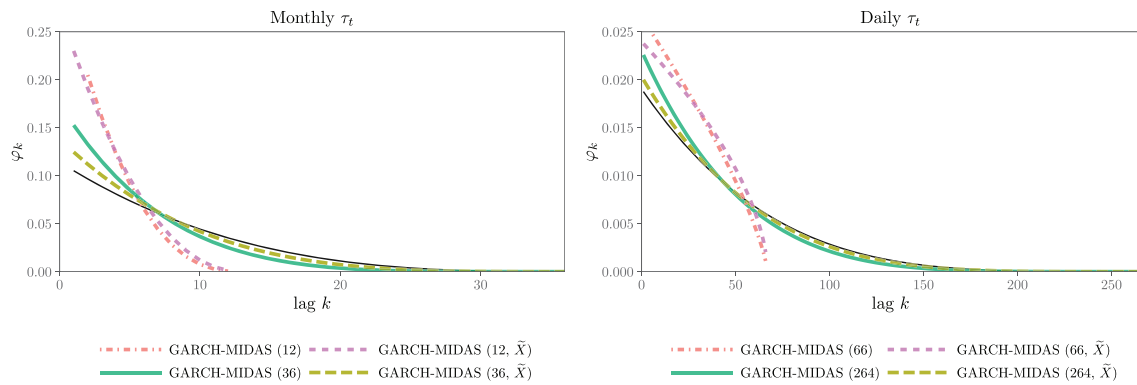
*Note.* The table reports the average bias of parameter estimates and the corresponding standard errors across 2,000 Monte Carlo simulations. We provide results for both daily and monthly long-term components. In curly brackets, empirical standard deviations of parameter estimates are reported. Entries in parentheses correspond to the square root of average Wang and Ghysels (2015) asymptotic variances. The parameter estimates are based on (the first) 20 years of observations (i.e. the in-sample period). In both long-term components (see Equations (7) and (9)), we choose  $\theta = 0.3$  and  $w_1 = 1$ . We use  $m = 0.1$  and  $w_2 = 4$  in the monthly  $\tau_t$  and  $m = -0.1$  and  $w_2 = 5$  in the daily  $\tau_t$ . The long-term component is assumed to depend on  $K = 36$  monthly or  $K = 264$  daily observations. The covariate  $X_t$  is modeled as an AR(1) process; that is,  $X_t = \phi X_{t-1} + \xi_t$ ,  $\xi_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\xi^2)$ , with  $\phi = 0.9$ ,  $\sigma_\xi^2 = 0.3^2$  for a monthly, and  $\phi = 0.98$ ,  $\sigma_\xi^2 = 0.2^2$  for a daily  $\tau_t$ . The parameters of the short-term component are in both cases given by  $\alpha = 0.06$ ,  $\beta = 0.91$  and  $\gamma = 0$ . For each model that is estimated based on the true value of  $X_t$ , we also incorporate estimations in which  $X_t$  is replaced by a noisy proxy  $\tilde{X}_t$ . It is modeled as  $\tilde{X}_t = X_t + \mathcal{N}(0, 0.2 + 0.8|X_t|)$  in the case of the monthly varying  $\tau_t$  and  $\tilde{X}_t = X_t + \mathcal{N}(0, 0.5 + 0.8|X_t|)$  in the case of a daily varying  $\tau_t$ . The column “ $\kappa - 3$ ” presents the mean excess kurtosis of the standardized residuals from each model.

the empirical standard deviation of the estimated parameters. In Table 1 these figures are presented in curly brackets. The numbers in parentheses are the average asymptotic standard errors based on the results in Wang and Ghysels (2015). A comparison of these numbers shows that the asymptotic standard errors are close to the empirical standard deviation of estimated parameters. The only exception is the specification with monthly  $\tau_t$  where the asymptotic standard errors of  $w_2$  appear to be too big. Nevertheless, the overall performance of the asymptotic standard errors is very satisfying. That is, the Wang and Ghysels (2015) asymptotic standard errors that were derived under the assumption that  $X_t = \sum_{j=0}^{J-1} \varepsilon_{t-j}^2$  are applicable more generally.

### 3.2.2 | Misspecified models: Bias

Next, we investigate the effect of model misspecification. First, we consider specifications with a smaller lag length than the true one.<sup>14</sup> Choosing a lag length that is too small ( $K = 12$  for monthly  $\tau_t$  or  $K = 66$  for daily  $\tau_t$ ) does not lead to a

<sup>14</sup>We do not report results for  $K$  being chosen too large as the Beta weighting scheme is flexible enough to downweight uninformative lags to almost zero.



**FIGURE 3** Weighting schemes implied by mean parameter estimates. Estimated Beta weighting schemes (see Equation (9)) as implied by the mean parameter estimates reported in Table 1. The green (solid) line corresponds to the case of a correctly specified model, whereas the red (dot-dashed) line corresponds to a model with  $K$  being too small. With the brown (long dashed) and purple (short dashed) line, the corresponding cases of a GARCH-MIDAS with measurement error are reported. The black line shows the true weighting scheme [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

bias in the parameter estimates—with the exception of  $w_2$ . Now the QMLE of  $w_2$  is downwardly biased. As the estimated weighting schemes in Figure 3 show, the downward bias in  $w_2$  translates into biased weighting schemes. Second, we consider the case of observing the explanatory variable  $X_t$  with measurement error. This is a reasonable scenario because in practice the true  $X_t$  is either unknown to or unobservable for the researcher, who will base his analysis on a reasonable proxy. We denote the proxy by  $\tilde{X}_t$  and specify it as  $X_t$  plus conditionally heteroskedastic noise. In the case of monthly  $\tau_t$  the noise is given by  $\mathcal{N}(0, 0.2 + 0.8|X_t|)$  and in the case of daily  $\tau_t$  by  $\mathcal{N}(0, 0.5 + 0.8|X_t|)$ . The average correlation between  $X_t$  and  $\tilde{X}_t$  is 68.79%/62.71% for monthly/daily  $\tau_t$ . As before, only the QML estimates of  $w_2$  appear to be biased when  $X_t$  is replaced with  $\tilde{X}_t$ . Last, we estimate a misspecified one-component GARCH model that is obtained when restricting  $\tau_t$  to be constant. Despite the omitted long-term component, the parameter estimates of  $\alpha$  and  $\beta$  are essentially unbiased.

Note that the numbers in panel B of Table 1 are very similar to those in panel A. When replacing the normally distributed innovations with Student  $t$  distributed innovations, the density in the maximum likelihood estimation is misspecified and the estimator is truly QMLE. Nevertheless, this change hardly affects our findings. The only notable difference can be seen in the last column of Table 1, which shows the average excess kurtosis of the fitted standardized residuals. Those residuals are given by  $\varepsilon_{i,t}/\sqrt{\hat{\tau}_t \hat{g}_{i,t}}$  for the GARCH-MIDAS models and by  $\varepsilon_{i,t}/\sqrt{\hat{g}_{i,t}}$  for the GARCH model. While the excess kurtosis is essentially zero in panel A, in panel B there is still excess kurtosis, reflecting the fact that the innovations are Student  $t$  distributed.

### 3.3 | Forecast evaluation

Next, we evaluate the forecast performance of the different specifications. Based on the in-sample parameter estimates, we construct OOS volatility forecasts for the remaining 20 years. Keeping the parameter estimates fixed is usually referred to as a “fixed (forecasting) scheme.”<sup>15</sup> The forecast performance of the different models will be evaluated over the 2,000 Monte Carlo replications.

We compare the forecast performance of the correctly specified GARCH-MIDAS with all the misspecified models presented in Table 1. In addition, we consider the two-state MS-GARCH-TVI model that was introduced in Section 2.2.<sup>16</sup>

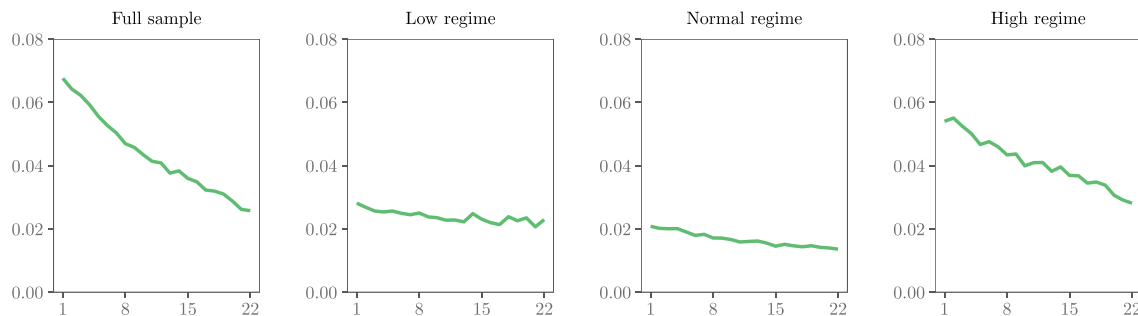
#### 3.3.1 | MZ regression

We first present the outcomes of MZ regressions. Figure 4 shows the  $R_k^2$  of MZ regressions for volatility forecasts,  $h_{k,t+1|t}$ , with  $k = 1, \dots, 22$  (i.e., for up to 1 month ahead). Forecast evaluation is based on the noisy proxy  $\varepsilon_{k,t+1}^2$ , whereby the data generating process is the GARCH-MIDAS with monthly  $\tau_t$  and normally distributed innovations. The forecasts are generated from the correctly specified GARCH-MIDAS model. We present the  $R_k^2$  for the full OOS period as well as for

<sup>15</sup>In contrast, in the empirical forecast evaluation in Section 4.4 we apply a “rolling scheme.” As we will discuss below, this is important because it takes into account the real-time nature of the data and allows for changes in the model parameters.

