Horizon-Targeted Estimation of Volatility Models: Application to a Misspecification Testing and Forecasting

Ekaterina Ugulava*

Sunday 6th July, 2025

Find the latest version here.

ABSTRACT

Multi-period volatility forecasting is crucial for financial decision-making. We consider a scenario where the decision-maker specifies an ex-ante loss function, such as the QLIKE, to assess the accuracy of multi-period volatility forecasts from a candidate volatility model. To reduce the impact of model misspecification on forecast accuracy, we introduce an estimator that is 'matched' to the specification of the forecast evaluation loss function. We examine the estimator's performance under a bias-variance trade-off, highlighting conditions where it is likely to offer improvements over standard estimation methods. We also propose a model misspecification test based on the Hausman principle, which exploits the fact that our estimator and the standard estimator are consistent for the true parameter under the null of correct specification but converge to different pseudo-true values under the alternative. In a Monte Carlo study, we examine the misspecification with respect to long-memory dynamics. Our results show that the misspecification test is reasonably sized and has power that increases with the degree of long-memory misspecification. Additionally, we recover multi-period volatility forecasts and find that under correct specification, both estimators perform equivalently; however, under misspecification, our estimator provides superior forecast accuracy. Finally, an out-of-sample analysis across ten return and realised measure series from 2001 to 2010 suggests three key findings: first, it is optimal for our estimator to match the estimation loss function to a shorter horizon than the forecasting horizon; second, our estimator provides greater accuracy gains for GARCH-type volatility models applied to realised measures of volatility compared to those applied to returns; and third, our estimator leads to greater accuracy gains for underparameterised models (which are more likely to be misspecified), highlighting the bias-variance trade-off.

Keywords: multi-period volatility forecasts, GARCH models, long-memory models, misspecification test, realised measures.

*University of Amsterdam and Tinbergen Institute, e-mail: e.ugulava@uva.nl. I would like to thank my supervisors Peter Boswijk, Sander Barendse and Paolo Gorgi for the idea, continuous guidelines and helpful discussions. I also thank participants at the UvA Econometrics internal seminar, participants of the 3rd International Econometrics PhD Conference, the Econometric Institute (EI) at Erasmus University Rotterdam, the Virtual Workshop for Junior Researchers in Time Series, IAAE 2024 in Thessaloniki and Xiamen, and the 12th SIdE Workshop for PhD students, in particular my discussant Massimiliano Caporin, for helpful comments and suggestions. I am very grateful to Peter Reinhard Hansen for his discussion at Pre-Conference for Young Scholars at 17th Annual SoFiE Conference.

1. INTRODUCTION

Volatility forecasting is central in asset allocation, risk management, and pricing of derivatives across various financial institutions. Particularly relevant to financial decisions are volatility forecasts of cumulative (multi-period) returns across various time horizons, ranging from one week to several months, depending on the application and the asset class. To ensure consistency in assessing these forecasts, financial institutions typically maintain their decision-making processes around a baseline model evaluated using a specified ex-ante loss function by a decision-maker. A widely used class of models for generating volatility forecasts is the GARCH class, pioneered by Engle (1982) and Bollerslev (1986). Parameters of these models are commonly estimated using the Gaussian quasi-maximum likelihood (QML) method. The resulting parameter estimates are subsequently used to generate one-period-ahead forecasts, which are iterated forward to achieve the desired horizon of volatility forecasts.

However, this approach of constructing forecasts is sensitive to model misspecification: even minor model misspecification can adversely affect the performance of a volatility model, as highlighted by Andersen et al. (2004) and Sizova (2011). As an example, consider the workhorse GARCH(1,1) model, which imposes mean reversion with a geometric rate on volatility, contradicting the presence of higher persistence in volatility known as long-memory (Baillie et al., 1996; Granger & Ding, 1996; Mandelbrot & Van Ness, 1968). In this case, the iteration of one-period-ahead GARCH(1,1) forecasts using the dynamically misspecified functional form implied by mean reversion may result in significant inaccuracies in forecasts, especially when considering large forecasting horizons. Since the true model is generally unknown, model misspecification is practically unavoidable.

In this paper, we propose a parameter estimation method for the GARCH class of volatility models to improve the accuracy of forecasts. To optimize a potentially misspecified model, we propose estimating the parameters of a given model by minimising the same loss function used to evaluate forecasts. In particular, our estimation method focuses on minimizing a quasi-likelihood (QLIKE) loss function specified directly for the volatility of cumulative returns, as the QLIKE function is typically used to assess the accuracy of volatility forecasts (Bauwens et al., 2012). By matching estimation and evaluation objectives, we ensure that our estimation method is consistent with respect to the forecasting objective, as established by Hansen and Dumitrescu (2022). In alignment with its purpose, we refer to our method as the horizon-matched (HM) method, with the resulting estimator also named horizon-matched.

A crucial feature of our approach is that, while we target the cumulative return variance in the estimation, the baseline volatility model remains unchanged, continuing to specify the volatility dynamics of daily returns. Doing that allows us to establish consistency of the proposed estimator for the true parameter under the correct model specification. However, targeting cumulative variance in the estimation increases the estimator's variance, whether or not the model is correctly specified. Building on this, we theoretically examine the forecasting performance of our estimator in a bias-variance trade-off framework. First, we show that when the model is correctly specified, the improvement potential of our estimator over the standard estimator is limited, as the difference in expected forecasting losses is driven by the difference in variances of the estimators. Second, when the model is misspecified, our estimator enhances the model fit by construction. However, the benefits for forecasting depend on ensuring that this improved fit is not offset by increased estimation error. Since volatility models used by financial institutions are typically misspecified, despite this bias-variance trade-off, we believe and show that our estimator is of practical use.

The horizon-matched estimation method applies not only to the standard GARCH class of volatility models, which specify the conditional distribution of returns, but also to GARCH-type models for

the conditional distribution of realised measures of volatility derived from high-frequency intraday data. Inspired by Andersen and Bollerslev (1998), recent research has increasingly popularised models integrating high-frequency information to forecast the volatility of daily or cumulative returns, as it has been established that, under some regularity conditions, realised measures serve as consistent estimators of daily squared volatility and are less noisy than squared daily returns. We demonstrate the application of our method across a range of models, starting with the simplest MEM model introduced by Engle (2002) and Engle and Gallo (2006) and extending to more advanced models such as the HEAVY model by Shephard and Sheppard (2010) and the Realised GARCH model by Hansen et al. (2012). Additionally, our method can be applied to pure-time series models for realised measures, as categorised by Hansen and Lunde (2011), including the widely used HAR model developed by Corsi (2009).

This paper is not the first to propose a method for mitigating the impact of model misspecification on forecasts. A 'direct' approach has been introduced in the forecasting literature with this purpose. This approach involves estimating parameters, defining a model for multi-period variables of interest (e.g., cumulative returns), and constructing forecasts directly. This contrasts sharply with the standard 'iterated' approach, where parameters from a baseline model for daily variables are estimated, and then forecasts are iterated until the horizon of interest is reached. Extensive comparisons between these two forecasting methodologies have been conducted for linear models, leading to the conclusion that direct forecasts are more robust to model misspecification than their iterated counterparts (Marcellino et al., 2006; Pesaran et al., 2011). This robustness arises from the direct minimisation of the in-sample loss function aligned (in terms of horizon) with the evaluation loss function. However, studies comparing these approaches theoretically and empirically for non-linear models (such as GARCH) are limited. Existing studies predominantly favour the iterated approach, whether the focus is on pure volatility forecasting or forecasting of risk measures (De Nicolò & Lucchetta, 2017; Ghysels et al., 2019; Kole et al., 2017; Mancini & Trojani, 2011). For a thorough overview of the literature on this topic, we refer the reader to the work by Ruiz and Nieto (2023).

Our approach integrates elements of the 'iterated' and 'direct' methods. While we retain the baseline volatility model for daily data, similar to the 'iterated' approach, we also target the evaluation loss function in the estimation step, similar to the 'direct' approach. This alignment of estimation and evaluation loss functions can enhance forecast accuracy, as indicated by Granger (1969), Granger (1993), Patton (2020) and Hansen and Dumitrescu (2022). As recommended by Gneiting (2011), once the evaluation loss is established, a forecaster is free to optimise a potentially misspecified baseline model, with one strategy being to target this loss during the parameter estimation step.

The work most closely related to ours is perhaps by Oh and Patton (2024), which, like our approach, does not alter the baseline model but 'tilts' its parameters to improve forecasts. The authors motivate this by arguing that financial institutions build their decision-making process around established statistical models because adopting newer models may be costly and challenging due to various barriers, at least in the short term. In contrast to our approach, they propose a non-parametric approach as an alternative estimation to standard parametric methods (e.g., QML), leveraging information from a state variable correlated with model misspecification. Their study theoretically explores the potential benefits of this non-parametric approach, highlighting the standard bias-variance trade-off in forecasting. Ultimately, our paper and theirs arrive at a similar conclusion: the new estimation method adds variance to forecasts to reduce the bias associated with a misspecified model.

