Using Daily Stock Returns to Estimate the Unconditional and Conditional Variances of Lower-Frequency Stock Returns

Abstract

Using daily returns to construct realized measures of the variances of lower-frequency returns is a natural alternative to using intraday returns for this purpose when transaction-level price data are unavailable. Notably, a suitable application of this approach yields realized measures that are unbiased estimators of the unconditional and conditional variances of holding-period returns for any investment horizon. I use a long sample of daily S&P 500 index returns to investigate the merits of constructing realized measures in this fashion. First, I conduct a Monte Carlo study using a data generating process that reproduces the key dynamic properties of index returns. The results of the study suggest that using realized measures constructed from daily returns to estimate the conditional and unconditional variances of lower-frequency returns should lead to substantial increases in efficiency. Next, I fit a multiplicative error model to the realized measures for weekly and monthly index returns to obtain out-of-sample forecasts of their conditional variances. Using the forecasts produced by a generalized autoregressive conditional heteroscedasticity model as a benchmark, I find that the forecasts produced by the multiplicative error model always generate the smallest losses. Furthermore, the performance advantage of forecasts that are based on realized measures is statistically significant in most cases.

Keywords: forecasts, GARCH, realized volatility, realized kernel, multiplicative errors

1. Introduction

In the financial econometrics literature, realized measures of volatility are typically constructed using high-frequency log returns for the trading day (see, e.g., Andersen and Bollerslev, 1998; Andersen et al., 2003; Barndorff-Nielsen et al., 2008). Researchers seldom feel the need to differentiate between simple returns and log returns in this setting because doing so is unnecessary from an empirical perspective. If the holding period for a stock or stock index is a single day, then the difference between the variance of a simple return and the variance of the corresponding log return will typically be negligible. However, the differences between the statistical properties of simple returns and those of log returns become more pronounced as the holding period increases. Thus they are unlikely to be negligible for research that addresses asset pricing, portfolio optimization, and related topics, which is usually conducted using simple returns for weekly, monthly, or quarterly holding periods (see, e.g., Avramov et al., 2006; Yogo, 2006; Kirby and Ostdiek, 2012).

More broadly, it is important to note that the high-frequency data needed to construct daily realized variances may not be available for the entire sample period of interest. The first year of the Trade and Quote data provided by the New York Stock Exchange is 1993. In contrast, the coverage of the daily stock file of the Center for Research in Security Prices begins in 1926. The widespread availability of daily historical data makes it well suited for estimating the unconditional and conditional variances of lower-frequency stock returns. I investigate the performance of this approach using a new technique for constructing realized measures. Unlike the conventional construction technique pioneered by Andersen and Bollerslev (1998), the new technique delivers realized measures that are unbiased estimators of the unconditional and conditional variances of simple returns in a discrete time setting under relatively mild assumptions that are frequently invoked in the volatility modeling literature.

I begin by conducting a Monte Carlo study of the relative estimation errors that result from using the new and conventional realized measures as estimators of the unconditional and conditional variances of simple returns and log returns for a range of different holding periods. The study demonstrates that my technique for constructing realized measures of the variances of simple returns works as intended. I find no evidence of bias for any holding period and the proposed realized measures deliver improvements in efficiency that are comparable to those produced by conventional realized measures of the variances of log returns.

Next, I use S&P 500 index data to investigate the performance of the new realized measures in the context of modeling and forecasting the conditional variances of weekly and monthly index returns. By specifying a multiplicative error model (MEM) of the type introduced by Engle (2002) for the realized measures, I obtain a sequence of pseudo out-of-sample variance forecasts. The accuracy of the MEM forecasts is evaluated relative to that of the variance forecasts produced by a similarly-parameterized specification of the generalized autoregressive conditional heteroskedasticity (GARCH) model of Bollerslev (1986). This is accomplished using the Giacomini and White (2006) test of equal predictive ability.

As anticipated, MEM forecasts produce smaller mean errors, smaller mean absolute errors, and smaller root mean square errors than the GARCH forecasts for every forecast horizon under consideration. This is the case for both weekly index returns and monthly index returns. But the results for monthly returns are stronger from the standpoint of statistical significance. I find that the smallest t-statistics produced by the test of equal predictive ability are 2.36 for monthly log returns and 2.61 for monthly simple returns. Because I reject the hypothesis that the GARCH forecasts are just as accurate as the MEM forecasts at the 1% significance level for monthly simple returns, irrespective of the forecast horizon, I conclude that the proposed realized measures of the variances of simple returns deliver meaningful performance gains.

2. Realized Measures

Suppose that $P(t_i)$ denotes the price of a stock or stock index at time t_i. Further suppose that the price is recorded at a fixed frequency such that there is always one time period between successive elements of the sequence $\{P(t_i)\}_{i=0}^{KT}$, where $K > 1$ and $T > 0$ are integers to be specified later. To develop realized measures that are unbiased estimators of the variance of simple returns in a discrete time setting, I invoke assumptions that eliminate the need to employ the type of fill-in asymptotics that underpin the arguments of Andersen et al. (2003). Henceforth, $\mathcal{F}(t_i)$ denotes the information set that contains all prices realized prior to time t_{i+1}. I presume throughout the discussion that log returns and simple returns are weakly stationary.

2.1. Realized measures computed from log returns

Andersen and Bollerslev (1998) pioneered the use of high-frequency log returns to construct realized measures. It is easy to formulate discrete-time analogs of the basic arguments that motivate their methodology. Let $\tilde{r}(t_i, t_{i+k}) = \log P(t_{i+k}) - \log P(t_i)$ denote the log return for the k-period interval that begins at time t_i and ends at time t_{i+k}, where $0 \leq k \leq K$. I assume for simplicity that $\mathrm{E}[\tilde{r}(t_i, t_{i+1})] = 0$ and use $\tilde{\sigma}_K^2 := \mathrm{var}(\tilde{r}(t_i, t_{i+K}))$ to denote the variance of K-period log returns.

The starting point is to consider a scenario in which the single-period log returns are serially uncorrelated. Because $\tilde{r}(t_i, t_{i+K})$ can be expressed as $\tilde{r}(t_i, t_{i+K}) = \sum_{j=1}^{K} \tilde{r}(t_{i+j-1}, t_{i+j})$, it follows immediately that

$$\tilde{\sigma}_K = \left(\mathrm{E}\left[\sum_{j=1}^{K} \tilde{r}^2(t_{i+j-1}, t_{i+j}) \right] \right)^{1/2}, \tag{1}$$

where $\tilde{r}^2(t_{i+j-1}, t_{i+j})$ denotes the square of $\tilde{r}(t_{i+j-1}, t_{i+j})$. Thus $\tilde{v}(t_i, t_{i+K}) = (\sum_{j=1}^{K} \tilde{r}^2(t_{i+j-1}, t_{i+j}))^{1/2}$ is a realized measure of volatility that satisfies $\mathrm{E}[\tilde{v}^2(t_i, t_{i+K})] = \tilde{\sigma}_K^2$.

It is also easy to see that $T^{-1} \sum_{j=1}^{T} \tilde{v}^2(t_{(j-1)K}, t_{jK})$ and $T^{-1} \sum_{j=1}^{T} \tilde{r}^2(t_{(j-1)K}, t_{jK})$ are unbiased estimators of $\tilde{\sigma}_K^2$. But the former is a lot more efficient than the latter in general. More broadly, $\tilde{v}(t_i, t_{i+K})$ satisfies $\mathrm{E}[\tilde{v}^2(t_i, t_{i+K}) | \mathcal{F}(t_i)] = \mathrm{var}(\tilde{r}(t_i, t_{i+K}) | \mathcal{F}(t_i))$ under suitable assumptions about the dynamic properties of log returns. To grasp the basic requirements for conditional unbiasedness, let $\tilde{\sigma}^2(t_i, t_{i+K}) := \mathrm{var}(\tilde{r}(t_i, t_{i+K}) | \mathcal{F}(t_i))$ and think about a data generating process (DGP) of the form

$$\tilde{r}(t_i, t_{i+1}) = \tilde{\sigma}(t_i, t_{i+1}) \tilde{z}(t_i, t_{i+1}), \quad i = 0, 1, \ldots, KT - 1, \tag{2}$$

where $\tilde{\sigma}(t_i, t_{i+1}) \in \mathcal{F}(t_i)$, $\mathrm{E}[\tilde{z}(t_i, t_{i+1}) | \mathcal{F}(t_i)] = 0$, and $\mathrm{E}[\tilde{z}^2(t_i, t_{i+1}) | \mathcal{F}(t_i)] = 1$ for all i. For example, the DGP could be a GARCH(1,1) model (see Bollerslev, 1986). Because a DGP of this form implies that $\mathrm{E}[\tilde{r}(t_i, t_{i+1}) \tilde{r}(t_{i+j}, t_{i+j+1}) | \mathcal{F}(t_i)] = 0$ for all $j \neq 0$, it follows that $\tilde{\sigma}(t_i, t_{i+K}) = (\sum_{j=1}^{K} \mathrm{E}[\tilde{r}^2(t_{i+j-1}, t_{i+j}) | \mathcal{F}(t_i)])^{1/2}$ by iterated expectations.