<sup>16</sup>In-sample parameter estimates for the MS-GARCH-TVI model can be found in the Supporting Information Appendix, Table B.1. The median estimates of  $\alpha$  and  $\beta$  are close to the true values. The estimates of  $\omega_1$  and  $\omega_2$  represent a low- and a high-volatility regime. As measured by  $\varpi = p_{1,1} + p_{2,2} - 1$ , the degree of persistence in the long-term component is very high.





**FIGURE 4** MZ  $R^2$ —monthly  $\tau_t$ —evaluation based on  $\varepsilon^2_{k,t+1}$ . The figure shows the average  $R^2_k$  of MZ regressions based on the predictions from the correctly specified GARCH-MIDAS model over all 2,000 Monte Carlo replications. The true volatility is proxied by  $\varepsilon^2_{k,t+1}$ . Besides the full out-of-sample period, we consider low-, normal-, and high-volatility regimes. For a definition of the regimes see Section 3.3.1 [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/ae.2742)]

three different volatility regimes: *low*, *normal*, and *high*. Volatility regimes are defined as follows. We consider the empirical distribution of daily realized variances during the OOS period. A forecast falls into the low/normal/high-volatility regime if the level of the realized variance on the day the forecast has been issued is below the 25% quantile, between the 25% and 75% quantile, or above the 75% quantile of the empirical distribution. In line with our theoretical result in Proposition 4, the  $R^2_k$ s for the full sample are decreasing with increasing forecast horizon. As expected,  $R^2_1$  is below the upper bound of one-third (see Equation (15)). Among the three regimes, we observe the highest  $R^2$ s in the high-volatility regime. Clearly, the high  $R^2$ s in the high-volatility regime do not reflect an improved absolute forecast performance but rather an improved relative forecast performance. Further, note that for almost all forecast horizons the  $R^2_k$ s in the full sample are higher than in each subsample.

For empirical applications, cumulative volatility forecasts are of greater importance than  $k$ -step-ahead forecasts. Hence in Figure 5 we present the  $R^2_{1:k}$  of MZ regressions for cumulative volatility forecasts,  $h_{1:k,t+1|t}$ , with  $k = 1, \dots, 22$ . Note that, by construction, the volatility forecasts are nonoverlapping. We now present forecasts from the correctly specified and the misspecified GARCH-MIDAS models as well as from the MS-GARCH-TVI and the nested GARCH. Forecast evaluation is based on the precise proxy  $RV_{1:k,t+1}$ . Panels (a)/(b) show the results for monthly/daily  $\tau_t$ . Based on Figure 5, we are able to rank the different models' forecast performance. While the performance of all GARCH-MIDAS models is essentially indistinguishable, the one-component GARCH and the MS-GARCH-TVI models lead to a lower  $R^2_{1:k}$ . Differences between models are most pronounced in the low and normal regime.

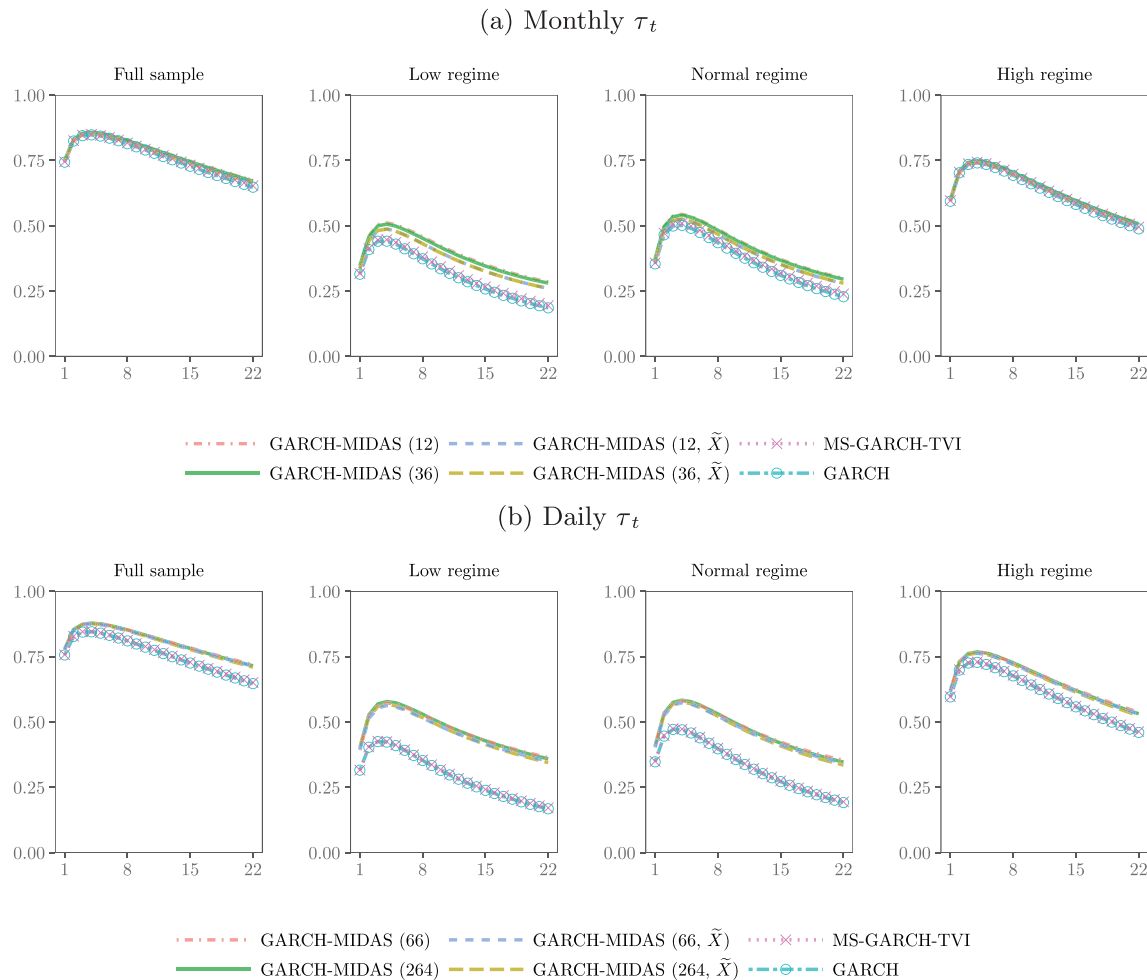
### 3.3.2 | Model confidence sets

Next, we formally test for superior predictive ability. We base our analysis on the MCS approach introduced by Hansen et al. (2011). Following the arguments in Patton (2011), we use the QLIKE loss as the evaluation criterion. For a  $k$ -step-ahead volatility forecast, the QLIKE is defined as

$$\text{QLIKE} \left( \sigma^2_{k,t+1}, h_{k,t+1|t} \right) = \sigma^2_{k,t+1} / h_{k,t+1|t} - \ln \left( \sigma^2_{k,t+1} / h_{k,t+1|t} \right) - 1. \quad (22)$$

The QLIKE is the only robust loss function that depends solely on the standardized forecast error,  $\sigma^2_{k,t+1} / h_{k,t+1|t}$ . As discussed in Patton (2011), the QLIKE is less sensitive with respect to extreme observations than the squared error loss. Further, it can be shown that the moment conditions required for Diebold and Mariano (1995) or Giacomini and White (2006) type tests are weaker under QLIKE than under squared error loss (see Patton, 2006).

We consider the following forecasting schemes. Based on the information available at the last day of the current month, cumulative volatility forecasts are computed for horizons of 1 day (1d), 2 weeks (2w), and 1 month (1m), as well as forecasts of volatility in 2 months (2m) and 3 months (3m). Whenever the forecast horizon is longer than the frequency of the long-term component, the optimal forecast requires predicting the long-term component. Instead, we simply fix the long-term component at its current level (see Section 2.4). Forecast evaluation is now based on the precise proxy  $RV_{1:k,t+1}$ . Next, we explain how the MCS is obtained. Denote by  $\mathcal{M}$  the set of all competing models. We define



**FIGURE 5** MZ  $R^2_{1:k}$ —monthly and daily  $\tau_t$ —evaluation based on  $RV_{1:k,t+1}$ : (a) monthly  $\tau_t$ ; (b) daily  $\tau_t$ . For each model the figure shows the average  $R^2_{1:k}$  of the MZ regressions over the 2,000 Monte Carlo replications. The true volatility is proxied by  $RV_{1:k,t+1}$ . The upper/lower panels display the case of monthly/daily long-term components. Besides the full out-of-sample period, we consider low-, normal-, and high-volatility regimes. For a definition of the regimes see Section 3.3.1 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

$$d_{i,j}(s, k) = \text{QLIKE} \left( RV_{1:k,t+s}, \hat{h}_{1:k,t+s|t}^{(i)} \right) - \text{QLIKE} \left( RV_{1:k,t+s}, \hat{h}_{1:k,t+s|t}^{(j)} \right)$$

as the difference in the QLIKE loss of models  $i$  and  $j$ . For example, when  $s = 1$  and  $k \in \{1, 5, 22\}$  the forecast  $\hat{h}_{1:k,t+s|t}^{(i)}$  denotes the cumulative forecast for the first (1d), the first 5 (1w), or all 22 (1m) days in the following month while for  $s \in 2, 3$  and  $k = 22$  we obtain the forecast for 2 (2m) and 3 (3m) months in the future. We compute the average loss difference,  $\bar{d}_{i,j}$ , and calculate the test statistic:

$$t_{ij} = \bar{d}_{i,j} / \sqrt{\hat{\text{var}}(\bar{d}_{i,j})} \quad \text{for all } i, j \in \mathcal{M}. \quad (23)$$

The MCS test statistic is then given by  $T_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} |t_{i,j}|$  and has the null hypothesis that all models have the same expected loss. Under the alternative, there is some model  $i$  that has an expected loss greater than the expected loss of all other models  $j \in \mathcal{M} \setminus i$ . If the null hypothesis is rejected, the worst-performing model is eliminated. The test is performed iteratively, until no further model can be eliminated. We denote the final set of surviving models by  $\mathcal{M}_{\text{MCS}}$ . This final set contains the best forecasting model with confidence level  $1 - \nu$ . We set  $\nu = 0.1$ . This choice is common practice in the literature. See, for example, Laurent, Rombouts, and Violante (2013) and Liu, Patton, and Sheppard (2015).