In addition to our main contribution to improving forecasts, as a second contribution, we propose a misspecification test comparing our HM estimator and the QML estimator based on a testing framework similar to that of Hausman (1978). Indeed, both estimators are consistent for the true parameter

under the null of correct specification yet have distinct asymptotic variances. Under the alternative of misspecification, the estimators converge in probability to different pseudo-true values, making our test sensitive to misspecification. To derive a test statistic, we derive the joint asymptotic distribution of the two estimators. Importantly, the overlapping observations in the estimation loss function used for our HM estimator introduce autocorrelation in the score contributions. We propose to use two types of heteroskedasticity-and-autocorrelation-consistent (HAC) variance-covariance estimators to correct for this: a nonparametric estimator by Newey and West (1987), and a parametric estimator by West (1997), which exploits an exact order of moving-average structure of the score contributions, and thus, has to be derived for each baseline model separately.

We perform a Monte Carlo study to analyse the size and power properties of a misspecification test and evaluate the forecasting performance of our HM and the QML estimators under various scenarios – both when the model is correctly specified and when it is misspecified. We also analyse the finite-sample behaviour of the estimator under correct specification in more detail. In this study, we focus on misspecification related to long memory, a concept initially explored by Mandelbrot and Van Ness (1968), Granger and Ding (1996), Baillie et al. (1996). As a baseline model correctly specified under the null hypothesis, we consider a short-memory MEM-GARCH(1,1) model by Engle (2002) and Engle and Gallo (2006). Under the alternative of misspecification, we consider the LMGARCH(1,d,1) model developed by Karanasos et al. (2004) as a data-generating process. For power analysis, we consider values of the memory parameter d between 0 and 0.45, with higher values of d indicating stronger long memory and greater misspecification of a baseline model. We consider a range of estimation horizons for our estimator for the entire Monte Carlo study, specifically between 3 and 66 periods.

We show that the test size is reasonably accurate at the 5% significance level, varying in accuracy depending on the estimation horizon of our estimator. In the power analysis, the misspecification test achieves the highest asymptotic and size-corrected power at the smallest estimation horizon, with power increasing with the value of d. Interestingly, this occurs despite the difference between the two estimators being smallest at this horizon. In the forecasting exercise, we consider forecasting horizons ranging from 5 to 66, allowing the HM estimator to be based on the forecasting horizon, or on a smaller horizon. When the model is correctly specified, we observe minimal out-of-sample forecasting differences between the two estimators, with the difference increasing monotonically with the estimation horizon. Conversely, under misspecification, our estimator yields more accurate out-of-sample forecasts starting from a moderate degree of misspecification, mainly when the estimation horizon is smaller than the forecasting one. This finding highlights the practical implications of the bias-variance trade-off.

In our forecasting exercise, we use ten real time series, specifically returns and realised measures from the paper by Gorgi et al. (2019), to assess the forecasting accuracy across competing GARCH-type models and parameter estimation methods. The horizons of interest span from weekly ($h = 5$) to quarterly ($h = 66$), with the estimation horizon for the HM estimator allowed to be smaller. In the first part of our analysis, we conduct stylised exercises using the standard GARCH(1,1) model and the MEM-GARCH(1,1) model. Our findings indicate that, as the forecasting horizon increases, the gap between the 'best' estimation horizon for the HM estimation and the forecasting horizon increases. Additionally, for the standard GARCH(1,1) model, the 'best' estimation horizon is shorter for a given forecasting horizon than for the MEM-GARCH(1,1). In the second part of our analysis, we investigate how different levels of model misspecification impact forecasting accuracy across horizons. For this purpose, we introduce a component GARCH (cGARCH) model developed by Ding and Granger (1996) and Engle and Lee (1999). We find that the amplitude of gains from the HM estimator is smaller for the cGARCH model compared to the GARCH(1,1), suggesting that our estimator is more effective for underparameterized

models, which are more likely to be misspecified.

The rest of the paper is organised as follows. Section 2 introduces the horizon-matched estimator and examines the bias-variance trade-off in forecasting. Section 3 demonstrates how to construct the horizon-matched estimator for different volatility models that specify the dynamics of both returns and realised measures. Section 4 presents a misspecification test and derives the asymptotic joint distribution for the horizon-matched and the QML estimators. Section 5 conducts a Monte Carlo study to analyse size and power properties of the test, as well as the forecasting accuracy of the two estimators under conditions of correct and incorrect model specification. Section 6 shifts focus to the forecasting exercise. Finally, Section 7 concludes.

2. HORIZON-MATCHED ESTIMATION AND FORECASTING

2.1. Fixing ideas

Let r_t denote a daily (single-period) log-return on an asset of interest generated between times $t-1$ to t, i.e., $r_t = 100 \ln(p_t/p_{t-1})$, where p_t is the price at time t. We denote by $\mathcal{F}_{t-1}$ the information set at the end of time $t-1$ and let it contain the set of past returns $\{r_{t-j}, j \geq 1\}$, and potentially include past observations of other variables measured at a higher frequency than r_t, such as intra-daily returns. We consider the data-generating process (DGP) for the return series to be of the form

$$r_t = \sqrt{g_t} z_t, \quad g_t = g(\mathcal{F}_{t-1}), \tag{1}$$

where the innovation z_t is i.i.d. sequence with $\mathbb{E}[z_t] = 0$ and $\mathbb{E}[z_t^2] = 1$, and $g_t \equiv \text{var}[r_t|\mathcal{F}_{t-1}]$ is the true volatility process known at time $t-1$. Note that the process above assumes $\mathbb{E}[r_t|\mathcal{F}_{t-1}] = 0$, which can be ensured by demeaning the return series.

We are interested in the conditional volatility prediction of the cumulative (multi-period) return over some future horizon h. Using the properties of log-returns, we define the cumulative h-day return as $\tilde{r}_{t,h} = \sum_{j=0}^{h-1} r_{t+j}$, which corresponds to a buy-and-hold strategy of buying an asset at time t and selling it at time $t+h-1$. Under the above assumption of zero conditional mean, daily returns exhibit zero serial correlation. Consequently, the volatility of the cumulative return $\tilde{r}_{t,h}$ conditional on $\mathcal{F}_{t-1}$, denoted by $\tilde{g}_{t,h}$, is given by

$$\tilde{g}_{t,h} \equiv \text{var}[\tilde{r}_{t,h}|\mathcal{F}_{t-1}] = \mathbb{E}[\tilde{r}_{t,h}^2|\mathcal{F}_{t-1}] = \sum_{j=0}^{h-1} g_{t+j|t-1}, \tag{2}$$

where $g_{t+j|t-1} \equiv \text{var}[r_{t+j}|\mathcal{F}_{t-1}]$ for $j = 0, \ldots, h-1$ with $g_{t|t-1} = g_t$.

2.2. Evaluating volatility forecasts of cumulative returns

To construct volatility forecasts cumulative h-day return, we specify a parametric model for the volatility process, denoted by $\sigma_t^2(\theta)$, where θ represents the model parameter vector. Once the forecast is constructed, it is evaluated using a loss function pre-specified by the decision-maker (Gneiting, 2011). In volatility modelling, one of the most common loss functions, in addition to the MSE, is the QLIKE (Bauwens et al., 2012). Both of these loss functions are 'consistent', meaning that their expected values are minimised at the true conditional volatility of cumulative return. However, unlike the MSE, the QLIKE loss depends on the ratio of the target variable to the forecast, making it less sensitive to extreme observations in the sample. Moreover, the QLIKE loss imposes a higher penalty on the under-prediction of volatility, which is of greater economic importance.

Given the volatility model $\sigma_t^2(\theta)$, the QLIKE loss for the conditional volatility of cumulative return $r_{t,h}$ is defined as

$$\text{QLIKE}(\tilde{r}_{t,h}^2, \tilde{\sigma}_{t,h}^2) = \ln \tilde{\sigma}_{t,h}^2(\theta) + \frac{\tilde{r}_{t,h}^2}{\tilde{\sigma}_{t,h}^2(\theta)}, \tag{3}$$

where $\tilde{\sigma}_{t,h}^2(\theta)$ denotes the conditional volatility of the cumulative return implied from a parametric model, and $\tilde{r}_{t,h}^2 = \sum_{j=0}^{h-1} r_{t+j}^2$ denotes the sum of squared returns, providing a less noisy measure of the targeted $\tilde{\sigma}_{t,h}^2(\theta)$ compared to the square of the sum of daily returns, $(r_t + \ldots + r_{t+h-1})^2$, which includes the cross-products between the returns. Given the definition of the loss in (3), the forecasting objective is then to minimise expected value of this loss - an estimable quantity using out-of-sample data:

$$\min_{\theta \in \Theta} \mathbb{E}[\text{QLIKE}(\tilde{r}_{t,h}^2, \tilde{\sigma}_{t,h}^2(\theta))|\mathcal{F}_{t-1}], \tag{4}$$

where $\Theta \subseteq \mathbb{R}^k$, where k denotes the dimension of the parameter vector θ.