2.2. Realized measures computed from simple returns

In a typical finance application (portfolio optimization, risk management, etc.), the analysis focuses on simple returns rather than log returns. Furthermore, the simple returns of interest are often measured at relatively low frequencies (monthly observations, quarterly observations, etc.). I therefore propose a new strategy for constructing realized measures that are unbiased estimators of the unconditional and conditional variances of simple returns. Henceforth, simple returns are just called returns.

Let $r(t_i, t_{i+k}) = P(t_{i+k})/P(t_i) - 1$ denote the return for the k-period interval that begins at time t_i and ends at time t_{i+k}. By straightforward algebra, this quantity can be expressed as

$$r(t_i, t_{i+k}) = \sum_{j=1}^{k} R(t_i, t_{i+j-1}) r(t_{i+j-1}, t_{i+j}), \tag{3}$$

where $R(t_i, t_{i+k}) = P(t_{i+k})/P(t_i)$ denotes the gross return for the k-period interval under consideration. I assume for simplicity that $\mathrm{E}[r(t_i, t_{i+1})] = 0$ and explain how to relax this assumption later on.

Now let $\sigma_K^2 := \text{var}(r(t_i, t_{i+K}))$, $\sigma^2(t_i, t_{i+K}) := \text{var}(r(t_i, t_{i+K})|\mathcal{F}(t_i))$, and consider a scenario in which single-period returns satisfy

$$\text{cov}(r(t_{i+j-1}, t_{i+j}), R(t_i, t_{i+j-1})r(t_{i+k-1}, t_{i+k})R(t_i, t_{i+k-1})) = 0 \tag{4}$$

for all $j > k \geq 1$. Under these circumstances,

$$v(t_i, t_{i+K}) = \left(\sum_{j=1}^{K} R^2(t_i, t_{i+j-1}) r^2(t_{i+j-1}, t_{i+j}) \right)^{1/2} \tag{5}$$

is a realized measure of σ_K that satisfies $\text{E}[v(t_i, t_{i+K})^2] = \sigma_K^2$. Furthermore, it is apparent that $v(t_i, t_{i+K})$ satisfies $\text{E}[v^2(t_i, t_{i+K})|\mathcal{F}(t_i)] = \sigma^2(t_i, t_{i+K})$ under suitable assumptions about the dynamic properties of returns. This is the case, for example, if the DGP takes the form

$$r(t_i, t_{i+1}) = \sigma(t_i, t_{i+1}) z(t_i, t_{i+1}), \quad i = 0, 1, \ldots, KT - 1, \tag{6}$$

where $\sigma(t_i, t_{i+1}) \in \mathcal{F}(t_i)$, $\text{E}[z(t_i, t_{i+1})|\mathcal{F}(t_i)] = 0$, and $\text{E}[z^2(t_i, t_{i+1})|\mathcal{F}(t_i)] = 1$ for all i. To see why, simply note that

$$\text{E}[R(t_i, t_{i+j-1}) r(t_{i+j-1}, t_{i+j}) R(t_i, t_{i+k-1}) r(t_{i+k-1}, t_{i+k})|\mathcal{F}(t_i)] = 0 \tag{7}$$

for all $j \geq 1$, $k \geq 1$, and $j \neq k$ under equation (6).

2.3. Some useful extensions

The methodology can easily be modified to address situations in which the maintained assumptions are deemed too restrictive. For instance, if single-period returns display serial correlation, then a realized kernel approach can be used to construct the realized measures. Barndorff-Nielsen et al. (2008) show that this is an effective way of addressing the presence of serial correlation that is due to microstructure effects.

The assumption that expected returns are equal to zero can also be relaxed. Suppose, for instance, that $\text{E}[r(t_i, t_{i+1})|\mathcal{F}(t_i)] = \mu$ for all i. In this case,

$$\text{E}[(1 + \mu)^{-k} R(t_i, t_{i+k}) - 1|\mathcal{F}(t_i)] = 0 \tag{8}$$

for all i and $k \geq 0$, so it is a simple matter to show that

$$\text{var}\left(\frac{R(t_i, t_{i+K})}{(1+\mu)^K} \middle| \mathcal{F}(t_i) \right) = \text{E}\left[\sum_{j=1}^{K} \left(\frac{R(t_i, t_{i+j-1})}{(1+\mu)^{j-1}} \right)^2 \left(\frac{r(t_{i+j-1}, t_{i+j}) - \mu}{1+\mu} \right)^2 \middle| \mathcal{F}(t_i) \right] \tag{9}$$

by mirroring the arguments for the $\mu = 0$ case. The realized measure

$$v^*(t_i, t_{i+K}) = \left(\sum_{j=1}^{K} (1+\mu)^{2(K-j)} R^2(t_i, t_{i+j-1}) (r(t_{i+j-1}, t_{i+j}) - \mu)^2 \right)^{1/2} \tag{10}$$

therefore satisfies $\text{E}[v^{*2}(t_i, t_{i+K})|\mathcal{F}(t_i)] = \text{var}(r(t_i, t_{i+K})|\mathcal{F}(t_i))$.

It is clear from these results that $v^2(t_i, t_{i+K})$ is a biased estimator of $\sigma^2(t_i, t_{i+K})$ for cases in which $\mu \neq 0$, just as $\tilde{v}^2(t_i, t_{i+K})$ is a biased estimator of $\tilde{\sigma}^2(t_i, t_{i+K})$ for cases in which $\text{E}[\tilde{r}(t_i, t_{i+1})|\mathcal{F}(t_i)] \neq 0$. But the results also show how to implement a simple bias correction for $v^2(t_i, t_{i+K})$. In particular, a bias-corrected realized measure for returns can be obtained by substituting a consistent estimator of $\text{E}[r(t_i, t_{i+1})]$ that is available at time t_i for μ in equation (10). The effect of this correction will typically be quite small for realized measures that are constructed from daily stock returns because the sample mean of daily returns is typically only a few basis points. This is the reason why studies that fit volatility models to daily stock returns often assume that expected returns are equal to zero (see, e.g., Visser, 2011).