TABLE 2 Model confidence set inclusion rates

		1d	2w	1m	2m	3m
<b>Panel A: <math>Z_{n,i,t}</math> normally distributed</b>						
Monthly $\tau_t$	GARCH-MIDAS (36)	0.850	0.758	0.770	0.795	0.792
	GARCH-MIDAS (12)	0.852	0.745	0.762	0.818	0.827
	GARCH-MIDAS (36, $\tilde{X}$ )	0.723	0.559	0.589	0.650	0.661
	GARCH-MIDAS (12, $\tilde{X}$ )	0.696	0.539	0.560	0.648	0.684
	MS-GARCH-TVI	0.765	0.560	0.603	0.664	0.673
	GARCH	0.477	0.221	0.216	0.260	0.310
Daily $\tau_t$	GARCH-MIDAS (264)	0.946	0.893	0.861	0.784	0.743
	GARCH-MIDAS (66)	0.850	0.796	0.836	0.890	0.878
	GARCH-MIDAS (264, $\tilde{X}$ )	0.843	0.672	0.646	0.663	0.688
	GARCH-MIDAS (66, $\tilde{X}$ )	0.763	0.614	0.664	0.778	0.831
	MS-GARCH-TVI	0.376	0.100	0.138	0.467	0.765
	GARCH	0.257	0.043	0.050	0.244	0.493
<b>Panel B: <math>Z_{n,i,t}</math> Student <math>t</math> distributed</b>						
Monthly $\tau_t$	GARCH-MIDAS (36)	0.912	0.790	0.772	0.761	0.764
	GARCH-MIDAS (12)	0.922	0.808	0.785	0.812	0.818
	GARCH-MIDAS (36, $\tilde{X}$ )	0.842	0.656	0.640	0.652	0.650
	GARCH-MIDAS (12, $\tilde{X}$ )	0.841	0.636	0.622	0.668	0.683
	MS-GARCH-TVI	0.875	0.666	0.654	0.675	0.664
	GARCH	0.734	0.331	0.267	0.280	0.309
Daily $\tau_t$	GARCH-MIDAS (264)	0.968	0.912	0.866	0.792	0.742
	GARCH-MIDAS (66)	0.918	0.839	0.862	0.885	0.854
	GARCH-MIDAS (264, $\tilde{X}$ )	0.927	0.769	0.712	0.694	0.685
	GARCH-MIDAS (66, $\tilde{X}$ )	0.877	0.726	0.731	0.812	0.822
	MS-GARCH-TVI	0.690	0.222	0.206	0.501	0.758
	GARCH	0.602	0.112	0.093	0.276	0.485

Note. The numbers are the empirical frequencies of a model being included in the 90% model confidence set at different forecast horizons: 1 day (1d), 2 weeks (2w), 1 month (1m), 2 months (2m), and 3 months (3m). Panel A corresponds to the simulation with normally distributed intraday returns and Panel B to standardized Student  $t$  distributed intraday returns with five degrees of freedom. The averages are taken across 2,000 Monte Carlo replications.

Since the asymptotic distribution of the test statistic  $T_{\mathcal{M}}$  is nonstandard, we approximate it by block-bootstrapping as proposed by Hansen et al. (2011), where the block length is determined by fitting an  $AR(p)$  process to the series of loss differences. In our analysis, 8,000 bootstrap replications at each stage were sufficient in order to obtain stable results.<sup>17</sup>

Table 2 reports how often a certain model is included in the MCS across the 2,000 replications. Panel A provides results for normally distributed innovations and panel B for Student  $t$  distributed innovations. For example, for normally distributed innovations, monthly  $\tau_t$ , and a forecast horizon of 1 day, the correctly specified GARCH-MIDAS (36) is included in the MCS in 85% of the replications. The table clearly shows that the misspecified one-component GARCH model is included less often in the MCS than the GARCH-MIDAS models. In particular, this is the case for daily  $\tau_t$ . Further, for daily  $\tau_t$  and forecast horizons of up to 2 months the MS-GARCH-TVI is less often part of the MCS than all GARCH-MIDAS models. Additionally, among the GARCH-MIDAS models the correctly specified one has the highest inclusion rates in the MCS when the forecast horizon is up to 1 month. At least for monthly  $\tau_t$ , it appears that a misspecification of the lag length is less severe than observing the explanatory variable with measurement error. Finally, at the longest forecast horizon (3m) all forecasts suffer from a misspecified forecast of the long-term component and hence it becomes increasingly difficult to distinguish between models.

In summary, independently of whether the long-term component is specified at a daily or monthly frequency, the correctly specified GARCH-MIDAS model as well as the GARCH-MIDAS with misspecified lag length clearly outperform the one-component GARCH as well as the MS-GARCH-TVI in terms of forecast performance. For models with daily long-term components this result also holds when the explanatory variable is observed with measurement error. Only for monthly long-term components and measurement error in  $X_t$ , we find that the MS-GARCH-TVI performs slightly better.

<sup>17</sup>For implementing the MCS procedure, we use the R package *rugarch* (Ghalanos, 2018), which includes the implementation used in the MFE Matlab Toolbox by Kevin Sheppard. See [https://www.kevin-sheppard.com/MFE\\_Toolbox](https://www.kevin-sheppard.com/MFE_Toolbox).

*Remark 2.* As discussed in Section 2.1, Assumption 3 is likely to hold for explanatory variables that are observed at a lower frequency than the daily returns. For certain daily explanatory variables (e.g., the VIX index) Assumption 3 might be violated. However, under reasonable assumptions the correlation between the innovations to the daily returns and  $X_t$  itself can be expected to be small. The correlation with future  $\tau_t$  will be even smaller. For a more detailed discussion see Supporting Information Appendix D, which also provides additional simulations. The simulations show that even if Assumption 3 is mildly violated all the previous findings still hold.

## 4 | EMPIRICAL ANALYSIS

Last, we turn to an empirical application of the GARCH-MIDAS models to S&P 500 return data. In Section 4.1 we introduce our data set. Full sample estimation results for various GARCH-MIDAS models are reported in Section 4.2. Thereafter, in Section 4.3 we explain how real-time volatility forecasts can be constructed when taking into account the release schedule of macroeconomic variables. The forecast comparison is carried out in Section 4.4, where we evaluate the GARCH-MIDAS volatility forecasts against forecasts from eight competitor models.

### 4.1 | Data

#### 4.1.1 | Stock market data

We consider daily log-returns on the S&P 500, calculated as  $r_{i,t} = 100 \cdot (\ln(p_{i,t}) - \ln(p_{i-1,t}))$ , for the 1971:M1 to 2018:M4 period. For evaluating the volatility forecasts, we employ daily realized variances,  $RV_{i,t}$ , defined as the sum of the squared 5-minute intraday log-returns on day  $t$  plus the squared overnight log-return. The latter is defined as the log of the open price on day  $t$  minus the log of the close price on day  $t - 1$ . This approach follows Bollerslev, Hood, Huss, and Pedersen (2018), among others. The data for constructing  $RV_{i,t}$  were obtained from the Realized Library of the Oxford-Man Institute of Quantitative Finance and are available from the year 2000 onwards (see Heber, Lunde, Shephard, & Sheppard, 2009).