Given this forecasting objective, it is natural to estimate the parameter vector θ of a given volatility model by minimising the in-sample loss defined in (3). In a less-than-ideal forecasting environment, such as one affected by model misspecification, aligning estimation and evaluation objective functions is one way of optimising a potentially misspecified model for the forecast objective. It has been shown that this alignment improves the forecasting performance for a given model, as demonstrated by Granger (1969), Granger (1993), Patton (2020) and Hansen and Dumitrescu (2022)[1]. In general, this improvement arises from favourably tilting the misspecificaiton bias of the parameter estimates to perform best at the horizon of interest (Patton, 2020).

2.3. Matching estimation and evaluation loss functions

Assume that sample runs from $t = 1, \ldots, T$ yielding T observations for estimation. At each time point t we can construct the conditional variance of the cumulative h-day return conditional on the past information set $\mathcal{F}_{t-1}$. The forward nature of the expression for $\tilde{\sigma}_{t,h}^2$ allows to construct only $T - h + 1$ cumulative h-day returns and their respective conditional variances. An estimator we propose is defined as

$$\hat{\theta}_T^c = \arg\min_{\theta \in \Theta} \frac{1}{T-h+1} \sum_{t=1}^{T-h+1} \text{QLIKE}(\tilde{r}_{t,h}^2, \tilde{\sigma}_{t,h}^2(\theta)), \tag{5}$$

where the superscript 'c' denotes 'cumulative', indicating that the estimation objective is based on 'cumulative' variables, such as the cumulative h-day return and its respective conditional volatility. We refer to $\hat{\theta}_T^c$ as the horizon-matched (HM) estimator.

When the baseline volatility model $\sigma_t^2(\theta)$ is correctly specified and point identified, that is there is the unique point θ_0 in the parameter space such that $\sigma_t^2(\theta_0) = g_t$ a.s. $\forall\, t$, it can be shown that

$$\mathbb{E}[s_t^c(\theta_0)|\mathcal{F}_{t-1}] = 0, \quad s_t^c(\theta) = (\tilde{\sigma}_{t,h}^2(\theta) - \tilde{r}_{t,h}^2) \frac{1}{\tilde{\sigma}_{t,h}^4(\theta)} \frac{\partial \tilde{\sigma}_{t,h}^2(\theta)}{\partial \theta}, \tag{6}$$

where $s_t^c(\theta_0)$ denotes the score evaluated at the true parameter vector θ_0, and the zero conditional mean equation holds since $\mathbb{E}[\tilde{r}_{t,h}^2|\mathcal{F}_{t-1}] = \tilde{\sigma}_{t,h}^2(\theta_0)$ under correct model specification, specifically when first two conditional moments are correctly specified. Despite the zero conditional mean assumption on the score, the sequence $\{s_{t-1}^c(\theta_0), \mathcal{F}_{t-1}\}$ is not a martingale difference sequence (mds). This is because $s_{t-1}^c(\theta_0)$ is not $\mathcal{F}_{t-1}$-measurable: the score depends on returns after time point $t-1$, except when $h = 1$. As a

[1]Hansen and Dumitrescu (2022) show that in some applications using different loss functions for estimation and evaluation may offer advantages, but only if the two loss functions are 'coherent' – meaning the parameter probability limits under them are the same (Def.1, p.538).

result, $s_t^c(\theta_0)$ is serially correlated. In Section 4 we show how to account for this time-series dependence in the score. Since, by the law of iterated expectations (LIE), $\mathbb{E}[s_t^c(\theta_0)] = 0$, under some regularity conditions[2] $\hat{\theta}_T^c$ converges at a rate $\sqrt{T}$ to a well-defined and unique probability limit θ^{c*}, and has a Normal asymptotic distribution:

$$\begin{aligned} \theta^{c*} &\equiv \underset{\theta\in\Theta}{\arg\min}\, \mathbb{E}[\text{QLIKE}(\tilde{r}_{t,h}^2, \tilde{\sigma}_{t,h}^2(\theta))] \\ \sqrt{T}(\hat{\theta}_T^c - \theta^{c*}) &\xrightarrow{d} \mathcal{N}\left(0, H_c^{-1} C_{cc} H_c^{-1}\right), \end{aligned} \tag{7}$$

where θ^{c*} is the minimiser of the population objective function, which also serves as the forecast evaluation function.

In contrast to (5), the traditional approach to estimating the parameter vector relies on the estimation objective function that differs from the forecasting objective of interest. This method assumes that the innovation z_t follows a standard Gaussian cdf, $z_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, a condition that can be challenging to justify in many empirical applications. Under this distributional assumption, the maximum likelihood estimation is interpreted as a quasi-maximum likelihood (QML) estimation. In this case, the model's parameter vector θ is estimated by maximising the Gaussian quasi-likelihood function. This, in turn, is equivalent to the minimization of the in-sample QLIKE loss function, which leads to the QML estimator defined as

$$\hat{\theta}_T^d = \underset{\theta\in\Theta}{\arg\min}\, \frac{1}{T}\sum_{t=1}^{T} \text{QLIKE}(r_t^2, \sigma_t^2(\theta)), \quad \text{QLIKE}(r_t^2, \sigma_t^2(\theta)) = \ln \sigma_t^2(\theta) + \frac{r_t^2}{\sigma_t^2(\theta)}, \tag{8}$$

where we use the superscript 'd' to denote 'daily', indicating that the estimation[3] objective function is based on 'daily' variables, such as the daily squared return r_t^2 and its respective conditional volatility.

In contrast to the HM estimator, it can be shown that the score $s_t^d(\theta)$ evaluated at the true parameter vector θ_0 is an mds with respect to $\mathcal{F}_{t-1}$, that is

$$\mathbb{E}[s_t^d(\theta_0)|\mathcal{F}_{t-1}] = 0, \quad s_t^d(\theta) = (\sigma_t^2(\theta) - r_t^2)\frac{1}{\sigma_t^4(\theta)}\frac{\partial \sigma_t^2(\theta)}{\partial \theta}, \tag{9}$$

and $s_{t-1}^d(\theta_0)$ is $\mathcal{F}_{t-1}$-measurable. Due to this mds feature, the score $s_t^d(\theta_0)$ is not autocorrelated. Hence, by an application of quasi-likelihood theory under standard regularity conditions the QML estimator $\hat{\theta}_T^d$ converges at a rate $\sqrt{T}$ to a well-defined and unique probability limit θ^{d*}

$$\begin{aligned} \theta^{d*} &\equiv \underset{\theta\in\Theta}{\arg\min}\, \mathbb{E}[\text{QLIKE}(r_t^2, \sigma_t^2(\theta))] \\ \sqrt{T}(\hat{\theta}_T^d - \theta^{d*}) &\xrightarrow{d} \mathcal{N}\left(0, H_d^{-1} C_{dd} H_d^{-1}\right), \end{aligned} \tag{10}$$

where θ^{d*} is the minimiser of the population objective function different from the one of forecasting interest. By the point-identification assumption, we then know that $\theta^{c*} = \theta^{d*}$, which is equal to θ_0. That is, when the model is correctly specified, the HM estimator is also consistent for the true parameter vector but at the expense of the larger asymptotic variance-covariance matrix induced by overlapping observations in the estimation objective loss function.

When the baseline volatility model $\sigma_t^2(\theta)$ is misspecified, it is possible to obtain the consistency results if we assume that θ^{c*} is a unique minimiser of the population objective function $\mathbb{E}[\text{QLIKE}(\tilde{r}_{t,h}^2, \tilde{\sigma}_{t,h}^2(\theta))]$,

[2] We may need to apply the central limit theorem (CLT) for near epoch dependent (NED) sequences.

[3] Patton (2011) defines the QLIKE loss function as $\text{QLIKE}(r_t^2, \sigma_t^2(\theta)) = \frac{r_t^2}{\sigma_t^2(\theta)} - \ln\frac{r_t^2}{\sigma_t^2(\theta)} - 1$, which is the QLIKE loss function in (8) up to additive and multiplicative constants. Furthermore, the definition of the loss by Patton (2011) is designed to yield the value of zero when the forecast matches the proxy's value.

and that θ^{d*} is a unique minimiser of the population objective function $\mathbb{E}[\text{QLIKE}(r_t^2, \sigma_t^2(\theta))]$. In this case, θ^{c*} and θ^{c*} can be interpreted as the pseudo-true parameters. Regarding the statistical inference, to obtain the asymptotic normality under misspecification we require that there is a CLT that can be applied to the scores. For instance, a CLT for near epoch dependent (NED) sequences. To summarise, when the model is misspecified, θ^{c*} is the pseudo-true parameter of interest because it minimises in population the objective function that a forecaster uses for model evaluation purposes.