3. Monte Carlo Analysis

I use Monte Carlo integration to document the properties of the unbiased variance estimators discussed in Section 2. The DGP for the study is a well-known variant of the GARCH(1,1) model of Bollerslev (1986). In particular, I generate the single-period log returns from the model

$$\tilde{r}(t_i, t_{i+1}) = \kappa\tilde{\sigma}^2(t_i, t_{i+1}) + \tilde{\sigma}(t_i, t_{i+1})\tilde{z}(t_i, t_{i+1}), \tag{11}$$

$$\tilde{\sigma}^2(t_i, t_{i+1}) = \omega + \beta\tilde{\sigma}^2(t_{i-1}, t_i) + \alpha(\tilde{z}(t_{i-1}, t_i) - \gamma\tilde{\sigma}(t_{i-1}, t_i))^2, \tag{12}$$

where $\omega \geq 0$, $\beta \geq 0$, $\alpha \geq 0$, $(\beta + \alpha\gamma^2) < 1$, and $\tilde{z}(t_i, t_{i+1})$ is an independent and identically distributed (i.i.d.) standard normal random variable. This specification is well suited to Monte Carlo work because it allows $\tilde{\sigma}^2_K$, σ^2_K, $\tilde{\sigma}^2(t_i, t_{i+K})$, and $\sigma^2(t_i, t_{i+K})$ to be computed analytically.[1]

Daily S&P 500 index data for the years 1946 through 2023 (19,835 observations) are used to calibrate the DGP. The data are from two sources: the daily stock file of the Center for Research in Security Prices for July 3, 1962 to December 29, 2023 and a dataset compiled by Schwert (1990) for January 2, 1946 to July 2, 1962.[2] First, I use the method of maximum likelihood to fit the model to daily log index returns subject to $\kappa = 0$ and $\omega = 0$.[3] Second, I set the values of α, β, and γ in equations (11) and (12) equal to their maximum likelihood estimates, generate $\{\tilde{r}(t_i, t_{i+1}\}_{i=0}^{KT-1}$ with $\kappa = 0$ and $\omega = 0$, and construct $\{r(t_i, t_{i+1}\}_{i=0}^{KT-1}$ by setting $\kappa = -1/2$ and computing $r(t_i, t_{i+1}) = \exp(\kappa\tilde{\sigma}^2(t_i, t_{i+1}) + \tilde{r}(t_i, t_{i+1})) - 1$ for all i.[4] Third, I use the simulated data to calculate $\tilde{v}^2(t_i, t_{i+K})$ and $v^2(t_i, t_{i+K})$ for each $i \in \{0, K, 2K, \ldots, (T-1)K\}$. Because there are roughly 252 trading days per year for the S&P 500 index, I consider $K = 5$, $K = 21$, $K = 63$, $K = 126$, and $K = 252$ to approximate weekly, monthly, quarterly, semiannual, and annual holding periods.

Table 1 summarizes the results for 10 million simulated observations (i.e., $T = 1000000$). Panel A examines the properties of the relative estimation errors for unconditional variances. The initial six columns report the mean, mean absolute, and root mean square values of $\tilde{r}^2(t_i, t_{i+K})/\tilde{\sigma}^2_K - 1$ and $r^2(t_i, t_{i+K})/\sigma^2_K - 1$ for the six values of K under consideration (denoted by ME, MAE, and RMSE). For $K = 1$, the results for log returns are nearly identical to those for returns. But differences emerge as K increases.

As anticipated, the mean errors are quite small (zero to three decimal places) because $\tilde{r}^2(t_i, t_{i+K})$ and $r^2(t_i, t_{i+K})$ are unbiased estimators of $\tilde{\sigma}^2_K$ and σ^2_K. The largest RMSEs correspond to $K = 21$ for log returns and $K = 5$ for returns. An increase in the RMSE is always indicative of an increase in kurtosis, which can be expressed as 1 plus the mean square error. The smallest MAEs and RMSEs correspond to $K = 252$.

Now consider the results in the final six columns of panel A, which contain the mean, mean absolute, and root mean square values of $\tilde{v}^2(t_i, t_{i+K})/\tilde{\sigma}^2_K - 1$ and $v^2(t_i, t_{i+K})/\sigma^2_K - 1$ for the six values of K under consideration.[5] The realized measures show no indications of bias and are clearly much more efficient estimators of $\tilde{\sigma}^2_K$ and σ^2_K for $K > 1$ than $\tilde{r}^2(t_i, t_{i+K})$ and $r^2(t_i, t_{i+K})$. Notice, for example, that replacing $r^2(t_i, t_{i+K})$ with $v^2(t_i, t_{i+K})$ reduces the RMSE from 1.666 to 0.856 with $K = 5$. This is a reduction of 48.6%. Furthermore, the improvements in efficiency become more pronounced as K increases. The RMSE drops from 1.407 to 0.194 for the $K = 252$ case, which is a reduction of 86.2%.

---

[1] See the note to Table 1 for specifics.

[2] Downloadable from https://www.billschwert.com/dstock.htm.

[3] The latter constraint is imposed because the nonnegativity restriction on ω is binding for the sample under consideration. This is a common finding for this model in the literature (see, e.g., Christoffersen et al., 2013).

[4] The maximum likelihood estimates of the parameters are given in the note to Table 1. Notice that the simulated log returns and simulated returns have a population mean of zero by construction.

[5] The results for $K = 1$ are identical to those in the first six columns because $\tilde{v}^2(t_i, t_{i+1}) = \tilde{r}^2(t_i, t_{i+1})$ and $v^2(t_i, t_{i+1}) = r^2(t_i, t_{i+1})$.

Panel B examines the properties of the relative estimation errors for the conditional variances using the same layout as panel A. Once again, the mean errors are zero to three decimal places in all cases and there are large gains in efficiency from employing the realized measures. The reduction in the RMSEs relative to those reported in panel A is one indicator of the benefits exploiting conditioning information. The RMSEs drop by 0.203 (12.6%) in all cases for $K = 1$. As K increases, the drop always becomes smaller in raw numerical terms. But the percentage drop in the RMSE does not display a monotonic relation with K. For example, the RMSE for $v^2(t_i, t_{i+K})$ drops by 0.133(15.5%) for $K = 5$.

Overall the Monte Carlo evidence indicates that the proposed technique for constructing realized measures that are unbiased estimators of the variances of multiperiod returns works as intended. It achieves improvements in efficiency that are comparable to those achieved by the conventional technique for constructing realized measures of the variances of multiperiod log returns. Next, I turn to an empirical application that focuses on forecasting the conditional variances of weekly and monthly S&P 500 index returns.

4. Empirical Methodology

To lay the groundwork for the discussion, assume that the objective is to forecast the variance of a financial variable $y(t+s)$ using a realization of the sequence $\{y(1), \dots, y(t)\}$ for some $s \geq 1$. Because the GARCH(1,1) model of Bollerslev (1986) is known to perform well in a variety of settings, it is often used to construct such forecasts. If the DGP is a GARCH(1,1) specification, then $y(t + s)$ can be expressed as

$$y(t+s) = \mu + h^{1/2}(t+1,s)z(t+s), \tag{13}$$

$$h(t+1,s) = (1-\phi^{s-1})\eta + \phi^{s-1}h(t+1,1), \tag{14}$$

$$h(t+1,1) = \eta + \phi(h(t,1)-\eta) + \delta(y(t)-\mu)^2, \tag{15}$$

where $\mathrm{E}[z(t+s)|y(1), \dots, y(t)] = 0$ and $\mathrm{E}[z^2(t+s)|y(1), \dots, y(t)] = 1$. Thus $h(t+1,s)$ is a conditionally-unbiased s-step-ahead forecast of $e^2(t+s) = (y(t+s)-\mu)^2$.

Now consider an alternative s-step-ahead forecast of $e^2(t+s)$ that is constructed from a realization of the sequence $\{x(1), \dots, x(t)\}$, where $x(t)$ is a realized measure of the variance of $y(t)$ whose dynamics are described by an MEM of the type introduced by Engle (2002). If the DGP is a first-order MEM, which has a recursive structure similar to that of the GARCH(1,1) model, then $x(t+s)$ can be expressed as

$$x(t+s) = m(t+1,s)u(t+s), \tag{16}$$

$$m(t+1,s) = (1-\varphi^{s-1})\varsigma + \varphi^{s-1}m(t+1,1), \tag{17}$$

$$m(t+1,1) = \varsigma + \varphi(m(t,1)-\varsigma) + \lambda x(t), \tag{18}$$

where $\mathrm{E}[u(t+s)|x(1), \dots, x(t)] = 1$. Because $m(t+1,s)$ is a conditionally-unbiased s-step-ahead forecast of $x(t+s)$, it clearly has the potential to outperform $h(t+1,s)$ as a forecast of $e^2(t+s)$.

I focus on the case in which $y(t+s)$ is a weekly or monthly return on the S&P 500 index and the realized measure of its variance is constructed from daily returns. Presumably, variance forecasts based on realized measures should generally be more accurate than those based on weekly or monthly returns. I therefore use the pseudo out-of-sample forecasts produced by the GARCH(1,1) model to benchmark the performance of the pseudo out-of-sample forecasts produced by the MEM model. As in Giacomini and White (2006), the analysis is conducted using limited-memory estimators of the parameters of the models.