#### 4.1.2 | Explanatory variables

As explanatory variables we use daily measures of financial risk, a weekly measure of financial conditions and monthly macroeconomic variables. We employ backward- and forward-looking measures of daily volatility. The former is proxied by a rolling window of the average realized volatility (based on squared daily returns) over the previous 22 days,  $RVol(22)_{i,t} = \sqrt{1/22 \sum_{j=0}^{21} r_{i-j,t}^2}$ , and the latter by the VIX index (converted to a daily level by dividing it by  $\sqrt{252}$ ). In addition, we consider the difference between the VIX (divided by  $\sqrt{252}$ ) and  $RVol(22)$  as a proxy for the (square root of the) variance risk premium (VRP).<sup>18</sup>

We use the weekly National Financial Conditions Index (NFCI) as a measure for the tightness of financial conditions in the USA. The NFCI is a weighted average of 105 standardized financial indicators of risk, credit and leverage derived by dynamic factor analysis. Monthly macroeconomic conditions are measured by the Chicago Fed National Activity Index (NAI) and growth rates of industrial production (IP) and housing starts, both calculated as  $\Delta X_t = 100 \cdot (\ln(X_t) - \ln(X_{t-1}))$ . While the macroeconomic variables are included from 1971 onwards, the NFCI series begins in 1973 and the VIX is available from 1990 onwards.<sup>19</sup>

Before we estimate GARCH-MIDAS models, we employ the Conrad and Schienle (2018) Lagrange multiplier (LM) test for an omitted multiplicative component in one-component GARCH models. This test checks whether a simple GJR-GARCH(1, 1) is misspecified in the sense of neglecting a second component that is driven by an explanatory variable  $X$ . Since the test is of the LM type, it requires estimation of the model under the null hypothesis only. Assuming that under the alternative there is a second component which is driven by  $K$  lags of the variable  $X$ , the test statistic can be shown to be  $\chi^2$  with  $K$  degrees of freedom. An appealing property of the test is that it can be applied in settings where  $X$  is observed at the same frequency as the returns but also when  $X$  is observed at a lower frequency. Intuitively, the test checks whether

<sup>18</sup>Note that the conventional definition of the variance risk premium is the squared VIX minus realized variance. We are interested in expressing the quantity in volatility units. Because the realized VRP takes positive as well as negative values, we take the square root of both quantities before we take the difference.

<sup>19</sup>Table B.2 in the Supporting Information Appendix provides summary statistics for the stock returns and the seven explanatory variables. Figure C.2 in the Supporting Information Appendix shows the evolution of the corresponding time series. Further details on the data set are provided in Supporting Information Appendix F.

TABLE 3 LM test for misspecification of GJR-GARCH(1, 1)

$X_t$	VIX	RVol	NFCI	NAI	$\Delta$ IP	$\Delta$ Hous
$K = 1$	76.28 [<0.01]	14.38 [<0.01]	22.54 [<0.01]	15.25 [<0.01]	7.99 [<0.01]	0.18 [0.67]
$K = 2$	84.05 [<0.01]	19.03 [<0.01]	24.05 [<0.01]	17.34 [<0.01]	10.22 [<0.01]	0.18 [0.91]

Note. The table reports the test statistics and corresponding  $p$ -values of the Conrad and Schienle (2018) misspecification test for one-component GJR-GARCH(1, 1) models. The test is implemented using either one ( $K = 1$ ) or two ( $K = 2$ ) lags of the explanatory variable  $X_t$ . For VIX and RVol(22) the test is based on daily data from 1990 onwards; for the NFCI, NAI,  $\Delta$  IP, and  $\Delta$  Housing starts the test is based on weekly/monthly data from 1974 onwards.

the squared standardized residuals from the GJR-GARCH are predictable using (functions of) past values of  $X$ . Table 3 shows the outcome of the test when applied to each of our explanatory variables. When choosing either  $K = 1$  or  $K = 2$ , the test clearly rejects the null hypothesis that a GJR-GARCH is correctly specified for all variables except housing starts. Thus the LM test results suggest using GARCH-MIDAS models instead. The estimates for a GARCH-MIDAS model based on housing starts in Section 4.2.1 will show that housing starts are a leading indicator with respect to financial volatility. This implies that the choice of  $K = 1$  or  $K = 2$  is too small. When redoing the LM test for a lag length of up to  $K = 12$  the LM test indeed rejects the null hypothesis also for housing starts.

We can also apply the LM test jointly to several variables at the same time. However, all variables need to be observed at the same frequency. When including the NAI, industrial production and housing starts and selecting an appropriate lag length, the NAI and housing starts are individually significant, whereas industrial production is not. This suggests that among the macroeconomic variables the NAI and housing starts are most informative. We also aggregated the VIX and the NFCI to a monthly frequency and performed the LM test jointly for all variables. While the overall LM statistic is highly significant, the VIX, the NFCI and housing starts are the only variables that are individually significant.

## 4.2 | Full-sample parameter estimates

### 4.2.1 | One explanatory variable

We first estimate a GARCH-MIDAS model for each explanatory variable for the full sample. We include a constant in the mean equation; that is, returns are modeled as  $r_{i,t} = \mu + \varepsilon_{i,t}$ . After visual inspection of the estimated weighting schemes for alternative choices of  $K$ , we select a lag length that is rather too large than too small. As discussed in Section 3, the data will identify the optimal weighting scheme as long as  $K$  is chosen large enough. We choose  $K = 264$  for RVol(22),  $K = 3$  for the VIX/VRP and  $K = 52$  for the NFCI.<sup>20</sup> Thus, for the forward-looking VIX/VRP, only the most recent information appears to drive long-term volatility, while the backward-looking RVol(22) is smoothed over many lags. As in Conrad and Loch (2015), we choose  $K = 36$  for the monthly macroeconomic variables. The estimates for the parameters in the conditional variance are reported in Table 4. For all variables except housing starts, we find that a restricted Beta weighting scheme with  $w_1 = 1$  is the best choice; that is, the optimal weights are declining from the beginning. For housing starts, an unrestricted scheme that allows for “hump-shaped” weights is required. This confirms the finding in Conrad and Loch (2015) that housing starts are leading with respect to long-term volatility.<sup>21</sup> Note that the GARCH-MIDAS models based on the NFCI and the three macroeconomic variables employ return data for the 1974:M1 to 2018:M4 period, while the models with daily  $\tau_t$  employ data from 1990:M1 onwards. Hence models based on daily  $\tau_t$  cannot be compared to models based on weekly/monthly  $\tau_t$  in terms of log-likelihood or Bayesian information criterion (BIC).

Concerning the parameter estimates, it is interesting to observe that the GARCH-MIDAS models with daily  $\tau_t$  lead to lower estimates of  $\beta$  than models with weekly or monthly  $\tau_t$ . While for the models with daily  $\tau_t$  the estimates of  $\alpha$  are close to zero, there is strong evidence for asymmetry (as indicated by the highly significant  $\gamma$  parameter). These parameter estimates imply that the deviations of the short-term component from the long-term component are more short lived for GARCH-MIDAS models with daily  $\tau_t$ .<sup>22</sup> The signs of the estimated  $\theta$ s for realized volatility, the VIX, and the macroeconomic variables are in line with findings in the previous literature. Higher levels of financial volatility tend to increase

<sup>20</sup>For all variables, Figure C.3 in the Supporting Information Appendix shows the estimated weighting schemes for selected choices of  $K$ . The figure illustrates that the estimated weighting schemes no longer change once the selected lag length is sufficiently large. In all cases, our choice of the lag length is rather conservative.

<sup>21</sup>Figure C.4 in the Supporting Information Appendix shows the estimated weighting schemes.

<sup>22</sup>This behavior is also evident from Figure C.5 in the Supporting Information Appendix, which shows the evolution of the annualized long-term components and the conditional volatilities.



**TABLE 4** Full-sample estimation results: GARCH-MIDAS with one explanatory variable

	$\alpha$	$\beta$	$\gamma$	$m$	$\theta$	$w_1$	$w_2$	$K$	LLH	BIC	VR(X)
<b>Daily <math>\tau_t</math></b>											
RVol(22)	0.000 (0.008)	0.843*** (0.012)	0.192*** (0.015)	-1.261*** (0.112)	1.177*** (0.096)	1	3.049*** (0.675)	264	-9,201	18,465	42.78
VIX	0.000 (0.010)	0.853*** (0.021)	0.095*** (0.015)	-2.129*** (0.086)	1.524*** (0.067)	1	3.470** (1.371)	3	-9,138	18,339	76.14
VRP	0.017** (0.007)	0.902*** (0.007)	0.128*** (0.011)	-0.384*** (0.137)	1.084*** (0.096)	1	5.571** (2.591)	3	-9,174	18,410	10.92
<b>Weekly <math>\tau_t</math></b>											
NFCI	0.017*** (0.006)	0.902*** (0.005)	0.115*** (0.007)	-0.101 (0.073)	0.252*** (0.048)	1	2.892 (2.314)	52	-15,103	30,271	11.42
<b>Monthly <math>\tau_t</math></b>											
NAI	0.019*** (0.006)	0.900*** (0.005)	0.116*** (0.007)	-0.058 (0.079)	-0.359*** (0.073)	1	9.066*** (3.312)	36	-14,569	29,202	14.14
$\Delta$ IP	0.019*** (0.006)	0.903*** (0.005)	0.113*** (0.007)	0.074 (0.089)	-0.650*** (0.161)	1	5.271*** (1.782)	36	-14,573	29,211	10.63
$\Delta$ Housing	0.019*** (0.005)	0.897*** (0.005)	0.119*** (0.007)	-0.079 (0.076)	-0.237*** (0.034)	1.695*** (0.383)	2.586*** (0.770)	36	-14,559	29,192	19.63
GARCH	0.021*** (0.005)	0.911*** (0.005)	0.103*** (0.007)	-0.073 (0.098)	—	—	—	—	-15,355	30,757	—

Note. Estimation results for GARCH-MIDAS models are reported for seven explanatory variables. Estimation using the NFCI, NAI, IP, and housing starts begins in 1974:M1 based on low-frequency observations reaching as far as 1971:M1 in line with the lag length  $K$ . Estimation of the GARCH-MIDAS models using RVol(22) and VIX as an explanatory variable employs daily return data starting in 1990:M1. For all explanatory variables except housing starts a restricted weighting scheme is chosen ( $w_1 = 1$ ). Bollerslev–Wooldridge standard errors are reported in parentheses, where asterisks indicate significance at the \*\*\*1%, \*\*5%, and \*10% level. LLF is the value of the maximized log-likelihood function and BIC is the Bayesian information criterion. The variance ratio  $VR(X) = \text{var}(\log(\tau_M^X)) / \text{var}(\log(\sigma_M^X))$  is calculated on monthly aggregates. Estimates for  $\mu$  are omitted.

long-term volatility, whereas an improvement in macroeconomic conditions decreases long-term volatility. The finding that a higher variance risk premium and tighter financial conditions (i.e., an increase in the NFCI) predict higher volatility is new. While the positive relation between realized/expected measures of volatility and long-term volatility might be viewed as “mechanical,” the NFCI as well as the macroeconomic variables can be considered fundamental drivers of financial volatility.