2.4. Forecasting bias-variance trade-off

In the previous subsection we have shown that the HM estimator $\hat{\theta}_T^c$ minimises the forecast evaluation objective function for the baseline volatility model $\sigma_t^2(\theta)$ in large samples, even under misspecification. Yet, in small samples, or in scenarios where the estimation sample does not increase with the sample size, the same conclusion for the HM estimator cannot always be drawn due to a bias-variance trade-off.

Using the population parameter definitions in (7) and (10), the unconditional out-of-sample loss evaluated at the population parameter θ^{c*} is weakly smaller than the loss evaluated at any other parameter:

$$\mathbb{E}[\text{L}(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta^{c*}))] \leq \mathbb{E}[\text{L}(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta))] \quad \forall\ \theta \in \Theta. \tag{11}$$

Evaluating the right-hand side at the QML population parameter θ^{d*}, we obtain:

$$\mathbb{E}[\text{L}(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta^{c*}))] \leq \mathbb{E}[\text{L}(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta^{d*}))], \tag{12}$$

which demonstrates that the out-of-sample average loss evaluated at the HM population parameter is weakly smaller than that evaluated at the QML population parameter.

Although 12 provides an ordering of the expected out-of-sample loss at the pseudo-true values, we are typically interested in the ordering of the expected loss at the estimated parameter values. By application of a second-order Taylor expansion we obtain the following decomposition

$$\begin{aligned}
&\mathbb{E}[L(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\hat{\theta}_T^c))] - \mathbb{E}[L(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\hat{\theta}_T^d))] \\
&\approx \underbrace{\mathbb{E}[L(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta^{c*}))] - \mathbb{E}[L(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta^{d*}))]}_{\leq 0,\ \text{loss difference at pseudo-true values}} \\
&\quad + \underbrace{\mathbb{E}\left[\frac{\partial L(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta^{c*}))}{\partial \theta}(\hat{\theta}_T^c - \theta^{c*})\right] - \mathbb{E}\left[\frac{\partial L(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta^{d*}))}{\partial \theta}(\hat{\theta}_T^d - \theta^{d*})\right]}_{\text{bias term}} \\
&\quad + \underbrace{\mathbb{E}\left[\frac{1}{2}\frac{\partial^2 L(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta^{c*}))}{\partial \theta^2}(\hat{\theta}_T^c - \theta^{c*})^2\right] - \mathbb{E}\left[\frac{1}{2}\frac{\partial^2 [L(\tilde{r}_{T+1,h}^2, \tilde{\sigma}_{T+1,h}^2(\theta^{d*}))}{\partial \theta^2}(\hat{\theta}_T^d - \theta^{d*})^2\right]}_{\text{variance term}},
\end{aligned} \tag{13}$$

where, for the sake of brevity, we denote the QLIKE loss by L. Under correct specification, it follows that the loss difference at the true values and the bias term take value zero, with the latter following by the conditional mean zero assumption on the scores evaluated at the true values (see Subsection 2.3). Hence, the difference in expected losses is then driven by the variance of the estimators. Under misspecification, both the bias and variance terms are expected to be smaller for larger estimation samples, since the estimators will converge to their pseudo-true values. In particular, for the bias term the magnitude of the estimation error is $O_p(1/\sqrt{T})$, while for the variance term it is $O_p(1/T)$, indicating that the estimation error in the variance term is of smaller magnitude than that in the bias term. Hence, in large samples under misspecification the loss difference is predominantly determined by the loss difference evaluated at the pseudo-true values.

The decomposition in (13) shows that under correct specification, when the estimation sample T is small, the variance term in (13) may become positive due to the inherently larger variance of the HM estimator, leading to worse expected out-of-sample performance of the horizon-based estimator $\hat{\theta}_T^c$ compared to the QML estimator $\hat{\theta}_T^d$. However, in the presence of misspecification, the HM estimator, despite being less biased, where the bias is measured as the distance to the population parameter θ^{c*} obtained under the relevant evaluation loss function, still may have larger variance than the QML estimator. If the reduction in bias is outweighed by the increase in variance, the HM estimator may generate higher expected loss. Consequently, it is possible that an efficient (low variance) but biased QML estimator could outperform the less efficient but unbiased HM estimator in terms of forecasting accuracy, particularly in finite samples.

3. MODEL EXAMPLES

3.1. Standard GARCH models

We demonstrate how the horizon-matched estimation method works for specific models. As an example, we consider the GARCH(1,1), which is proven hard to beat against a variety of alternatives (see Hansen and Lunde (2005)). The GARCH(1,1) model for the conditional variance of asset return r_t is given by:

$$\begin{aligned} r_t &= \sigma_t z_t, \\ \sigma_t^2(\theta) &= \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{14}$$

where $\theta = (\omega, \alpha, \beta)'$. To implement the horizon-matched estimation method, as shown in (5), we need to find an expression for $\tilde{\sigma}_{t,h}^2(\theta)$. Given the model, one can show that the conditional variance of the cumulative return $\tilde{r}_{t,h}$ conditional on $\mathcal{F}_{t-1}$ is equal to

$$\tilde{\sigma}_{t,h}^2(\theta) = \sum_{j=0}^{h-1} \sigma_{t+j|t-1}^2(\theta), \tag{15}$$

where $\sigma_{t+j|t-1}^2(\theta)$ denotes the conditional variance of a daily return at time point $t+j$ given information set $\mathcal{F}_{t-1}$. It is a well-known expression that the variance of a daily return for the GARCH(1,1) model is equal to

$$\sigma_{t+j|t-1}^2(\theta) = \overline{\sigma}^2 + (\alpha+\beta)^j(\sigma_{t|t-1}^2 - \overline{\sigma}^2), \quad \overline{\sigma}^2 = \frac{\omega}{1-\alpha-\beta}, \quad j = 1, \ldots, h-1, \tag{16}$$

where $\sigma_{t|t-1}^2$ follows in a closed-form from the model.

We now show how to implement our method for a more flexible GARCH model, such as the component GARCH (cGARCH) model introduced by Ding and Granger (1996) and Engle and Lee (1999). The cGARCH model decomposes the conditional variance into the time-varying long-run (permanent) and short-run (transitory) components. The cGARCH model is defined as

$$\begin{aligned} \sigma_t^2(\theta) &= q_t + s_t \\ s_t &= \alpha(r_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}) \\ q_t &= \omega + \rho q_{t-1} + \varphi(r_{t-1}^2 - \sigma_{t-1}^2), \end{aligned} \tag{17}$$

where q_t denotes the long-run component, s_t denotes the short-run component, and $\theta = (\omega, \alpha, \beta, \rho, \varphi)'$. One can rewrite the short-run component in the autoregressive form, $s_t = (\alpha+\beta)s_{t-1} + \alpha(r_{t-1}^2 - \sigma_{t-1}^2)$, which shows that the sum $\alpha+\beta$ determines the persistence rate of s_t, while ρ does so for q_t. In parallel

to the GARCH(1,1) model, we derive an analytical expression for $\sigma^2_{t+j|t-1}(\theta)$ which now consists of the two components:

$$\begin{aligned} \sigma^2_{t+j|t-1}(\theta) &= q_{t+j|t-1} + s_{t+j|t-1} \\ s_{t+j|t-1} &= (\alpha+\beta)^j s_{t|t-1}, \\ q_{t+j|t-1} = \overline{q} + \rho^j(q_{t|t-1} - \overline{q}), \quad \overline{q} &= \frac{\omega}{1-\rho}, \quad j = 1, \dots, h-1. \end{aligned} \tag{18}$$

3.2. GARCH models augmented with high-frequency information

We demonstrate the versatility of the proposed horizon-matched estimation approach by extending its application to GARCH models with realised measures. As an example, we focus on 'parallel GARCH' models, including the MEM model by Engle (2002) and Engle and Gallo (2006), and the HEAVY model by Shephard and Sheppard (2010). These models are classified as 'parallel GARCH' because their primary goal is to infer the conditional variance of returns with the help of improved measurements, such as realised measures. Also, included in this category are models with the measurement equation for the realised measure, such as the Realized GARCH model by Hansen et al. (2012).