To illustrate, suppose $W + s - 1 > 0$ is the number of observations in a rolling window of weekly or monthly returns. For each choice of s and value of $N \in \{1, \dots, T-W-s+1\}$, I construct an estimate of $h(t+1,s)$ for $t = N + W - 1$ using the estimate of $\boldsymbol{\theta} := (\mu, \eta, \phi, \delta)$ obtained by minimizing

$$Q_h(\boldsymbol{\theta}; s, N) = \sum_{t=N}^{N+W-1} \frac{1}{2}\log(h(t,s)) + \frac{1}{2}\left(\frac{y(t+s-1)-\mu}{h(t,s)}\right)^2 \tag{19}$$

subject to $h(N, 1) = \eta$, $\mu = \hat{\mu}$, and $\eta = \hat{\eta}$, where $\hat{\mu} = W^{-1}\sum_{t=N}^{N+W-1} y(t)$ and $\hat{\eta} = W^{-1}\sum_{t=N}^{N+W-1}(y(t) - \hat{\mu})^2$. Similarly, for each choice of s and value of N, I construct an estimate of $m(t+1, s)$ for $t = N + W - 1$ using the estimate of $\boldsymbol{\vartheta} := (\varsigma, \varphi, \lambda)$ obtained by minimizing

$$Q_m(\boldsymbol{\vartheta}; s, N) = \sum_{t=N}^{N+W-1} \log(m(t, s)) + \frac{x(t+s-1)}{m(t, s)} \tag{20}$$

subject to $m(N, 1) = \varsigma$ and $\varsigma = \hat{\varsigma}$, where $\hat{\varsigma} = W^{-1}\sum_{t=N}^{N+W-1} x(t)$. The resultant estimated values of μ, $h(t+1, s)$, and $m(t+1, s)$ are denoted by $\hat{\mu}(t+1, s)$, $\hat{h}(t+1, s)$, and $\hat{m}(t+1, s)$.

Several features of this procedure are worthy of further comment. First, apart from an excluded additive constant, $-Q_h(\boldsymbol{\theta}; s, N)$ and $-Q_m(\boldsymbol{\vartheta}; s, N)$ are the quasi log likelihood functions that result from treating $z(t)$ as i.i.d. $N(0, 1)$ and $u(t)$ as an i.i.d. exponential random variable with a rate parameter of one. Thus the resultant estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ are consistent under the usual regularity conditions for quasi maximum likelihood estimation. Second, I use the sample mean of $y(t)$, sample variance of $y(t)$, and sample mean of $x(t)$ that are computed from the initial W observations of the rolling window as estimators of μ, η, and ς. This targeting approach simplifies optimization. Third, the procedure produces horizon-tuned forecasts because the estimates of ϕ, δ, φ, and λ are specific to the value of s under consideration.[6]

To formally compare the accuracy of $\hat{h}(t+1, s)$ and $\hat{m}(t+1, s)$ as s-step-ahead forecasts of $\hat{e}^2(t+s) = (r(t+s) - \hat{\mu}(t+1, s))^2$, I use the unconditional version of the Giacomini and White (2006) test of equal predictive ability. The test is based on the criterion

$$\Delta L(t+s) = \left|\frac{\hat{e}^2(t+s)}{\hat{h}(t+1, s)} - 1\right| - \left|\frac{\hat{e}^2(t+s)}{\hat{m}(t+1, s)} - 1\right|, \quad t = W, W+1, T-s, \tag{21}$$

which is the difference between the absolute error losses produced by $\hat{h}(t+1, s)$ and $\hat{m}(t+1, s)$.[7] The null hypothesis for the test is $H_0\colon \mathrm{E}[\Delta L(t+s)] = 0$. Hence, inference is conducted using the t-statistic for

$$\Delta \bar{L}(s) = \frac{1}{T - W - s + 1} \sum_{t=W}^{T-s} \Delta L(t+s). \tag{22}$$

If $\Delta\bar{L}(s)$ is positive and statistically significant, then the test indicates that the s-step-ahead MEM forecasts outperform the s-step-ahead GARCH(1,1) forecasts under the specified loss function.[8]

The weekly and monthly index returns along with their realized variances are computed from daily index data for the years 1946 through 2023. As is typical in the finance literature, I use the actual number of trading days in a given week or given month rather than a fixed value of K for the computations. Because the daily index returns display some evidence of negative first-order serial correlation, I account for the impact of this feature by computing the realized measures as

$$v^2(t_i, t_{i+D}) = \sum_{j=1}^{D} R^2(t_i, t_{i+j-1}) r^2(t_{i+j-1}, t_{i+j}) + 2\sum_{j=1}^{D-1} R(t_i, t_{i+j-1}) r(t_{i+j-1}, t_{i+j}) R(t_i, t_{i+j}) r(t_{i+j}, t_{i+j+1}) \tag{23}$$

rather than as shown in Section 2.[9] Here D denotes the number of trading days for the week or month in

[6]This approach to constructing multi-step-ahead variance forecasts is discussed in detail by Shephard and Sheppard (2010).

[7]I use absolute error loss rather than squared error loss to mitigate the impact of the pronounced excess kurtosis of S&P 500 index returns, which substantially inflates the variance of $\hat{e}^2(t+s)$.

[8]I use the Newey and West (1987) estimator with a lag length of $s-1$ to estimate the long-run variance of $\Delta\bar{L}(s)$.

[9]Technically, the autocorrelation correction in equation (23) could cause $v^2(t_i, t_{i+D})$ to be negative. But this never occurs in the

question. I specify $W = 2820$ for the weekly data and $W = 480$ for the monthly data (50% of the number of available observations in each case). To aid in interpreting the findings, I also conduct tests of equal predictive ability using weekly and monthly observations of log returns and their realized variances.

5. Empirical results

Table 2 examines the properties of the sequence of parameter estimates produced by the rolling-window optimizations for each specification. Panels A and B present the results for weekly log returns and weekly returns. Not surprisingly, the average estimates of ϕ and φ for $s = 1$ point to strong persistence in the conditional variances for both log returns and returns. The results also indicate that the estimates of ϕ and φ are quite stable over time. In panel A, for example, the estimate of ϕ for $s = 1$ ranges from 0.943 to 0.977 and the estimate of φ for $s = 1$ ranges from 0.961 to 0.976.

The results for δ and λ in panel A display some interesting patterns. First, the average estimate of δ is somewhat smaller than the average estimate of λ for $s = 1$, $s = 3$, and $s = 6$. This finding suggests the conditional variance process of weekly log returns displays a weaker response to shocks under the GARCH specification than under the MEM specification. Second, the average estimate of δ declines monotonically with s, whereas the average estimate of λ does not. But there is a sharp drop in the average estimate of λ for $s = 12$. Although the underlying mechanism that leads to this finding is not immediately apparent, the findings for weekly returns mirror those for weekly log returns in all respects.

Panels C and D present the results for monthly log returns and monthly returns. As anticipated, the average estimates of ϕ and φ are somewhat lower than the corresponding values in panels A and B, which is consistent with returns following a stationary stochastic process. But the results still point to substantial degree of persistence in the conditional variances. There is also more variation in the estimates of ϕ and φ over time for the monthly observations, which is obviously associated with the sharp reduction in the number of observations in the rolling window used for estimation purposes.

Perhaps the most intriguing aspect of the results in panels C and D is that the average estimate of δ is considerably smaller than the average estimate of λ for $s = 1$, $s = 6$, and $s = 12$. This pattern suggests that the GARCH specification produces a smoother sequence of conditional variance forecasts than the MEM specification, which could indicate that the latter specification has an advantage in tracking the conditional variances. Notice that the average estimate of λ for $s = 3$ is relatively low by comparison. Because $s = 3$ for monthly observations is roughly equivalent to $s = 12$ for weekly observations, the relation between the average estimate of λ and the forecast horizon is similar at both frequencies.

To develop further insights, I plot the conditional volatility forecasts for weekly returns and monthly returns. Figure 1 shows side-by-side plots of the GARCH and MEM forecasts for weekly returns. The upper panels are for $s = 1$ and lower panels are for $s = 12$. Although the side-by-side comparisons highlight the broad similarities in the forecasts for both forecast horizons, it is easy to spot a few differences. For instance, the spike in the one-step-ahead forecast of conditional volatility that follows the 1987 stock market crash is larger for the MEM specification than for the GARCH specification. But it is clear from the plots that the GARCH and MEM forecasts are highly correlated as a general rule.