We gauge the importance of the variation in the long-term component for the overall expected variation in return volatility by the variance ratio introduced in Equation (8). To facilitate comparison across models, we focus on the monthly variation of volatility. That is, for all models we denote the monthly aggregate volatility by  $\sigma_M^X$ . For models with monthly long-term components, we have that  $\tau_M^X = \tau_t^X$ . For models with daily or weekly long-term components,  $\tau_M^X$  refers to monthly aggregates of the daily/weekly long-term component. We then calculate  $VR(X) = \text{var}(\log(\tau_M^X)) / \text{var}(\log(\sigma_M^X))$ , where  $X$  indicates that the variance ratio is based on a specific explanatory variable. As Table 4 shows, the models with daily  $\tau_t$  achieve much higher variance ratios than the models with a weekly/monthly long-term component. Among the models with daily long-term components, the variance ratio of 76.14% for the VIX-based model is by far the highest and implies that three quarters of the expected variation in return volatility can be traced back to variation in the VIX. In Section 4.4 we will investigate whether a high variance ratio necessarily translates into good OOS predictive performance.

#### 4.2.2 | Two explanatory variables

The GARCH-MIDAS setting allows us to include two or more explanatory variables in the long-term component. Based on the results in the previous section, the VIX appears to be better suited to capture daily movements in the long-term component than RVol(22) or the VRP. Since the NFCI and, in particular, the macroeconomic variables capture lower frequency movements, it is natural to estimate GARCH-MIDAS models with the VIX and one of those variables jointly in the long-term component. This allows us to formally check whether the NFCI and the three macroeconomic variables contain information that is complementary to the VIX. The long-term component for those models is given by

$$\log \tau_{i,t} = m + \theta^{VIX} \sum_{l=1}^{K^{VIX}} \varphi_l(1, w_2^{VIX}) VIX_{i-l,t} + \theta^X \sum_{l=1}^{K^X} \varphi_l(w_1^X, w_2^X) X_{i-l,t}. \quad (24)$$

**TABLE 5** Full-sample estimation results: VIX combined with second explanatory variable

	$\alpha$	$\beta$	$\gamma$	$m$	$\theta^X$	$w_1^X$	$w_2^X$	$\theta^{VIX}$	$w_2^{VIX}$	$K^X$	LLH	BIC	VR(VIX, X)
<b>Daily <math>\tau_t</math></b>													
VIX	0.000 (0.010)	0.853*** (0.021)	0.095*** (0.015)	-2.129*** (0.086)	—	—	—	1.524*** (0.067)	3.470** (1.371)	3	-9,138	18,339	76.14
<b>Weekly <math>\tau_t</math></b>													
NFCI	0.000 (0.010)	0.852*** (0.020)	0.099*** (0.016)	-1.993*** (0.143)	0.118 (0.085)	1	2.252 (4.152)	1.451*** (0.093)	3.617** (1.518)	52	-9,110	18,300	75.84
<b>Monthly <math>\tau_t</math></b>													
NAI	0.000 (0.009)	0.870*** (0.018)	0.092*** (0.015)	-2.032*** (0.100)	-0.108** (0.046)	1	119.372 (326.330)	1.431*** (0.079)	3.775** (1.594)	36	-9,133	18,346	75.06
$\Delta$ IP	0.000 (0.009)	0.876*** (0.018)	0.084*** (0.014)	-2.133*** (0.096)	-0.043 (0.089)	1	8.960 (34.803)	1.528*** (0.072)	3.806** (1.520)	36	-9,139	18,357	75.91
$\Delta$ Housing	0.000 (0.009)	0.863*** (0.019)	0.097*** (0.015)	-2.035*** (0.094)	-0.061** (0.024)	1.001 (0.743)	2.139 (2.462)	1.446*** (0.074)	3.605** (1.503)	36	-9,135	18,359	74.99

Note. Estimation results for GARCH-MIDAS models are reported, in which the daily VIX is combined with the low-frequency variables reported in Table 4—that is, the NFCI, NAI, and changes in industrial production and housing starts. The estimates are based on daily return data from 1990:M1 to 2018:M4. For comparison, the estimation results using only the VIX as a covariate from Table 4 are included in the first row. All parameters with a superscript  $X$  relate to the second explanatory variable.  $K^{VIX}$  is always equal to 3. Bollerslev–Wooldridge standard errors are reported in parentheses, where asterisks indicate significance at the \*\*\*1%, \*\*5%, and \*10% level. LLF is the value of the maximized log-likelihood function and BIC is the Bayesian information criterion. The variance ratio  $VR(VIX, X) = \text{var}(\log(\tau_M^{VIX, X})) / \text{var}(\log(\sigma_M^{VIX, X}))$  is calculated on monthly aggregates. Estimates for  $\mu$  are omitted.

Estimation results are presented in Table 5. Note that  $K^{VIX}$  and  $K^X$  are chosen as in Table 4. For all models the estimation period is now determined by the availability of the VIX. When controlling for the VIX, the  $\theta^X$  parameter turns out to be significant for the NAI and housing starts. Thus macroeconomic variables appear to contain information that is complementary to the one included in the VIX. However, none of the models that include two variables achieves a higher VR than the model based on the VIX alone.

### 4.2.3 | More than two explanatory variables

As an extension to Section 4.2.2, one could employ more than two covariates. We experimented with combining three variables in the long-term component but found no further improvements in terms of model fit. Moreover, GARCH-MIDAS models including more than two variables in the long-term component are difficult to estimate because the likelihood is relatively insensitive with respect to changes in the weighting parameters. Instead, in Section 4.4 on OOS forecasting, we will aggregate the information in the different variables by simply calculating the average forecast across all GARCH-MIDAS models with one explanatory variable.

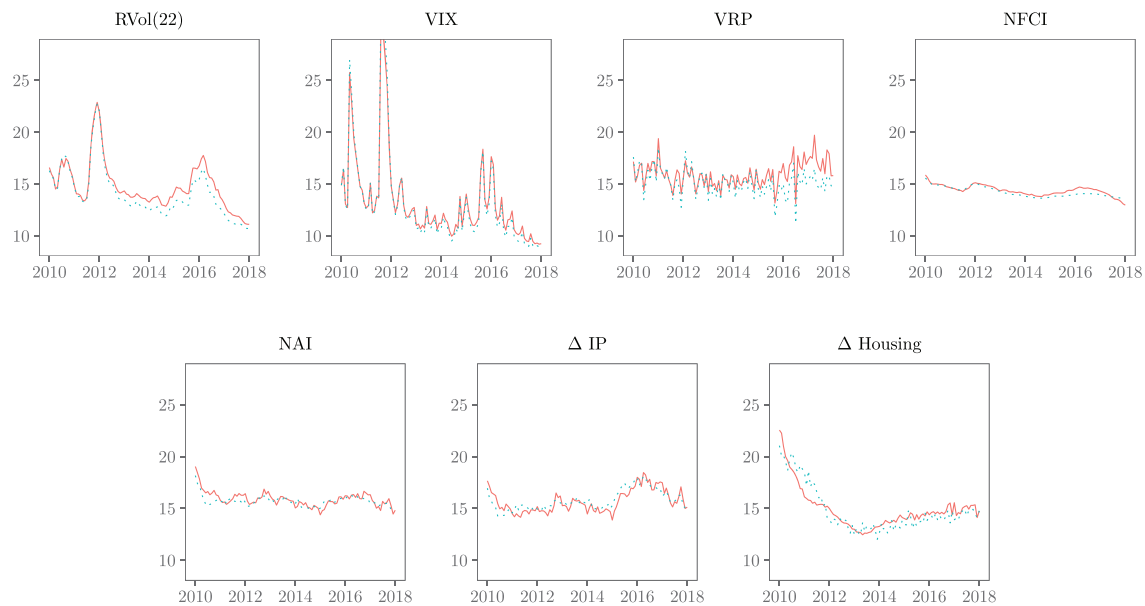
## 4.3 | Real-time estimates

In the following, we make use of vintage data. This allows for a realistic evaluation of the GARCH-MIDAS models' ability to describe the behavior of long-term financial volatility in real time.<sup>23</sup> In order to compare full-sample estimates of the long-term component with corresponding real-time estimates, we reestimate all GARCH-MIDAS models from Table 4 on a daily basis. Estimation is performed on a rolling window. For each explanatory variable, the window size is determined by the length of the first estimation period ending in 2009:M12. The period 2010:M1 to 2018:M4 will be used as the OOS period for the forecast evaluation in Section 4.4. In order to ensure that our estimates of the long-term component are feasible in real time, we employ vintage data that are available for the NFCI, the NAI, IP, and housing starts from the ALFRED database hosted by the St. Louis Fed.<sup>24</sup> When using real-time data, the long-term component no longer changes its value at the beginning of a week/month but whenever a new data release becomes available.