We begin with the MEM model. For simplicity, we assume that σ^2_t has a conditionally unbiased estimator, denoted by x_t, which represents the realised measure. By conditional unbiasedness we mean that $\mathbb{E}[x_t|\mathcal{F}_{t-1}] = \sigma^2_t$. The MEM model for the realised measure with the standard GARCH(1,1) specification for σ^2_t is defined as

$$\begin{aligned} x_t = \sigma^2_t u_t, \quad u_t &\overset{\text{i.i.d.}}{\sim} D^+(1, \sigma^2) \\ \sigma^2_t(\theta) &= \omega + \alpha x_{t-1} + \beta \sigma^2_{t-1}, \end{aligned} \tag{19}$$

where $\mathcal{F}_t = \sigma\{x_s : s \le t\}$ and u_t is an iid component with non-negative support and unit mean assumption, $\mathbb{E}[u_t|\mathcal{F}_{t-1}] = 1$, such that σ^2_t is identified as the conditional mean of the realised measure. A possible distribution for u_t is a Gamma distribution $\Gamma(a, b)$ with a shape parameter a and a scale parameter b, where the unit mean assumption imposes a restriction on the scale, that is $b = 1/a$.

Estimating the parameters of the MEM model is aligned with optimization of the QLIKE function. Assuming that $a = 1$, which corresponds to the exponential distribution, the log-likelihood function (not quasi), after omitting constants and rescaling, is given by (8), where r^2_t is substituted with the realised measure x_t as a proxy for the conditional variance. As shown by Engle and Gallo (2006), exactly the same estimation objective can be used for other types of Gamma distribution with $a \neq 1$, and the estimator obtained from this objective's minimisation can be interpreted as the QML.

To implement the horizon-matched estimator, we define $\tilde{\sigma}^2_{t,h}(\theta)$ as

$$\tilde{\sigma}^2_{t,h}(\theta) = \mathbb{E}[\tilde{x}_{t,h}|\mathcal{F}_{t-1}] = \sum_{j=0}^{h-1} \mathbb{E}\left[x_{t+j}|\mathcal{F}_{t-1}\right]. \tag{20}$$

While $\mathbb{E}[x_t|\mathcal{F}_{t-1}] = \sigma^2_{t|t-1}$ follows from the model, to find expressions for $\mathbb{E}[x_{t+j}|\mathcal{F}_{t-1}]$ for $t \ge 1$ we apply the LIE:

$$\mathbb{E}[x_{t+j}|\mathcal{F}_{t-1}] = \mathbb{E}[\mathbb{E}[x_{t+j}|\mathcal{F}_{t+j-1}]|\mathcal{F}_{t-1}] = \mathbb{E}[\sigma^2_{t+j}|\mathcal{F}_{t-1}], \quad \text{for} \quad j = 1, \dots, h-1, \tag{21}$$

which, combined with the given model's specification, implies that

$$\mathbb{E}[x_{t+j}|\mathcal{F}_{t-1}] = \omega + \alpha\mathbb{E}[x_{t+j-1}|\mathcal{F}_{t-1}] + \beta\mathbb{E}[\sigma^2_{t+j-1}|\mathcal{F}_{t-1}], \tag{22}$$

yielding the recursive relation

$$\mathbb{E}[x_{t+j}|\mathcal{F}_{t-1}] = \omega + (\alpha+\beta)\mathbb{E}[x_{t+j-1}|\mathcal{F}_{t-1}]. \tag{23}$$

Finally, we use the backward substitution in (23) to express it via the $\sigma^2_{t|t-1}$ and the long-run unconditional mean $\overline{\sigma}^2 = \omega/(1-\alpha-\beta)$, assuming that the process for σ^2_t is covariance-stationary, leading to

$$\mathbb{E}[x_{t+j}|\mathcal{F}_{t-1}] = \overline{\sigma}^2 + (\alpha+\beta)^j(\sigma^2_{t|t-1} - \overline{\sigma}^2), \quad \text{for} \quad j = 1,\dots,h-1, \tag{24}$$

which is a parallel expression to (16) for a standard GARCH(1,1) model. This shows that for the class of MEM models, the expression for $\tilde{\sigma}^2_{t,h}$ resembles that of the standard GARCH. Hence, if we now specify the MEM model with the cGARCH structure for volatility, the resulting expression for $\tilde{\sigma}^2_{t,h}$ will again mirror that in (18).

As another example of a 'parallel GARCH' model, we consider the HEAVY model:

$$\mu_t(\theta^{RM}) = \omega^{RM} + \alpha^{RM} RM_{t-1} + \beta^{RM}\mu_{t-1}, \quad \mu_t = \mathbb{E}[RM_t|\mathcal{F}_{t-1}] \tag{25}$$

$$\sigma^2_t(\theta^R) = \omega^R + \alpha^R RM_{t-1} + \beta^R\sigma^2_{t-1}, \quad \sigma^2_t = \text{var}[r_t|\mathcal{F}_{t-1}] \tag{26}$$

where $\mathcal{F}_{t-1}$ combines information from both daily returns and realised measures, and $\theta^R = (\omega^R, \alpha^R, \beta^R)'$ and $\theta^{RM} = (\omega^{RM}, \alpha^{RM}, \beta^{RM})'$. We note that (26) is identical to the dynamic equation of the MEM model with GARCH(1,1) structure. However, in contrast to the MEM model, Shephard and Sheppard (2010) assume that (26) models close-to-close conditional variance, while (25) models conditional expectation of open-to-close variance, which implies that the realised measure is likely to be a downward biased measure (due to overnight effects) of the squared return, that is $\mathbb{E}[RM_t|\mathcal{F}_{t-1}] \leq \text{var}[r_t|\mathcal{F}_{t-1}]$. Since the return equation in (26) contains the exogenous variable (realised measure), which is not an unbiased estimator for the conditional variance of interest, the specification of the dynamics for the realised measure is necessary to produce multi-day ahead conditional variance forecasts, and thus construct the variance of interest $\tilde{\sigma}^2_{t,h}$. Expression for $\sigma^2_{t+j|t-1}$ and $\mu_{t+j|t-1}$ are shown in eq.(11) in Shephard and Sheppard (2010).

It is important to note that one can also apply the horizon-matched estimation approach to a pure time-series model for the realised measure, such as the popular HAR model by Corsi (2009). However, in this case, to infer the variance of returns, an assumption about the realised measure is required. Typically, the assumption is that the realised measure is an unbiased estimator of the conditional variance – similar to what is done in the MEM model. Alternatively, as seen in the HEAVY model, if the realised measure is not unbiased, the dynamics of its conditional mean should be incorporated.

4. MODEL MISSPECIFICATION TEST

4.1. Theoretical framework

In this section we derive the misspecification test that is based on a similar null hypothesis as the Hausman test (Hausman, 1978). Under the null hypothesis of correct specification, the estimators $\hat{\theta}^c_T$ and $\hat{\theta}^d_T$ are both consistent estimators for the true value θ_0, but have distinct asymptotic distributions. Under the alternative hypothesis of misspecification, each estimator will converge in probability to different psuedo-true values.

Formally, Hausman (1978) considers the following testing situation:

$$\begin{aligned} \mathbb{H}_0 &: \text{plim}(\hat{\theta}^d_T - \hat{\theta}^c_T) = 0 \\ \mathbb{H}_1 &: \text{plim}(\hat{\theta}^d_T - \hat{\theta}^c_T) \neq 0. \end{aligned} \tag{27}$$

The Hausman test statistic is then simply a quadratic form around the contrast between the two estimators and is given by

$$\hat{H}_T = T(\hat{\theta}^d_T - \hat{\theta}^c_T)'\hat{\Omega}^-_T(\hat{\theta}^d_T - \hat{\theta}^c_T), \tag{28}$$

where $\Omega := \text{Avar}(\sqrt{T}(\hat{\theta}_T^d - \hat{\theta}_T^c))$ and $\hat{\Omega}_T^-$ is a consistent estimator of some generalised inverse of Ω, denoted by Ω^-. Assuming $\hat{\Omega}_T^- \xrightarrow{p} \Omega^-$, under $\mathbb{H}_0$

$$\hat{H}_T \xrightarrow{d} \chi^2_{\text{rk}(\Omega)}, \tag{29}$$

where for a given square matrix A, $\text{rk}(A)$ denotes its rank. Under $\mathbb{H}_1$, the test statistic diverges, $\hat{H}_T \xrightarrow{p} \infty$. These two conditions guarantee that the misspecification test has the asymptotic size control under the null, and is consistent under the alternative.

The question is how to construct $\hat{\Omega}_T^-$ to ensure that the test statistic converges in distribution. When Ω is full rank, then by the application of the Continuous Mapping Theorem a valid choice of $\hat{\Omega}_T^-$ is given by the inverse of $\hat{\Omega}_T$ assuming that $\hat{\Omega}_T \xrightarrow{p} \Omega$. If the covariance matrix is rank deficient, then the Moore-Penrose inverse of $\hat{\Omega}_T$, denoted as $\hat{\Omega}_T^+$, converges to the Moore-Penrose inverse of Ω, that is $\hat{\Omega}_T^+ \xrightarrow{p} \Omega^+$, iff $\hat{\Omega}_T$ has rank converging in probability to the one of Ω, see Theorem 2 of Andrews (1987). In our application, we do not face the problem of reduced rank, and thus use the standard inverse to construct the test statistic.