Of course, this finding does not necessarily imply that the differences in the predictive ability of GARCH and MEM forecasts is negligible. If the MEM forecasts are more efficient than the GARCH forecasts, then they should have a performance advantage in formal statistical tests provided that the sample size is sufficiently large. Furthermore, the results of the Monte Carlo analysis indicate that gains from employing realized measures are inversely related to the investment horizon employed for the analysis.

Consider the side-by-side plots of the GARCH and MEM forecasts for monthly returns, which are shown in Figure 2. The visual differences in the plots are certainly more pronounced in this case. Not only are the

---

empirical application. Under a realized kernel approach, the second summation would be multiplied by 1 rather than by 2.

one-step-ahead GARCH forecasts relatively smooth, they are also confined to a much narrower range than one-step-ahead MEM forecasts. These features are broadly consistent with a scenario in which the realized measures are more efficient estimators of the conditional variances than the squared demeaned returns.

The tests for equal predictive ability provide formal evidence in this regard. The results of the tests are presented in Table 3. The initial eight columns of the table report the mean, mean absolute, root mean square, mean square values of $\hat{e}^2(t+s)/\hat{h}(t,s) - 1$ and $\hat{e}^2(t+s)/\hat{m}(t,s) - 1$ for the four choices of s: 1, 3, 6, and 12. The final three columns report $\Delta\bar{L}(s)$, its t-statistic, and the associated p-value.

The results in panel A are for weekly log returns. Notably, the MEM forecasts produce smaller MEs, MAEs, and RMSEs than the GARCH forecasts at every forecast horizon. The largest difference in the RMSE corresponds to $s = 3$: 3.604 versus 2.440. But the test of equal predictive ability produces a p-value of 0.129 in this case. Broadly speaking, however, the test favors the MEM forecasts. Note that it produces a t-statistic of 1.75 ($p = 0.079$) for $s = 6$ and 2.30 ($p = 0.021$) for $s = 12$.

The results in panel B are for weekly returns. Once again, the MEM forecasts produce smaller MEs, MAEs, and RMSEs than the GARCH forecasts at every forecast horizon. The other findings are also similar to those for weekly log returns. The test of equal predictive ability favors the MEM forecasts, yielding a t-statistic of 1.93 ($p = 0.054$) for $s = 6$ and 2.31 ($p = 0.021$) for $s = 12$.

The results in panel C and D are for monthly log returns and monthly returns. The overall pattern of the MAEs, and RMSEs mirrors that in panels A and B. However, the evidence regarding the superiority of the MEM foecasts is considerably stronger at the monthly frequency. The smallest t-statistics in panels C and D are 2.36 and 2.61, which have p-values of 0.018 and 0.009. Hence, the null hypothesis of equal predictive ability is rejected at the 1% level for every forecast horizon for monthly returns. This finding highlights the extent to which the new realized measures for returns deliver meaningful performance gains.

6. Conclusions

The availability of high-frequency data on stock prices has transformed the volatility modeling literature over the past 25 years. But there are still good arguments for using daily returns to forecast the volatility of longer-horizon returns, especially for sample periods that begin prior to 1993. Because the statistical properties of log returns differ from those of returns and the differences increase with the investment horizon, I show how to construct realized measures that are unbiased estimators of the unconditional and conditional variances of returns in a discrete time setting, provided that the DGP satisfies relatively mild assumptions that are often invoked in the volatility modeling literature. The empirical evidence indicates that using the proposed realized measures to compute out-of-sample forecasts of the variances of weekly and monthly returns on the S&P 500 index leads to significant improvements in forecast accuracy. Hence, the measures should be useful in research that addresses asset pricing, portfolio optimization, and related topics, which is typically conducted using returns for weekly, monthly, or quarterly holding periods.

References

Andersen, T., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885–905.

Andersen, T., Bollerslev, T., Diebold, F., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.

Avramov, D. , Chordia, T., Goyal, A., 2006. Liquidity and autocorrelations in individual stock returns. *The Journal of Finance* 61, 2365–2394 .

Barndorff-Nielsen O., Hansen, P., Lunde, A., Shephard N., 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Christoffersen, P., Heston, S., Jacobs K., 2013. Capturing option anomalies with a variance-dependent pricing kernel. *Review of Financial Studies* 26, 1963–2006.

Engle, R., 2002. New frontiers for ARCH models. *Journal of Applied Econometrics* 17, 425–446.

Giacomini, R., White H., 2006. Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.

Heston, S., Nandi S., 2000. A closed-form GARCH option valuation model. *The Review of Financial Studies* 13, 585–625.

Kirby, C., Ostdiek B., 2012. It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis* 47, .437–467.

Newey, W., West K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Schwert, G. 1990. Indexes of U.S. stock prices from 1802 to 1987. *The Journal of Business* 63, 399–426.

Shephard, N., Sheppard K., 2010. Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 25, 197–231.

Visser, M., 2011. GARCH parameter estimation using high-frequency data. *Journal of Financial Econometrics* 9, 162–197

Yogo, M., 2006. A consumption-based explanation of expected stock returns. *The Journal of Finance* 61, 539–580.

Table 1
Monte Carlo Study of the Relative Estimation Errors for Unconditional and Conditional Variances

Panel A	$\tilde{r}^2(t_i, t_{i+K})/\tilde{\sigma}^2_K - 1$			$r^2(t_i, t_{i+K})/\sigma^2_K - 1$			$\tilde{v}^2(t_i, t_{i+K})/\tilde{\sigma}^2_K - 1$			$v^2(t_i, t_{i+K})/\sigma^2_K - 1$		
K	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
1	−0.000	1.015	1.617	−0.000	1.015	1.618	−0.000	1.015	1.617	−0.000	1.015	1.618
5	0.000	1.027	1.692	0.000	1.025	1.666	0.000	0.615	0.864	0.000	0.612	0.856
21	0.000	1.013	1.734	0.000	1.005	1.635	0.000	0.432	0.576	0.000	0.417	0.548
63	0.000	0.997	1.698	0.000	0.984	1.532	0.000	0.317	0.412	0.000	0.284	0.362
126	0.000	0.987	1.617	0.000	0.972	1.446	0.000	0.244	0.313	0.000	0.204	0.258
252	−0.000	0.978	1.537	−0.000	0.966	1.407	0.000	0.181	0.229	0.000	0.153	0.194
Panel B	$\tilde{r}^2(t_i, t_{i+K})/\tilde{\sigma}^2(t_i, t_{i+K}) - 1$			$r^2(t_i, t_{i+K})/\sigma^2(t_i, t_{i+K}) - 1$			$\tilde{v}^2(t_i, t_{i+K})/\tilde{\sigma}^2(t_i, t_{i+K}) - 1$			$v^2(t_i, t_{i+K})/\sigma^2(t_i, t_{i+K}) - 1$		
K	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
1	−0.000	0.968	1.414	−0.000	0.968	1.415	−0.000	0.968	1.414	−0.000	0.968	1.415
5	−0.000	0.990	1.559	−0.000	0.988	1.533	0.000	0.537	0.732	0.000	0.534	0.723
21	0.000	0.995	1.676	0.000	0.988	1.579	0.000	0.380	0.506	0.000	0.362	0.476
63	0.000	0.993	1.684	0.000	0.980	1.519	0.000	0.300	0.391	0.000	0.265	0.338
126	0.000	0.985	1.613	0.000	0.971	1.441	−0.000	0.238	0.306	0.000	0.197	0.249
252	−0.000	0.978	1.536	−0.000	0.965	1.406	0.000	0.178	0.227	0.000	0.150	0.191