Figure 6 shows the estimated long-term components based on the full-sample estimates (as reported in Table 4, dotted lines) and based on the rolling window real-time estimates (solid lines). For RVol(22), VIX, and VRP the long-term component estimates in the full sample might differ from the rolling window estimates, because they are based on distinct sample periods (rolling window vs. full sample). For the NFCI, NAI, IP, and housing starts, the two long-term components

<sup>23</sup>To the best of our knowledge, Lindblad (2017) appears to be the only other paper that makes use of real-time data when estimating GARCH-MIDAS models.

<sup>24</sup>For more details on real-time data availability see Supporting Information Appendix F.



**FIGURE 6** Comparison of rolling window and full-sample long-term components. For each explanatory variable, the monthly averaged long-term volatility components,  $\sqrt{\tau_t}$ , are depicted for the period 2010:M1 to the end of 2018:M1, the last month of issuing forecasts and hence real-time estimation. The long-term component obtained from the full-sample estimates is given in green (dotted). Real-time estimates of the most recently fitted  $\sqrt{\tau_t}$  are depicted in red (solid). Volatilities are presented on an annualized scale [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

are not only based on distinct sample periods but also on different data vintages (real-time vs. final). Figure 6 shows that for RVol(22), VIX, and VRP the rolling window estimate of the long-term component is often somewhat higher than the full-sample estimate. For the macroeconomic variables the real-time estimates of the long-term component are occasionally below or above the full-sample estimates. However, the average absolute differences are quite sizable. For example, the average absolute difference between the full-sample and real-time estimates based on industrial production is 6.80%. To put this into context, for industrial production the mean absolute revision from the initial release to the latest available data was 2.18% during the 1965:Q3 to 2006:Q4 period (see Croushore, 2011). Among the variables considered in Croushore (2011), this is the highest value (even higher than for GDP). Similar numbers in terms of changes in the long-term component are obtained for the other variables: 9.35% for housing starts, 4.78% for the NAI, and 2.68% for the NFCI. In summary, these figures highlight the importance of using real-time instead of final data releases for the macroeconomic variables for a realistic forecast evaluation.

#### 4.4 | Forecast evaluation

Finally, we evaluate the predictive performance of the GARCH-MIDAS models in the 2010:M1 to 2018:M4 OOS period. As before, we consider cumulative volatility forecasts for horizons up to 3 months. When computing the forecasts, we keep the long-term component fixed at its current level. Volatility forecasts are based on the real-time rolling window parameter estimates as obtained in Section 4.3 (i.e., we apply a “rolling (forecasting) scheme”).

##### 4.4.1 | Competitor models

For forecast comparison, we use an extensive range of competitor models which are either extensions of the simple GARCH specification or which model the realized variance directly.

First, we consider the simple one-component GARCH(1, 1) model and a no-change (or random-walk) forecast, which simply scales the realized variance on the last day of period  $t$  to the appropriate horizon:  $h_{1:k,t+s|t} = k \cdot RV_{n,t}$ . Second, we use the MS-GARCH-TVI model that we employed in Section 3.3. The only difference is that we now use a GJR-GARCH specification in both regimes. In addition, we use an MS-GARCH model that consists of two GARCH equations with individual intercepts and individual ARCH and GARCH parameters. We incorporate asymmetric effects in the low-volatility

**TABLE 6** QLIKE losses and model confidence sets:  
full out-of-sample period

	1d	2w	1m	2m	3m
<b>GARCH-MIDAS</b>					
RVol(22)	0.306	0.246	0.271	0.387	0.428
VIX	0.275	0.215	0.240	0.359	0.414
VRP	0.291	0.227	0.260	0.384	0.430
NFCI	0.324	0.248	0.264	0.363	0.393
NAI	0.343	0.266	0.283	0.391	0.424
$\Delta$ IP	0.345	0.267	0.285	0.395	0.438
$\Delta$ Housing	0.328	0.252	0.264	<b>0.347</b>	<b>0.380</b>
VIX and NFCI	0.274	0.213	0.236	0.349	0.399
VIX and NAI	0.275	0.215	0.241	0.358	0.409
VIX and $\Delta$ IP	0.274	0.214	0.239	0.355	0.409
VIX and $\Delta$ Housing	0.275	0.218	0.243	0.351	0.405
Avg.	0.317	0.246	0.264	0.364	0.400
<b>Competitor models</b>					
GARCH	0.342	0.263	0.282	0.395	0.434
MS-GARCH-TVI	0.362	0.292	0.315	0.426	0.488
MS-GARCH-TVC	0.355	0.271	0.283	0.387	0.421
Real GARCH	0.260	<b>0.206</b>	<b>0.233</b>	0.356	0.390
HEAVY	0.277	0.238	0.299	0.539	0.662
HAR	0.254	0.210	0.243	0.368	0.419
HAR (lev.)	<b>0.238</b>	0.207	0.245	0.371	0.419
No-change	0.358	0.498	0.636	1.157	1.292

*Note.* Numbers reported are the average out-of-sample QLIKE losses for each model for 1-day- (1d), 2-week- (2w), 1-month- (1m), 2-month- (2m), and 3-month-ahead (3m) variance forecasts. Bold entries indicate the model with the lowest average QLIKE loss per horizon. Shaded entries indicate that the respective model is included in the 90% model confidence set. The average forecast (avg.) is the mean forecast across all GARCH-MIDAS models employing one explanatory variable. The out-of-sample evaluation period spreads from 2010:M1 to 2018:M4.

regime only.<sup>25</sup> We refer to this model as MS-GARCH with time-varying coefficients (MS-GARCH-TVC). Further, we use the HEAVY model by Shephard and Sheppard (2010) and the realized GARCH model by Hansen et al. (2012). The specifications of the HEAVY and realized GARCH models employ a measure of pure intraday realized variance,  $RV_{i,t}^{int}$  (defined as the sum of squared intraday returns). Third, we consider two specifications that directly model the realized variance,  $RV_{i,t}$ , (including squared overnight returns) and allow us to compute direct (as compared to iterated) volatility forecasts. We employ the HAR model of Corsi (2009) and the HAR model with leverage effect proposed in Corsi and Reno (2012).

For more details on the specification of the competitor models, their estimation, and volatility forecasting see Supporting Information Appendix G.<sup>26</sup> For the OOS forecast evaluation all competitor models are reestimated on a rolling window basis.

#### 4.4.2 | Forecast error statistics and model confidence set

As in Section 3.3.2, we base the comparison of the forecast performance of the different models on the QLIKE loss. Table 6 reports the average QLIKE loss for each model and forecast horizons of 1 day (1d), 2 weeks (2w), 1 month (1m), 2 months (2m), and 3 months (3m). We use the MCS approach to test whether there are one or several models that significantly outperform the others. As in Section 3.3.2, we rely on 90% model confidence sets.<sup>27</sup>

*MCS for full OOS period.* Shaded areas in Table 6 indicate that for the corresponding forecast horizon the respective model is included in the final set,  $\mathcal{M}_{MCS}$ . For example, for a forecast horizon of 1 day the only model that is included in

<sup>25</sup>Initially, we estimated a GJR-GARCH specification in both regimes. However, it turned out that the asymmetry term was only significant in the component that represents the low-volatility regime. In addition, we select this specification because it is much more stable in the rolling window estimation than the one with two GJR-GARCH regimes.

<sup>26</sup>Table B.3 in the Supporting Information Appendix shows the full-sample parameter estimates for the competitor models.

<sup>27</sup>As a robustness check, we present the corresponding results for a 95% MCS in Supporting Information Appendix H. Essentially all findings remain unaffected.

TABLE 7 QLIKE losses and model confidence sets: low/normal/high-volatility regimes