4.2. Asymptotic joint distribution

The estimators $\hat{\theta}_T^d$ and $\hat{\theta}_T^c$ are obtained by the first-order conditions, respectively:

$$\sum_{t=1}^{T} s_t^d(\hat{\theta}_T^d) = 0 \quad \text{and} \quad \sum_{t=1}^{T-h+1} s_t^c(\hat{\theta}_T^c) = 0, \tag{30}$$

where $s_t^d(\hat{\theta}_T^d)$ and $s_t^c(\hat{\theta}_T^c)$ are the scores evaluated at the estimators. To obtain the joint asymptotic distribution of the two estimators under correct specification, we firstly define $s_t^c(\hat{\theta}_T^c) = 0$ for $t = T-h+2, \ldots, T$, which are the missing observations. Then, we expand the stacked score vectors from (30) in a Taylor series around the true parameter θ_0. Under several conditions, it can be shown that

$$\sqrt{T}\begin{pmatrix} \hat{\theta}_T^d - \theta_0 \\ \hat{\theta}_T^c - \theta_0 \end{pmatrix} = -\begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} \frac{\partial s_t^d(\theta_0)}{\partial \theta'} & 0_{k\times k} \\ 0_{k\times k} & \frac{1}{T}\sum_{t=1}^{T} \frac{\partial s_t^c(\theta_0)}{\partial \theta'} \end{bmatrix}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} s_t^d(\theta_0) \\ s_t^c(\theta_0) \end{pmatrix} + o_p(1), \tag{31}$$

from where the asymptotic joint distribution is found.

As shown in Subsection 2.3, the scores evaluated at the true parameter value are such that $\mathbb{E}[s_t^d(\theta_0)] = \mathbb{E}[s_t^c(\theta_0)] = 0$. With that result, it can be shown that the stacked scores evaluated at θ_0 obey the central limit theorem (CLT):

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} s_t^d(\theta_0) \\ s_t^c(\theta_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0_{2k\times 1}, C\right), \quad C = \begin{bmatrix} C_{dd} & C_{dc} \\ C_{cd} & C_{cc} \end{bmatrix}, \tag{32}$$

where

$$C = \lim_{T\to\infty} \mathbb{E}\left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} s_t^d(\theta_0) \\ s_t^c(\theta_0) \end{pmatrix} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} s_t^d(\theta_0) \\ s_t^c(\theta_0) \end{pmatrix} \right)' \right] \tag{33}$$

with

$$
\begin{aligned}
C_{dd} &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[s_t^d(\theta_0)s_t^d(\theta_0)'] \\
C_{dc} &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[s_t^d(\theta_0)s_t^c(\theta_0)'\right] + \frac{1}{T}\sum_{j=1}^{h-1}\sum_{t=j+1}^{T} \mathbb{E}[s_t^d(\theta_0)s_{t-j}^c(\theta_0)'] \\
C_{cc} &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[s_t^c(\theta_0)s_t^c(\theta_0)'] \\
&\quad + \frac{1}{T}\sum_{j=1}^{h-1}\sum_{t=j+1}^{T} \mathbb{E}[s_t^c(\theta_0)s_{t-j}^c(\theta_0)'] + \frac{1}{T}\sum_{j=-1}^{-h+1}\sum_{t=-j+1}^{T} \mathbb{E}[s_{t+j}^c(\theta_0)s_t^c(\theta_0)'] \\
C_{cd} &= C_{dc}',
\end{aligned}
\tag{34}
$$

where the limiting variance of the 'daily' score, C_{dd}, accounts only for the outer-product because $\{s_t^d(\theta_0), \mathcal{F}_t\}$ forms a martingale difference sequence. The 'cumulative' score, on the contrary, forms a moving average process of order $h-1$, and thus we have to account for non-zero autocovariances: $E[s_t^c(\theta_0)s_{t-j}^c(\theta_0)'] \neq 0$ for $j = -h+1, \ldots, 0, \ldots, h-1$.

Together with (32), by the application of the Continuous Mapping Theorem and Slutsky's Lemma (31) leads to the joint asymptotic distribution of the two estimators:

$$
\begin{aligned}
\sqrt{T}\begin{pmatrix} \hat{\theta}_T^d - \theta_0 \\ \hat{\theta}_T^c - \theta_0 \end{pmatrix} &\xrightarrow{d} \mathcal{N}\left(0, H^{-1}CH^{-1}\right), \\
H = \begin{bmatrix} H_d & 0_{k\times k} \\ 0_{k\times k} & H_c \end{bmatrix} &= \begin{bmatrix} \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{\partial s_t^d(\theta_0)}{\partial\theta'}\right] & 0_{k\times k} \\ 0_{k\times k} & \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{\partial s_t^c(\theta_0)}{\partial\theta'}\right] \end{bmatrix}.
\end{aligned}
\tag{35}
$$

4.3. Variance-covariance matrix estimator

The estimator for the asymptotic variance-covariance matrix of $\hat{\theta}_T = (\hat{\theta}_T^{d\prime}, \hat{\theta}_T^{c\prime})'$ is given by

$$
\widehat{\mathrm{Avar}}(\hat{\theta}_T) = \frac{1}{T}\hat{H}^{-1}\hat{C}\hat{H}^{-1} = \frac{1}{T}\begin{bmatrix} \hat{H}_d^{-1}\hat{C}_{dd}\hat{H}_d^{-1} & \hat{H}_d^{-1}\hat{C}_{dc}\hat{H}_c^{-1} \\ \hat{H}_c^{-1}\hat{C}_{cd}\hat{H}_d^{-1} & \hat{H}_c^{-1}\hat{C}_{cc}\hat{H}_c^{-1} \end{bmatrix},
\tag{36}
$$

with the estimator for the Hessian defined as

$$
\hat{H} = \begin{bmatrix} \hat{H}_d & 0_{k\times k} \\ 0_{k\times k} & \hat{H}_c \end{bmatrix}, \quad \hat{H}_d = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial s_t^d(\hat{\theta}_T^d)}{\partial\theta'}\right], \quad \hat{H}_c = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial s_t^c(\hat{\theta}_T^c)}{\partial\theta'}\right],
\tag{37}
$$

where $\hat{H}_d$ and $\hat{H}_c$ are the sample counterparts of H_d and H_c from (35), respectively. As for the estimator for C, we propose to use two heteroskedasticity-and-autocorrelation-consistent (HAC)-type estimators, which we discuss below. To obtain the test statistic, from $\widehat{\mathrm{Avar}}(\hat{\theta}_T)$ we obtain the estimator for the asymptotic variance-covariance matrix of the contrast:

$$
\begin{aligned}
\hat{\Sigma}_T &:= \widehat{\mathrm{Avar}}(\hat{\theta}_T^d - \hat{\theta}_T^c) = S\widehat{\mathrm{Avar}}(\hat{\theta}_T)S' \\
&= \frac{1}{T}[\hat{H}_d^{-1}\hat{C}_{dd}\hat{H}_d^{-1} + \hat{H}_c^{-1}\hat{C}_{cc}\hat{H}_c^{-1} - \hat{H}_d^{-1}\hat{C}_{dc}\hat{H}_c^{-1} - \hat{H}_c^{-1}\hat{C}_{cd}\hat{H}_d^{-1}],
\end{aligned}
\tag{38}
$$

where $S = (I_k, -I_k)$ is a differencing matrix and I_k is the identity matrix of dimension k. Finally, the test statistic is constructed as

$$
\hat{H}_T = (\hat{\theta}_T^d - \hat{\theta}_T^c)'\hat{\Sigma}_T^{-}(\hat{\theta}_T^d - \hat{\theta}_T^c),
\tag{39}
$$

where $\hat{\Sigma}_T^{-}$ is the inverse of $\hat{\Sigma}_T$. Below we show how to obtain the estimator for C.

4.3.1. Newey-West estimator

We firstly introduce a non-parametric estimator proposed by Newey and West (1987). Let $s_t(\hat{\theta}_T) = \left(s_t^d(\hat{\theta}_T^d)', s_t^c(\hat{\theta}_T^c)'\right)'$ for $t = 1, \ldots, T$. Then, the estimator for C is defined as

$$\hat{C} = \sum_{j=-l+1}^{l-1} \omega_j \hat{\Gamma}_j, \quad \hat{\Gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} s_t(\hat{\theta}_T) s_{t-j}(\hat{\theta}_T)', \quad j = 0, \ldots, l-1, \tag{40}$$

with $\hat{\Gamma}_j = \hat{\Gamma}'_{-j}$ for $j = -l+1, \ldots, -1$. In terms of the autocovariance matrices $\hat{\Gamma}_j$ for $j = 1, \ldots, l-1$, $\hat{C}$ in (40) can be written as

$$\hat{C} = \hat{\Gamma}_0 + \sum_{j=1}^{l-1} \omega_j (\hat{\Gamma}_j + \hat{\Gamma}'_j). \tag{41}$$

The weights ω_j can be taken from any symmetric around 0 kernel function with the normalisation $\omega_0 = 1$. We opt for the Bartlett kernel proposed by Newey and West (1987). In this case, the weights are tent-shaped weights and defined as $\omega_j = 1 - j/l$ for $j = 1, \ldots, l-1$. As we can see, the weights decrease linearly with the step size of $1/l$, hence, we may inflate the lag structure by picking the lag length l larger than h to reduce the bias in the estimator. Overall, the downside of using this estimator is that, first, the choice of l is determined empirically; and second, we inevitably account for autocovariances up to lag $l-1$ of the 'daily' score $s_t^d(\hat{\theta}_T^d)$, which should be zero in theory at the true value.