I use Monte Carlo integration to document the performance of competing estimators of $\tilde{\sigma}^2_K := \text{var}(\tilde{r}(t_i, t_{i+K}))$, $\sigma^2_K := \text{var}(r(t_i, t_{i+K}))$, $\tilde{\sigma}^2(t_i, t_{i+K}) := \text{var}(\tilde{r}(t_i, t_{i+K})|\mathcal{F}(t_i))$, and $\sigma^2(t_i, t_{i+K}) := \text{var}(r(t_i, t_{i+K})|\mathcal{F}(t_i))$, where $\tilde{r}(t_i, t_{i+K})$ and $r(t_i, t_{i+K})$ denote the log return and return for the K-period interval that begins at time t_i and ends at time t_{i+K} and $\mathcal{F}(t_i) = \{\tilde{r}(t_0, t_1), \ldots, \tilde{r}(t_{i-1}, t_i)\}$. The columns labeled ME, MAE, and RMSE report the mean, mean absolute, and root mean square values of the relative estimation errors for the indicated estimators. The analysis is carried out using a data generating process (DGP) of the form

$$\tilde{r}(t_i, t_{i+1}) = \kappa\tilde{\sigma}^2(t_i, t_{i+1}) + \tilde{\sigma}(t_i, t_{i+1})\tilde{z}(t_i, t_{i+1}),$$
$$\tilde{\sigma}^2(t_i, t_{i+1}) = \omega + \beta\tilde{\sigma}^2(t_{i-1}, t_i) + \alpha(\tilde{z}(t_{i-1}, t_i) - \gamma\tilde{\sigma}(t_{i-1}, t_i))^2,$$

where $\omega = 0$, $\beta = 0.8754$, $\alpha = 4.554 \times 10^{-6}$, $\gamma = 127.0$, and $\tilde{z}(t_i, t_{i+1}) \sim \text{NID}(0, 1)$. First, I generate the sequence $\{\tilde{r}(t_i, t_{i+1}\}_{i=0}^{KT-1}$ with $\kappa = 0$ and $T = 10000000$. Next, I construct the sequence $\{r(t_i, t_{i+1}\}_{i=0}^{KT-1}$ by setting $\kappa = -1/2$ and computing $r(t_i, t_{i+1}) = \exp(\kappa\tilde{\sigma}^2(t_i, t_{i+1}) + \tilde{r}(t_i, t_{i+1})) - 1$ for all i. Finally, I calculate

$$\tilde{v}^2(t_i, t_{i+K}) = \sum_{j=1}^{K} \tilde{r}^2(t_{i+j-1}, t_{i+j}) + 2\sum_{j=1}^{K-1} \tilde{r}(t_{i+j-1}, t_{i+j})\tilde{r}(t_{i+j}, t_{i+j+1}),$$
$$v^2(t_i, t_{i+K}) = \sum_{j=1}^{K} R^2(t_i, t_{i+j-1})r^2(t_{i+j-1}, t_{i+j}) + 2\sum_{j=1}^{K-1} R(t_i, t_{i+j-1})r(t_{i+j-1}, t_{i+j})R(t_i, t_{i+j})r(t_{i+j}, t_{i+j+1}),$$

$\tilde{\sigma}^2(t_i, t_{i+K})$, and $\sigma^2(t_i, t_{i+K})$ for $i \in \{0, K, 2K, \ldots, (T-1)K\}$, where $R(t_i, t_{i+j}) = 1 + r(t_i, t_{i+j})$. Simple algebra yields $\tilde{\sigma}^2(t_i, t_{i+K}) = K\tilde{\sigma}^2_1 + (1-\rho)^{-1}(1-\rho^K)(\tilde{\sigma}^2(t_i, t_{i+1}) - \tilde{\sigma}^2_1)$ and $\tilde{\sigma}^2_K = K\tilde{\sigma}^2_1$, where $\rho = \beta + \alpha\gamma^2$ and $\tilde{\sigma}^2_1 = (1-\rho)^{-1}(\omega + \alpha)$. To obtain analytic expressions for $\sigma^2(t_i, t_{i+K}) = \text{E}[R^2(t_i, t_{i+K})|\mathcal{F}(t_i)] - 1$ and $\sigma^2_K = \text{E}[R^2(t_i, t_{i+K})] - 1$, I rely on results from the option pricing literature (see Heston and Nandi, 2000, for details). Specifically, it is well known that

$$\text{E}[R^\tau(t_i, t_{i+K})|\mathcal{F}(t_i)] = \exp(a_K(\tau) + b_K(\tau)\tilde{\sigma}^2(t_i, t_{i+1}))$$

under the DGP for the study, where $a_K(\tau)$ and $b_K(\tau)$ are given by the recurrence relations

$$a_K(\tau) = a_{K-1}(\tau) + \omega b_{K-1}(\tau) - \frac{1}{2}\log(1 - 2\alpha b_{K-1}(\tau)) \quad \text{and} \quad b_K(\tau) = \tau(\kappa+\gamma) - \frac{1}{2}\gamma^2 + \beta b_{K-1}(\tau) + \frac{(1/2)(\tau-\gamma)^2}{1 - 2\alpha b_{K-1}(\tau)}$$

with $a_0(\tau) = b_0(\tau) = 0$. Setting $\tau = 2$ produces an expression for $\sigma^2(t_i, t_{i+K}) + 1$, which in turn yields

$$\sigma^2_K + 1 = \exp(a_K(2))\text{E}[\exp(b_K(2)\tilde{\sigma}^2(t_i, t_{i+1}))]$$

by the law of iterated expectations. Because $\text{E}[\exp(\xi(z+v)^2)] = (1-2\xi)^{-1/2}\exp(v^2\xi(1-2\xi)^{-1})$ given that $z \sim \text{N}(0,1)$, the law of iterated expectations also implies that

$$\text{E}[\exp(b_K(2)\tilde{\sigma}^2(t_i, t_{i+1}))] = \exp\left(\sum_{j=1}^{\infty} \omega c_{j-1} - (1/2)\log(1 - 2\alpha c_{j-1})\right),$$

where c_j satisfies the recurrence relation $c_j = \beta c_{j-1} + \alpha c_{j-1}(1 - 2\alpha c_{j-1})^{-1}\gamma^2$ with $c_0 = b_K(2)$.

Table 2
Selected Properties of Rolling-Window Parameter Estimates for GARCH and MEM Specifications

	GARCH with s-step-ahead conditional variances						MEM with s-step-ahead conditional variances					
	ϕ			δ			φ			λ		
s	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Panel A: Weekly log returns												
1	0.943	0.963	0.977	0.098	0.129	0.198	0.959	0.966	0.975	0.143	0.173	0.200
3	0.937	0.968	0.992	0.051	0.116	0.194	0.936	0.964	0.975	0.160	0.210	0.307
6	0.949	0.976	0.994	0.027	0.093	0.208	0.930	0.961	0.976	0.140	0.205	0.313
12	0.968	0.983	0.994	0.019	0.054	0.091	0.971	0.984	0.992	0.027	0.046	0.064
Panel B: Weekly returns												
1	0.946	0.964	0.977	0.097	0.126	0.193	0.961	0.967	0.976	0.143	0.170	0.194
3	0.941	0.968	0.991	0.056	0.116	0.191	0.938	0.965	0.976	0.159	0.208	0.304
6	0.952	0.976	0.994	0.029	0.093	0.195	0.933	0.962	0.977	0.140	0.199	0.307
12	0.968	0.983	0.994	0.020	0.054	0.091	0.972	0.984	0.992	0.028	0.048	0.069
Panel C: Monthly log returns												
1	0.874	0.940	0.975	0.056	0.087	0.132	0.821	0.867	0.922	0.472	0.552	0.667
3	0.867	0.942	0.971	0.091	0.129	0.173	0.872	0.936	0.963	0.151	0.196	0.251
6	0.861	0.934	0.964	0.084	0.141	0.175	0.882	0.929	0.960	0.212	0.449	0.812
12	0.749	0.898	0.939	0.204	0.347	0.760	0.874	0.921	0.955	0.242	0.351	0.423
Panel D: Monthly returns												
1	0.880	0.942	0.975	0.060	0.091	0.136	0.845	0.889	0.937	0.404	0.493	0.613
3	0.861	0.941	0.971	0.102	0.137	0.180	0.875	0.937	0.963	0.154	0.211	0.276
6	0.853	0.931	0.963	0.094	0.145	0.185	0.888	0.934	0.963	0.214	0.356	0.569
12	0.749	0.890	0.938	0.178	0.321	0.762	0.885	0.923	0.954	0.245	0.342	0.416