	Low-volatility regime					Normal-volatility regime					High-volatility regime				
	1d	2w	1m	2m	3m	1d	2w	1m	2m	3m	1d	2w	1m	2m	3m
<b>GARCH-MIDAS</b>															
RVol(22)	0.364	0.264	0.305	0.399	0.409	0.271	0.232	0.260	0.400	0.463	0.273	0.241	0.217	0.313	0.365
VIX	0.332	0.210	0.250	0.367	0.354	0.233	0.204	0.231	0.355	0.454	0.262	0.259	0.243	0.347	0.437
VRP	0.349	0.237	0.288	0.405	0.424	0.252	0.215	0.245	0.375	0.440	0.266	0.238	0.237	0.360	0.414
NFCI	0.400	0.274	0.304	0.389	0.402	0.272	0.228	0.248	0.364	0.417	0.292	0.244	0.217	0.293	0.297
NAI	0.438	0.308	0.338	0.432	0.460	0.284	0.240	0.260	0.384	0.427	0.292	0.241	0.216	0.309	0.322
Δ IP	0.441	0.313	0.343	0.437	0.468	0.287	0.241	0.262	0.389	0.447	0.288	<b>0.236</b>	<b>0.213</b>	0.310	0.335
Δ Housing	0.402	0.277	0.300	0.386	0.406	0.279	0.234	0.249	<b>0.331</b>	<b>0.385</b>	0.295	0.249	0.219	0.300	0.298
VIX and NFCI	0.331	0.212	0.250	0.364	<b>0.351</b>	0.235	<b>0.203</b>	<b>0.228</b>	0.345	0.440	<b>0.254</b>	0.249	0.229	0.321	0.391
VIX and Δ Indpro	0.332	0.211	0.251	0.363	0.351	0.234	0.204	0.230	0.354	0.450	0.257	0.253	0.237	0.338	0.424
VIX and NAI	0.333	0.212	0.254	0.367	0.354	0.234	0.205	0.231	0.358	0.449	0.259	0.254	0.237	0.334	0.417
VIX and Δ Housing	0.330	0.213	0.254	0.369	0.359	0.237	0.209	0.234	0.343	0.439	0.260	0.261	0.242	0.333	0.409
Avg.	0.396	0.273	0.303	0.391	0.403	0.269	0.226	0.247	0.362	0.418	0.272	0.240	0.217	0.306	0.335
<b>Competitor models</b>															
GARCH	0.430	0.296	0.325	0.419	0.452	0.285	0.241	0.263	0.394	0.441	0.300	0.252	0.232	0.340	0.370
MS-GARCH-TVI	0.468	0.338	0.370	0.452	0.519	0.303	0.270	0.298	0.429	0.488	0.286	0.246	0.233	0.348	0.405
MS-GARCH-TVC	0.461	0.318	0.335	0.414	0.437	0.290	0.245	0.263	0.386	0.432	0.295	0.239	0.218	0.324	0.350
Real GARCH	0.237	<b>0.182</b>	<b>0.239</b>	0.380	0.409	0.256	0.208	0.229	0.358	0.408	0.331	0.261	0.229	<b>0.287</b>	<b>0.289</b>
HEAVY	0.272	0.223	0.326	0.591	0.759	0.262	0.228	0.273	0.498	0.593	0.339	0.305	0.313	0.535	0.642
HAR	0.234	0.189	0.254	<b>0.359</b>	0.385	0.243	0.212	0.238	0.374	0.430	0.340	0.257	0.230	0.371	0.470
HAR (lev.)	<b>0.226</b>	0.187	0.258	0.362	0.387	<b>0.232</b>	0.211	0.240	0.378	0.429	0.286	0.245	0.227	0.373	0.470
No-change	0.418	0.821	1.143	2.213	2.310	0.304	0.297	0.336	0.532	0.715	0.382	0.320	0.314	0.481	0.555

Note. Numbers reported are the average out-of-sample QLIKE losses for each model for 1-day- (1d), 2-week- (2w), 1-month- (1m), 2-month- (2m), and 3-month-ahead (3m) variance forecasts across three different volatility regimes; forecasts are issued at a day for which the daily realized volatility is below the empirical 25% quantile (low regime), between the 25% and 75% quantile (normal regime), or above the 75% quantile (high regime). Bold entries indicate the model with the lowest average QLIKE loss per regime and horizon. Shaded entries indicate that the respective model is included in the 90% model confidence set. The average forecast (avg.) is the mean forecast across all GARCH-MIDAS models employing one explanatory variable. The out-of-sample evaluation period spreads from 2010:M1 to 2018:M4.



the final MCS is the HAR model with leverage. Thus, at the very short horizon of 1 day, the HAR with leverage dominates all other models. At forecast horizons of 2 weeks the MCS includes both HAR models, the realized GARCH, and GARCH-MIDAS specifications that either include the VIX alone or in combination with the NFCI/NAI/IP. At the 1-month horizon only the realized GARCH and the GARCH-MIDAS that combines the VIX and the NFCI are included. The picture changes at horizons of 2 and 3 months. At these horizons GARCH-MIDAS models that either combine the VIX with the NFCI/housing starts or models based on housing starts alone are included in the MCS. These results illustrate that the performance of a GARCH-MIDAS model strongly depends on choosing the best horizon-specific explanatory variable. In summary, the HAR model with leverage and the realized GARCH achieve the lowest QLIKE at forecast horizons of 1 day and 2 weeks/1 month, respectively. In contrast, the GARCH-MIDAS model based on housing starts performs best at horizons of 2 and 3 months ahead (see the bold entries).

*MCS for volatility regimes.* In addition to the results for the full OOS period, we also provide MCS for subsamples of low, normal, and high volatility. We define these regimes in the same way as outlined in Section 3.3. Quantiles are now computed based on the empirical distribution of full-sample realized variances. In total, we have 764 observations in the low, 961 in the normal, and 304 in the high regime. Table 7 presents the regime-specific analysis.

Interestingly, in the low-volatility regime the realized GARCH and the two HAR models are the only models in the MCS for short horizons of 1 day and 2 weeks. For a forecast horizon of 1 month, various GARCH-MIDAS models are included in the MCS. For 3 months ahead, two GARCH-MIDAS specifications based on the VIX are the only models in the MCS. The results for the normal-volatility regime are even more in favor of the GARCH-MIDAS models. At essentially all horizons GARCH-MIDAS models based on the VIX are included in the MCS. As for the full OOS period, GARCH-MIDAS based on housing starts is the only model in the 3-month MCS. Finally, in the high-volatility regime and for horizons of 2 weeks and 1 month, essentially all models are included in the MCS. This result may be driven by the fact that the intermediate-term

TABLE 8 Mincer–Zarnowitz  $R^2$

	Panel A: Full out-of-sample period					Panel B: Volatility regimes		
	1d	2w	1m	2m	3m	Low 1m	Normal 1m	High 1m
<b>GARCH-MIDAS</b>								
RVol(22)	0.312	0.367	0.340	0.086	0.008	0.037	0.061	0.314
VIX	0.347	0.346	0.321	0.145	0.047	0.071	0.099	0.297
VRP	0.343	<b>0.404</b>	0.354	0.128	0.030	0.041	0.083	0.324
NFCI	0.295	0.375	0.354	0.146	0.062	0.030	0.073	0.341
NAI	0.294	0.373	0.352	0.143	0.062	0.025	0.071	0.339
$\Delta$ IP	0.296	0.374	0.348	0.124	0.029	0.017	0.065	0.341
$\Delta$ Housing	0.293	0.372	0.355	<b>0.168</b>	<b>0.102</b>	0.031	0.077	0.334
VIX and NFCI	<b>0.353</b>	0.359	0.333	0.147	0.050	0.072	0.100	0.302
VIX and NAI	0.348	0.349	0.323	0.146	0.048	0.067	0.099	0.297
VIX and $\Delta$ IP	0.348	0.347	0.321	0.145	0.047	0.070	0.099	0.296
VIX and $\Delta$ Housing	0.347	0.346	0.321	0.153	0.056	0.064	0.099	0.295
Avg.	0.322	0.380	0.357	0.149	0.057	0.036	0.078	0.341
<b>Competitor models</b>								
GARCH	0.288	0.373	0.353	0.138	0.051	0.027	0.068	0.343
MS-GARCH-TVI	0.316	0.357	0.288	0.118	0.016	0.005	0.015	0.339
MS-GARCH-TVC	0.311	0.390	0.368	0.142	0.052	0.030	0.066	<b>0.374</b>
Real GARCH	0.318	0.394	<b>0.377</b>	0.146	0.070	<b>0.076</b>	<b>0.112</b>	0.303
HEAVY	0.297	0.322	0.272	0.061	0.004	0.028	0.084	0.173
HAR	0.312	0.394	0.374	0.125	0.052	0.058	0.087	0.315
HAR (lev.)	0.342	0.392	0.366	0.122	0.053	0.056	0.088	0.303
No-change	0.254	0.227	0.189	0.060	0.020	0.046	0.044	0.088

Note. We report coefficients of determination derived from MZ regressions. Bold entries indicate the models with the highest  $R^2$  for a specific forecast horizon. The last three columns correspond to the forecast evaluation divided into three volatility regimes; forecasts are issued at a day for which the daily realized volatility is below the empirical 25% quantile (low regime), between the 25% and 75% quantile (normal regime), or above the 75% quantile (high regime). The out-of-sample evaluation period spreads from 2010:M1 to 2018:M4.

forecast performance of all models substantially deteriorates during the high-volatility regime and, therefore, it becomes increasingly difficult to distinguish between models. Nevertheless, even in the high-volatility regime the GARCH-MIDAS models are very competitive for longer forecast horizons. Specifically, GARCH-MIDAS models based on the NFCI and housing starts are included in the MCS.

In summary, we find that the informative content of the explanatory variables depends on the volatility regime. While in low- and normal-volatility regimes GARCH-MIDAS models based on the VIX or VIX combined with another variable perform well, in high-volatility regimes models purely based on macroeconomic variables are very competitive. Because recessions typically coincide with regimes of high volatility, our results are consistent with the finding from the previous literature that macroeconomic variables are particularly useful to predict financial volatility during the onset of recessions (see, e.g., Paye, 2012). At the longest forecast horizons, housing starts and the NFCI become increasingly important. Among the competitor models it is again the realized GARCH which performs very well across volatility regimes.

#### 4.4.3 | MZ regressions

Lastly, we consider the outcome of MZ regressions. As Table 8 shows, for forecast horizons of 1 day and 2 weeks the highest  $R^2$  is achieved by GARCH-MIDAS-type models. This is in sharp contrast to the results from the previous section. However, for longer forecast horizons (1m–3m) the winning models according to the  $R^2$  are exactly the same as when using the MCS approach. Thus, at forecast horizons at which the correct modeling of the long-term component pays off, the  $R^2$  selects the same model as the MCS. Again, the last three columns of Table 8 show that the highest  $R^2$ s are obtained in the high-volatility regime.<sup>28</sup>

## 5 | CONCLUSION

We introduce and discuss the properties of a class of multiplicative volatility models. This class of models includes the GARCH-MIDAS but also a variant of the MS-GARCH. We show that multiplicative volatility models can generate an autocorrelation structure in the conditional variance that mimics the long-memory-type behavior that is often observed for realized variances. We also argue that the  $R^2$  of an MZ regression can be a misleading measure of forecast accuracy across volatility regimes because the  $R^2$  will be highest in the regime with the highest squared error loss. In a Monte Carlo simulation, we investigate the properties of the QMLE of the GARCH-MIDAS model and show that the estimator is unbiased and that the Wang and Ghysels (2015) asymptotic standard errors are valid in the presence of exogenous explanatory variables. We also reveal that forecast performance is relatively insensitive with respect to moderate misspecification of the explanatory variable and the true lag length.