4.3.2. West estimator

The second estimator we consider is a parametric estimator proposed by West (1997), which exploits the known structure of the dependence in the score vectors. Consider the 'cumulative' score, which can be decomposed into components, one of which is drives the dependence in the score:

$$\begin{aligned} s_t^c(\theta) &= (\tilde{r}_{t,h}^2 - \tilde{\sigma}_{t,h}^2) \frac{1}{\tilde{\sigma}_{t,h}^4} \frac{\partial \tilde{\sigma}_{t,h}^2}{\partial \theta} = z_t u_t, \\ \text{with} \quad z_t &= \frac{1}{\tilde{\sigma}_{t,h}^4} \frac{\partial \tilde{\sigma}_{t,h}^2}{\partial \theta} \quad \text{and} \quad u_t = \tilde{r}_{t,h}^2 - \tilde{\sigma}_{t,h}^2. \end{aligned} \tag{42}$$

Given that u_t is covariance stationary (due to the covariance-stationarity of σ_t^2), and moreover, satisfies $\mathbb{E}[u_t u_{t+j}] = 0$ for $j \geq h$, it follows that u_t can be represented as an MA$(h-1)$ process:

$$u_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \ldots + \psi_{h-1} \varepsilon_{t-h+1}, \tag{43}$$

where ε_t is white noise, but conditionally heteroscedastic like u_t.

West (1997) shows that C_{cc} can be expressed as

$$C_{cc} = \sum_{j=-h+1}^{h-1} \mathbb{E}[s_t^c(\theta) s_{t+j}^{c\prime}(\theta)] = \mathbb{E}[d_t d_t'], \quad d_t = \varepsilon_t \sum_{j=0}^{h-1} \psi_j z_{t+j}, \tag{44}$$

with $\varepsilon_t = r_{t+h-1}^2 - \sigma_{t+h-1}^2$ (see Appendix A5 for details). Moreover, we show that for the GARCH(1,1) model the coefficients ψ_j in (44) can be expressed in terms of the model's parameters:

$$\psi_0 = 1, \quad \psi_j = 1 + \alpha \frac{1 - (\alpha + \beta)^j}{1 - \alpha - \beta}, \quad j = 1, \ldots, h-1. \tag{45}$$

Note that expressing the coefficients ψ_j in terms of the model parameters can become more challenging for more complex models.

Consider now the 'daily' score $s_t^d(\theta)$. Using the definition of ε_t from above, the score can be decomposed into

$$s_t^d(\theta) = (r_t^2 - \sigma_t^2)\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta} = \varepsilon_{t-h+1} x_t, \quad \text{with} \quad x_t = \frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \theta}. \tag{46}$$

Using this decomposition, C_{dd} can be expressed as

$$C_{dd} = \mathbb{E}[\varepsilon_t^2 x_{t+h-1} x'_{t+h-1}]. \tag{47}$$

As shown in Appendix A5, C_{cd} can be then expressed as

$$C_{cd} = \sum_{j=0}^{h-1} \mathbb{E}[s_t^c s_{t+j}^{d\prime}] = \mathbb{E}\left[\varepsilon_t^2 \left(\sum_{j=0}^{h-1} \psi_j z_{t+j}\right) x'_{t+h-1}\right], \tag{48}$$

where by defining $z^*_{t+h-1} = \sum_{j=0}^{h-1} \psi_j z_{t+j}$, we obtain

$$C = \begin{bmatrix} C_{dd} & C_{dc} \\ C_{cd} & C_{cc} \end{bmatrix} = \mathbb{E}\left[\varepsilon_t^2 \begin{pmatrix} x_{t+h-1} \\ z^*_{t+h-1} \end{pmatrix} \begin{pmatrix} x_{t+h-1} \\ z^*_{t+h-1} \end{pmatrix}'\right]. \tag{49}$$

The natural estimator for C is its sample counterpart:

$$\hat{C}_T = \frac{1}{T}\sum_{t=1}^{T} (r_t^2 - \hat{\sigma}_t^2)^2 \begin{pmatrix} \hat{x}_t \\ \hat{z}^*_t \end{pmatrix} \begin{pmatrix} \hat{x}_t \\ \hat{z}^*_t \end{pmatrix}', \tag{50}$$

which shows that by construction the estimator is guaranteed to be positive (semi-)definite because it involves the summation over outer-products.

Overall, we acknowledge that West (1997) may be impractical due to its analytical complexities, which may arise for more complex models rather than the GARCH(1,1). That is, the estimator has to be derived for each model separately and depends on the model's structure. Alternatively, a more practical option is the estimator by Newey and West (1987).

5. MONTE CARLO STUDY

The type of misspecification takes many forms and is dependent on the true DGP and the model under consideration (e.g., see White (1996)). In this section, we focus on dynamic misspecification, referring to the incorrect specification of the dynamics in conditional volatility. As an example, we consider misspecification with respect to long memory, a well-documented empirical characteristic of volatility (Baillie et al., 1996; Granger, 1969). We consider a short-memory MEM-GARCH(1,1) model both when it is correctly specified and when it is misspecified. First, we present the size results of a test together with the Monte Carlo properties of both the QML and the HM estimators. Second, we illustrate the power of the test in detecting a specific form of misspecification with respect to long memory. Additionally, we assess the forecasting performance of the QML and the HM estimators when the MEM-GARCH(1,1) model is both correctly specified and misspecified.

5.1. Size properties of a misspecification test

To assess the size properties of a misspecification test, we consider as a data generating process (DGP) under the null hypothesis of correct specification the LMGARCH(1,d,1) model developed by Karanasos et al. (2004) with the restriction $d = 0$. Under this restriction, the original long-memory model simplifies

to the short-memory GARCH(1,1) model. The LMGARCH(1,d,1) for the realised measure x_t is defined as

$$\begin{aligned} x_t &= \sigma_t^2 u_t, \quad u_t \overset{\text{i.i.d}}{\sim} \Gamma(a, 1/a) \\ (1-\phi L)(1-L)^d(x_t - \overline{\sigma}^2) &= (1-\beta L)(x_t - \sigma_t^2), \end{aligned} \tag{51}$$

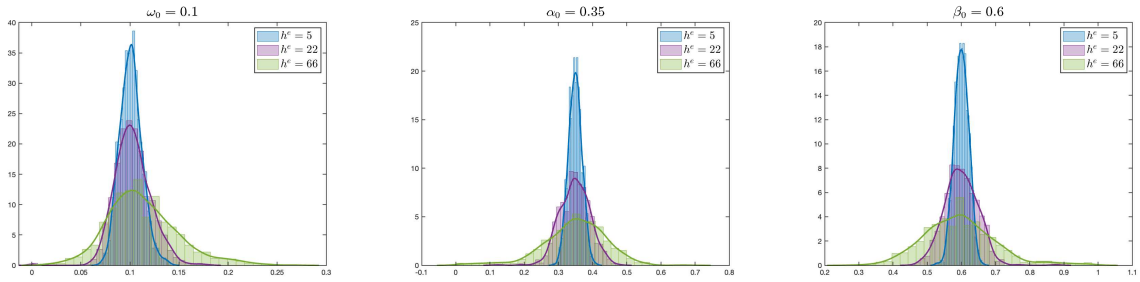
where $\mathbb{E}[x_{t+1}] = \overline{\sigma}^2$ and $(1-L)^d$ is a fractional difference operator with a memory parameter d. Unlike the well-known FIGARCH(1,d,1) model by Baillie et al. (1996), the LMGARCH(1,d,1) is specified in terms of deviations from the unconditional mean. This minor modification aligns it with the ARFIMA(1,d,1) model for the mean, and, consequently, ensures covariance stationarity for the series modelled.