The table reports selected properties of limited-memory parameter estimates that are computed using s-step-ahead forecasts of the conditional variances of weekly and monthly S&P 500 index returns. The forecasts of the conditional variances are generated by GARCH(1,1) and MEM specifications of the form

$$y(t+s) = \mu + h^{1/2}(t+1,s)z(t+s), \qquad x(t+s) = m(t+1,s)u(t+s),$$
$$h(t+1,s) = (1-\phi^{s-1})\eta + \phi^{s-1}h(t+1,1), \qquad m(t+1,s) = (1-\varphi^{s-1})\varsigma + \varphi^{s-1}m(t+1,1).$$
$$h(t+1,1) = \eta + \phi(h(t,1)-\eta) + \delta(y(t)-\mu)^2, \qquad m(t+1,1) = \varsigma + \varphi(m(t,1)-\varsigma) + \lambda x(t),$$

where $y(t)$ denotes a weekly log return, weekly return, monthly log return, or monthly return and $x(t)$ denotes the corresponding realized measure, which is constructed from either daily log returns (if $y(t)$ is a longer-horizon log return) or daily returns (if $y(t)$ is a longer-horizon return). I conduct the analysis using a quasi-maximum likelihood (QML) approach that employs a rolling window of $W+s-1$ observations to estimate the parameters. In particular, I construct the estimated values of $h(t+1,s)$ and $m(t+1,s)$ for $t = N+W-1$ using a window of observations that begins in period N and ends in period $N+W+s-1$, where $N \in \{1,\ldots,T-W-s+1\}$. To compute the log quasi-likelihood functions, I treat $z(t)$ as an i.i.d. $N(0,1)$ random variable and $u(t)$ as an i.i.d. exponential random variable with a rate parameter of one. Note that this methodology produces horizon-tuned forecasts of the conditional variances because the estimates of ϕ, δ, φ, and λ are specific to the value of s under consideration. I specify $W = 2820$ for the weekly data and $W = 480$ for the monthly data, which is 50% of the number of available observations in each case. The sample period is January 1946 to December 2023.

Table 3
Tests of Equal Predictive Ability Using Realized Measures Constructed from Daily Observations

	GARCH forecasts (s-step-ahead)				MEM forecasts (s-step-ahead)				H_0: E[$\Delta L(t+s)$] = 0		
s	ME	MAE	RMSE	MSE	ME	MAE	RMSE	MSE	$\Delta \bar{L}(s)$	t-stat	pval
Panel A: Weekly log returns											
1	0.067	1.083	2.172	4.716	0.039	1.061	2.127	4.522	0.022	1.41	0.157
3	0.167	1.198	3.604	12.992	0.096	1.124	2.440	5.953	0.075	1.52	0.129
6	0.211	1.262	3.802	14.459	0.170	1.216	3.481	12.118	0.046	1.75	0.079
12	0.205	1.269	3.671	13.476	0.153	1.235	3.615	13.068	0.034	2.30	0.021
Panel B: Weekly returns											
1	0.062	1.077	2.063	4.258	0.033	1.054	2.010	4.042	0.023	1.58	0.115
3	0.150	1.179	3.178	10.100	0.088	1.115	2.293	5.258	0.064	1.55	0.120
6	0.197	1.244	3.462	11.987	0.151	1.195	3.150	9.921	0.049	1.93	0.054
12	0.193	1.255	3.401	11.567	0.144	1.223	3.345	11.191	0.032	2.31	0.021
Panel C: Monthly log returns											
1	0.080	1.104	2.535	6.425	−0.059	0.971	1.900	3.612	0.133	2.83	0.005
3	0.153	1.174	2.655	7.049	−0.057	1.040	2.284	5.217	0.134	4.10	0.000
6	0.160	1.187	2.586	6.689	−0.010	1.068	2.130	4.536	0.119	2.36	0.018
12	0.157	1.195	2.705	7.316	−0.026	1.077	2.301	5.293	0.118	2.89	0.004
Panel D: Monthly returns											
1	0.064	1.080	2.177	4.740	−0.068	0.956	1.727	2.982	0.124	3.28	0.001
3	0.137	1.150	2.336	5.458	−0.052	1.029	2.028	4.114	0.121	4.04	0.000
6	0.136	1.158	2.272	5.163	−0.026	1.043	1.873	3.510	0.115	2.61	0.009
12	0.129	1.164	2.343	5.490	−0.022	1.066	2.043	4.173	0.098	3.02	0.003

The table reports the results of tests of equal predictive ability for the S&P 500 index. The tests are conducted using the s-step-ahead variance forecasts produced by GARCH(1,1) and MEM models for weekly and monthly observations. Under the models,

$$y(t+s) = \mu + h^{1/2}(t+1,s)z(t+s), \qquad x(t+s) = m(t+1,s)u(t+s),$$
$$h(t+1,s) = (1-\phi^{s-1})\eta + \phi^{s-1}h(t+1,1), \qquad m(t+1,s) = (1-\varphi^{s-1})\varsigma + \varphi^{s-1}m(t+1,1).$$
$$h(t+1,1) = \eta + \phi(h(t,1)-\eta) + \delta(y(t)-\mu)^2, \qquad m(t+1,1) = \varsigma + \varphi(m(t,1)-\varsigma) + \lambda x(t),$$

where $y(t)$ denotes a weekly log return, weekly return, monthly log return, or monthly return and $x(t)$ denotes the corresponding realized measure, which is constructed from either daily log returns (if $y(t)$ is a longer-horizon log return) or daily returns (if $y(t)$ is a longer-horizon return). I use a quasi maximum likelihood approach that employs a rolling window of $W + s - 1$ observations to estimate the parameters, which produces horizon-tuned forecasts because the estimates of ϕ, δ, φ, and λ are specific to the value of s under consideration. In particular, I construct the estimated values of $h(t+1,s)$ and $m(t+1,s)$ for $t = N + W - 1$ using a window of observations that begins in period N and ends in period $N + W + s - 1$, where $N \in \{1, \ldots, T - W - s + 1\}$. The estimated values of $h(t+1,s)$ and $m(t+1,s)$ are denoted by $\hat{h}(t+1,s)$ and $\hat{m}(t+1,s)$. I base the tests on the criterion

$$\Delta L(t+s) = \left|\frac{\hat{e}^2(t+s)}{\hat{h}(t+1,s)} - 1\right| - \left|\frac{\hat{e}^2(t+s)}{\hat{m}(t+1,s)} - 1\right|, \quad t = W, W+1, \ldots, T-s, \tag{24}$$

where $\hat{e}^2(t+s) = (r(t+s) - \hat{\mu}(t+1,s))^2$ and $\hat{\mu}(t+1,s)$ is the sample mean of $\{r(t-W+1), \ldots, r(t)\}$. The null hypothesis is H_0: E[$\Delta L(t+s)$] = 0. Hence, inference is conducted using the t-statistic for

$$\Delta \bar{L}(s) = \frac{1}{T - W - s + 1}\sum_{t=W}^{T-s} \Delta L(t+s). \tag{25}$$

If $\Delta \bar{L}(s)$ is positive and statistically significant, then the test indicates that the s-step-ahead MEM forecasts outperform the s-step-ahead GARCH(1,1) forecasts under the specified loss function. The initial eight columns report the mean, mean absolute, root mean square, and mean square values of $\hat{e}^2(t+s)/\hat{h}(t,s) - 1$ and $\hat{e}^2(t+s)/\hat{m}(t,s) - 1$ for the four choices of s: 1, 3, 6, and 12. I specify $W = 2820$ for the weekly data and $W = 480$ for the monthly data, which is 50% of the number of available observations in each case. The sample period is January 1946 to December 2023.