In an empirical application to S&P 500 stock returns, we compare the forecast performance of the GARCH-MIDAS model with a wide range of competitor models. As expected, relative forecast performance depends on the forecast horizon. Among all models, the HAR with leverage performs best at a one-day horizon. For longer forecast horizons the realized GARCH is very competitive and performs best at forecast horizons of 2 weeks and 1 month. The performance of GARCH-MIDAS models depends on the choice of explanatory variable. The best GARCH-MIDAS specifications generate volatility forecasts that are comparable to or improve upon the forecasts from the realized GARCH. Specifically, GARCH-MIDAS specifications that combine the VIX with the NFCI are included in the MCS for forecast horizons of 2 weeks up to 2 months. Most importantly, the GARCH-MIDAS based on housing starts achieves the lowest QLIKE at forecast horizons of 2 and 3 months ahead. Thus our results are useful for selecting the appropriate horizon-specific explanatory variable and suggest that models based on low-frequency information can be more useful than models that exploit high-frequency intraday data.

## ACKNOWLEDGMENTS

We would like to thank the co-editor, Eric Ghysels, and two anonymous referees for comments, which greatly improved our paper. We also thank Richard Baillie, Matei Demetrescu, Markus Haas, Karin Stürmer, Timo Teräsvirta, and Peter Winker for helpful comments and suggestions. A preliminary version of this paper was circulated under the title “On the statistical properties of multiplicative GARCH models.”

<sup>28</sup>For brevity, we now focus on a forecast horizon of 1 month.

## OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [<http://qed.econ.queensu.ca/jae/datasets/conrad002/>].

## REFERENCES

- Amado, C., Silvennoinen, A., & Teräsvirta, T. (2019). Models with multiplicative decomposition of conditional variances and correlations. In J. Chevallier, S. Goutte, D. Guerreiro, S. Saglio, & B. Sanhaji (Eds.), *Financial mathematics, volatility and covariance modelling*, Vol. 2, 217–260. Milton, UK: Routledge.
- Amado, C., & Teräsvirta, T. (2008). *Modelling conditional and unconditional heteroskedasticity with smoothly time-varying structure*, (SSE/EFI Working Paper Series in Economics and Finance 691). Stockholm: Stockholm School of Economics.
- Amado, C., & Teräsvirta, T. (2013). Modelling volatility by variance decomposition. *Journal of Econometrics*, 175, 142–153.
- Amado, C., & Teräsvirta, T. (2017). Specification and testing of multiplicative time-varying GARCH models with applications. *Econometric Reviews*, 36, 421–446.
- Amendola, A., Candila, V., & Scognamiglio, A. (2017). On the influence of US monetary policy on crude oil price volatility. *Empirical Economics*, 52, 155–178.
- Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39, 885–905.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71, 579–625.
- Andersen, T. G., Bollerslev, T., & Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73, 279–296.
- Asgharian, H., Hou, A. J., & Javed, F. (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., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74, 3–30.
- Billingsley, P. (1995). *Probability and Measure* (3rd ed.). New York, NY: Wiley.
- Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *Review of Financial Studies*, 31, 2730–2773.
- Borup, D., & Jakobsen, J. S. (2019). Capturing volatility persistence: A dynamically complete realized EGARCH-MIDAS model. *Quantitative Finance*, 19, 1839–1855. Advance online publication. <https://doi.org/10.2139/ssrn.3071822>
- Conrad, C., Custovic, A., & Ghysels, E. (2018). Long- and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management*, 11, 23.
- Conrad, C., & Loch, K. (2015). Anticipating long-term stock market volatility. *Journal of Applied Econometrics*, 30, 1090–1114.
- Conrad, C., Loch, K., & Rittler, D. (2014). On the macroeconomic determinants of long-term volatilities and correlations in U.S. stock and crude oil markets. *Journal of Empirical Finance*, 29, 26–40.
- Conrad, C., & Schienle, M. (2018). Testing for an omitted multiplicative long-term component in GARCH models. *Journal of Business and Economic Statistics*. Advance online publication. <https://doi.org/10.1080/07350015.2018.1482759>
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7, 174–196.
- Corsi, F., & Reno, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business and Economic Statistics*, 30, 368–380.
- Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature*, 49, 72–100.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253–263.
- Ding, Z., & Granger, C. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73, 185–215.
- Ding, Z., Granger, C., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1, 83–106.
- Dominicy, Y., & Vander Elst, H. (2015). Macro-driven VAR forecasts: From very high to very low-frequency data. Available at SSRN: <https://ssrn.com/abstract=2701256>
- Dorion, C. (2016). Option valuation with macro-finance variables. *Journal of Financial and Quantitative Analysis*, 51, 1359–1389.
- Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, 95, 776–797.
- Engle, R. F., & Lee, G. (1999). A long-run and short-run component model of stock return volatility. In R. Engle, & H. White (Eds.), *Cointegration, causality, and forecasting: A Festschrift in honour of Clive WJ Granger*. Oxford, UK: Oxford University Press, pp. 475–497.
- Engle, R. F., & Rangel, J. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21, 1187–1222.
- Francq, C., & Thieu, L. Q. (2019). QML Inference For volatility models with covariates. *Econometric Theory*, 35, 37–72.
- Ghalanos, A. (2018). rugarch: Univariate GARCH models. R package version 1.4-0. Available on CRAN: <https://cran.r-project.org/package=rugarch>

- Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74, 1545–1578.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of nominal excess return on stocks. *Journal of Finance*, 48, 1779–1801.
- Haas, M., Mittnik, S., & Paoletta, M. S. (2004). A new approach to markov-switching GARCH models. *Journal of Financial Econometrics*, 2, 493–530.
- Han, H. (2015). Asymptotic properties of GARCH-X processes. *Journal of Financial Econometrics*, 13, 188–221.
- Han, H., & Kristensen, D. (2015). Semiparametric multiplicative GARCH-X model: Adopting economic variables to explain volatility. Toulouse, France: Toulouse School of Economics.
- Han, H., & Park, J. Y. (2014). GARCH With omitted persistent covariate. *Economics Letters*, 124, 248–254.
- Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27, 877–906.
- Hansen, P. R., & Lunde, A. (2006). Consistent ranking of volatility models. *Journal of Econometrics*, 131, 97–121.
- Hansen, P. R., & Lunde, A. (2014). Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory*, 30, 60–93.
- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, 79, 453–497.
- Heber, G., Lunde, A., Shephard, N., & Sheppard, K. K. (2009). *Oxford-Man Institute's Realized Library*. Oxford: Oxford-Man Institute, University of Oxford.
- Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129, 121–138.
- Karanasos, M. (1999). The second moment and the autocovariance function of the squared errors of the GARCH model. *Journal of Econometrics*, 90, 63–76.
- Kleen, O. (2017). alfred: Downloading time series from ALFRED database for various vintages. R package version 0.1.7. Available on CRAN: <https://cran.r-project.org/package=alfred>
- Kleen, O. (2018). mfGARCH: Mixed-frequency GARCH models. R package version 0.1.8. Available on CRAN: <https://cran.r-project.org/package=mfGARCH>
- Laurent, S., Rombouts, J. V. K., & Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*, 173, 1–10.
- Lindblad, A. (2017). Sentiment indicators and macroeconomic data as drivers for low-frequency stock market volatility. (Discussion Paper No.413). Helsinki, Finland: Helsinki Center of Economic Research.
- Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? a comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187, 293–311.
- Mincer, J., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. Mincer (Ed.), *Economic forecasts and expectations: Analysis of forecasting behavior and performance*. New York, NY: National Bureau of Economic Research, pp. 3–46.
- Opschoor, A., van Dijk, D., & van der Wel, M. (2014). Predicting volatility and correlations with financial conditions indexes. *Journal of Empirical Finance*, 29, 435–447.
- Pan, Z., Wang, Y., Wu, C., & Yin, L. (2017). Oil price volatility and macroeconomic fundamentals: A regime switching GARCH-MIDAS model. *Journal of Empirical Finance*, 43, 130–142.
- Patton, A. J. (2006). Volatility forecast comparison using imperfect volatility proxies. (Research Paper 175). Sydney, Australia: Quantitative Finance Research Centre, University of Technology Sydney.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160, 246–256.
- Patton, A. J., & Sheppard, K. (2009). Evaluating volatility and correlation forecasts. In T. Mikosch, J.-P. Kreiß, R. A. Davis, & T. G. Andersen (Eds.), *Handbook of financial time series*. Berlin, Germany: Springer, pp. 801–838.
- Paye, B. S. (2012). “Déjà Vol”: Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, 106, 527–546.
- Shephard, N., & Sheppard, K. (2010). Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25, 197–231.
- Wang, F., & Ghysels, E. (2015). Econometric analysis of volatility component models. *Econometric Theory*, 31, 362–393.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Conrad C, Kleen O. Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS models. *J Appl Econ*. 2020;35:19–45. <https://doi.org/10.1002/jae.2742>