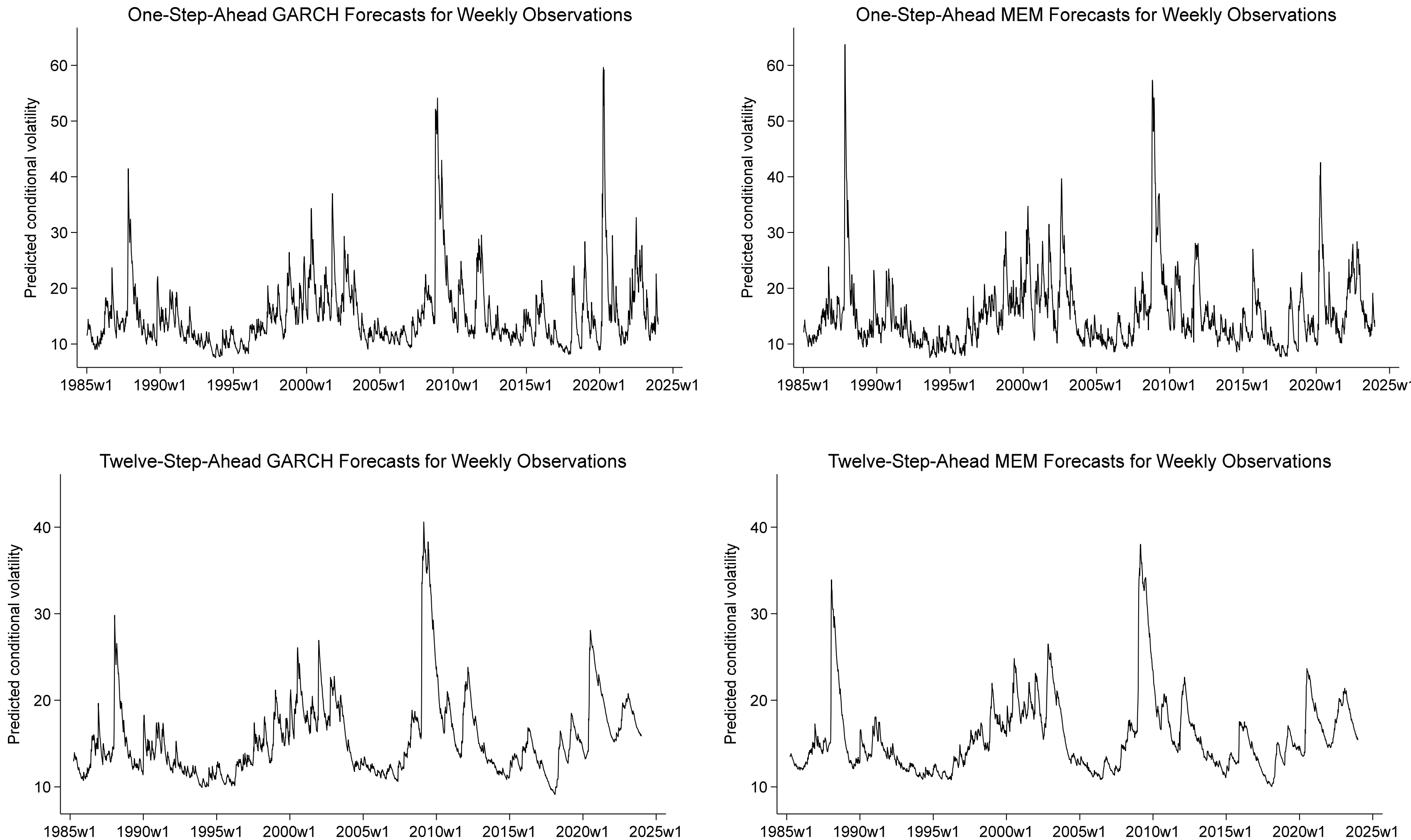

Figure 1: Out-of-Sample Forecasts of Conditional Volatility from GARCH and MEM Specifications for Weekly S&P 500 Index Returns
I use a rolling window of 2820 weekly observations to estimate the parameters of the GARCH and MEM specifications via quasi-maximum likelihood. The forecasts of conditional volatility are expressed as an annualized percentage rate. The out-of-sample forecasts are either for week one of January, 1985 to week four of December 2023 (top tow panels) or for week three of March, 1985 to week four of December 2023 (bottom two panels). The overall sample period is January 1946 to December 2023.

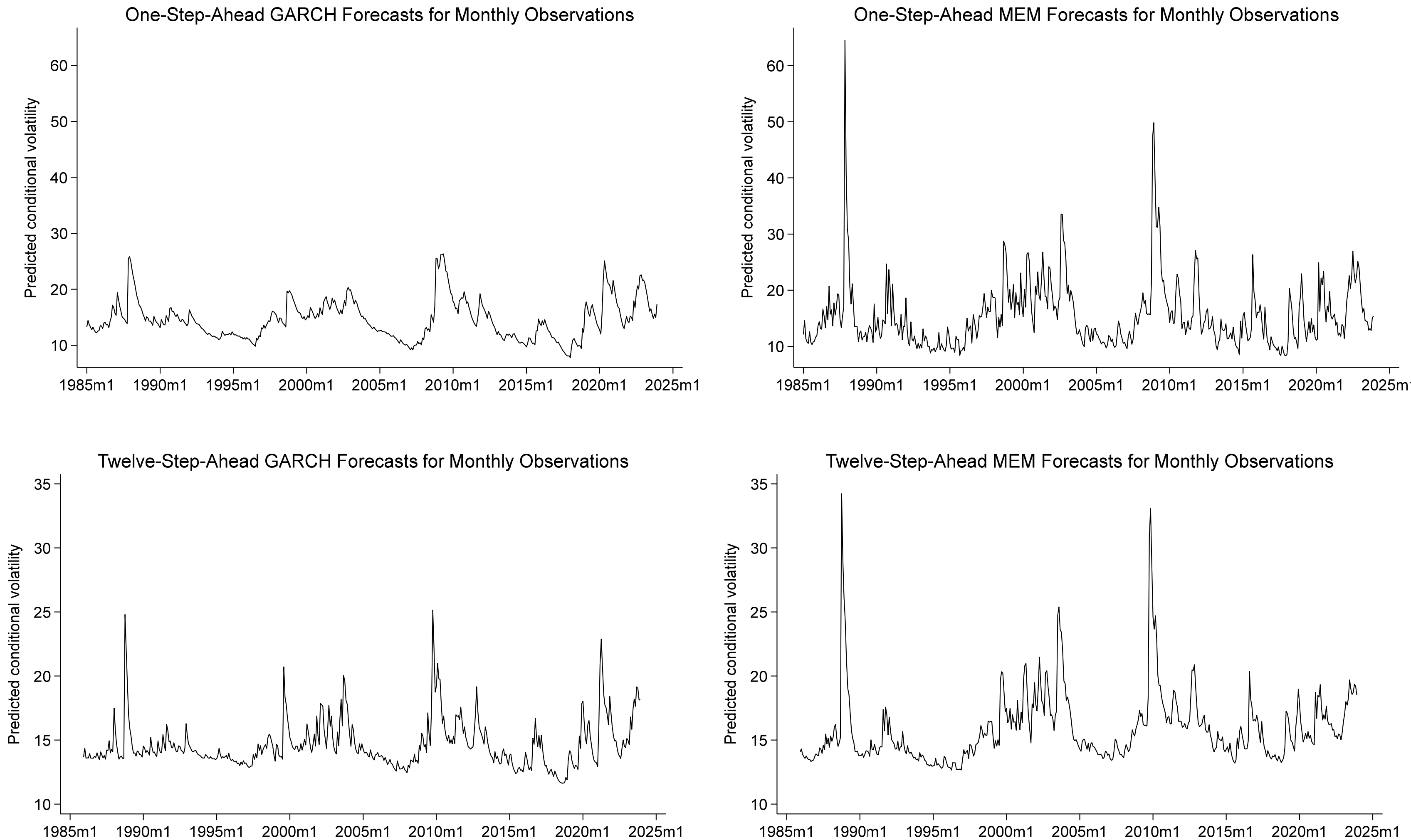

Figure 2: Out-of-Sample Forecasts of Conditional Volatility from GARCH and MEM Specifications for Monthly S&P 500 Index Returns
I use a rolling window of 480 monthly observations to estimate the parameters of the GARCH and MEM specifications via quasi-maximum likelihood. The forecasts of conditional volatility are expressed as an annualized percentage rate. The out-of-sample forecasts are either for January, 1985 to December 2023 (top tow panels) or for December, 1985 to December 2023 (bottom two panels). The overall sample period is January 1946 to December 2023.