We simulate the realised measure series through the ARCH(∞) representation for σ_t^2:

$$\sigma_{t+1}^2 = \overline{\sigma}^2 + \Psi(L)(x_{t+1} - \overline{\sigma}^2), \quad \Psi(L) = \sum_{i=1}^{\infty} \psi_i L^i, \quad \Psi(L) = 1 - \frac{(1-L)^d(1-\phi L)}{(1-\beta L)}, \tag{52}$$

with $d = 0$, $\beta = 0.6$, $\phi = 0.95$, implying essentially a short-memory MEM-GARCH(1,1) model with $\beta = 0.6$ and $\alpha = \phi - \beta = 0.35$. Shape parameter a in gamma distribution is set to $a = 2$ and $\overline{\sigma}^2 = 2$. All parameter values for the simulated series are consistent with the parameter estimates obtained from the model estimated on real dataset. We simulate $N = 1000$ series of length $5 \times T$, discarding the initial $4 \times T$ observations as a burn-in period to avoid a long lasting effect of the initialisation, which is particularly useful in cases when $d \neq 0$ simulated in the next subsection for power analysis. This disregard of data results in a final length of $T = 5000$ observations.

Figure 1. Pdf-Normalised Histograms of Estimated Parameters for the HM Estimator.



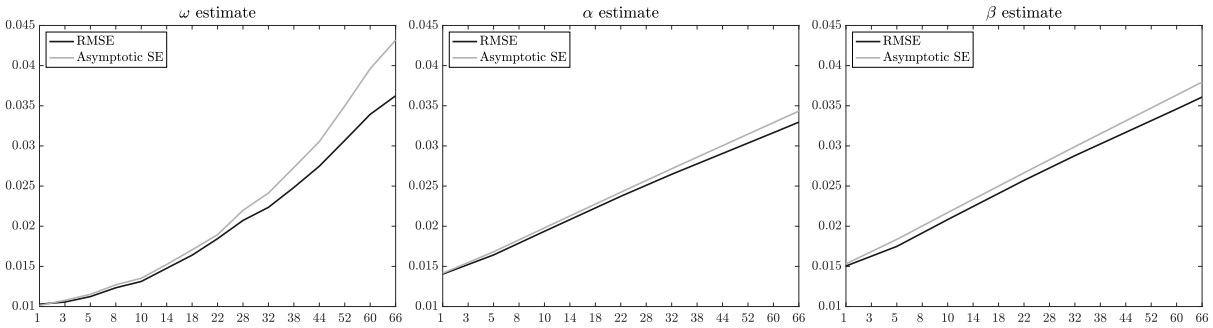
Notes: Histograms with the pdf normalisation of estimated parameters for the MEM-GARCH(1,1). Each figure corresponds to a different parameter, with three histograms per figure representing different estimation horizons h^e in the HM estimator: $h^e \in \{5, 22, 66\}$. True parameter values are equal to $\omega_0 = 0.1, \alpha_0 = 0.35, \beta_0 = 0.6$.

Figure 1 illustrates the sampling distributions of the HM estimator for three estimation horizons: $h^e = 5$ (small), $h^e = 22$ (medium) and $h^e = 66$ (large). We choose these estimation horizons to highlight the differences in sampling distributions when the horizon varies from small to large. As expected under correct specification, the sampling distributions are centered around the true parameter value for each horizon, with slight biases arising for larger estimation horizons. However, these finite sample biases are expected to vanish as the sample size increases. For smaller estimation horizons, such as $h^e = 5$, the sampling distributions are narrower with a sharper peak at the true parameter value, indicating smaller variance, that is more precise estimates. In contrast, for larger estimation horizons, such as $h^e = 66$, the sampling distribution is flatter around the true parameter value, indicating higher variance, that is less accurate estimates.

In Figure 2, we analyse in more detail the variance of the sampling distribution for the QML and HM estimators. The goal is to assess in finite samples the quality of the asymptotic approximation for

the variance of the estimators. For the HM estimator, we consider estimation horizons beyond those shown in Figure 1. To address our goal, we compare the root mean square error of parameter estimates with the corresponding asymptotic standard error, which is independently computed using the West (1997) variance estimator for a MEM-GARCH(1,1) model as discussed in Section 4. Figure 2 reveals two key findings. First, the variance of the HM estimator is consistently higher than that of the QML estimator and increases with the estimation horizon. Second, as the estimation horizon increases, the asymptotic standard error tends to be systematically higher than the root mean square error, indicating a decline in the accuracy of finite sample approximations due to the true variance of the estimator being overestimated by the West (1997) estimator.

Figure 2. Asymptotic SE and RMSE for the QML and HM Estimators.



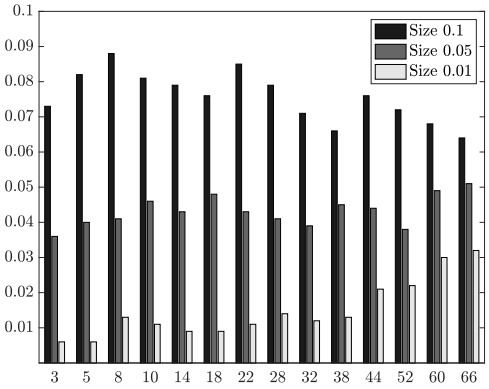
Notes: Figures compare asymptotic standard error (Asymptotic SE) and root mean square error (RMSE) for the QML and HM estimation methods. Each figure represents a separate parameter in the MEM-GARCH(1,1) model. The horizontal axis corresponds to the estimation horizon h^e. In particular, for the QML estimator $h^e = 1$, while for the HM estimator we consider $h^e \in \{3, 5, 8, 10, 14, 18, 22, 28, 32, 38, 44, 52, 60, 66\}$. For the purpose of illustration, we consider definitions of the RMSE and Asymptotic SE for ω: RMSE is $\left[\frac{1}{N}\sum_{i=1}^{N}(\hat{\omega}_T^{(i)} - \overline{\omega}_T)^2\right]^{1/2}$, where $\overline{\omega}_T$ is Monte Carlo empirical mean across $N = 1000$ simulations, and Asymptotic SE is $\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{\sqrt{T}}\left[\text{Avar}\left(\sqrt{T}(\hat{\omega}_T^{(i)} - \omega_0)\right)\right]^{1/2}\right]$, where ω_0 is the true parameter value.

Having discussed the Monte Carlo behaviour of the estimators, we present the empirical size results for the misspecification test in Figure 3. In general, the test is reasonably accurate in correctly not rejecting the null hypothesis of correct specification. In particular, the empirical sizes demonstrate a slight wave pattern as the estimation horizon increases. At the 1% significance level, the rejection frequency tends to increase with the estimation horizon, while at the 10% significance level, tends to decrease making the test more conservative. In contrast, the more accurate size results are observed at the 5% significance level.

Table 1 presents the out-of-sample mean QLIKE loss difference across Monte Carlo simulations when the MEM-GARCH(1,1) model is correctly specified. For each simulation, we use first $T = 2500$ observations to estimate parameters, followed by the rolling window forecasting for the horizons $h \in \{5, 10, 22, 44, 66\}$, with a re-estimation window of 50 observations (approximately two months). We estimate parameters using both the QML and HM estimation methods, where for the latter we allow the estimation horizon to be smaller than the forecasting horizon. As shown in the Table, the mean losses are quite comparable across different forecasting horizons, which is in line with the null hypothesis of correct specification that both estimators are consistent for the true value. However, as the estimation horizon for the HM estimator increases, the mean difference also increases, with a negative sign indicating worse performance for the HM estimator. Indeed, according to the decomposition in (13), for a correctly specified model only the variance term affects the mean difference in out-of-sample losses. Since the

estimation sample size is fixed and relatively large in this controlled forecasting exercise, the impact of the variance term on differences in mean QLIKE loss is small but increases with the estimation horizon.

Figure 3. Empirical Size of a Misspecification Test.



Notes: Rejection frequencies of the null hypothesis of correct specification for a misspecification test. The model under the null is the MEM-GARCH(1,1) model. The model used in the estimation is the MEM-GARCH(1,1) as well. Size results are shown for different estimation horizons $h^e \in \{3, 5, 8, 10, 14, 18, 22, 28, 32, 38, 44, 52, 60, 66\}$ (horizontal axis) used to compute the HM estimator, and for significance levels $\{0.1, 0.05, 0.01\}$. Results are based on the West (1997) variance-covariance estimator.

Table 1. Forecasting Results for a Correctly Specified Model.

	$h^e = 3$	$h^e = 5$	$h^e = 8$	$h^e = 10$	$h^e = 14$	$h^e = 18$	$h^e = 22$	$h^e = 28$	$h^e = 32$	$h^e = 38$	$h^e = 44$	$h^e = 52$	$h^e = 60$	$h^e = 66$
h = 5	-0.0001	-0.0001												
h = 10	-0.0001	-0.0002	-0.0003	-0.0004										
h = 22	-0.0001	-0.0002	-0.0005	-0.0006	-0.0009	-0.0012	-0.0016							
h = 44	-0.0001	-0.0003	-0.0006	-0.0008	-0.0011	-0.0014	-0.0018	-0.0022	-0.0027	-0.0031	-0.0037			
h = 66	-0.0002	-0.0003	-0.0008	-0.0009	-0.0012	-0.0015	-0.0018	-0.0022	-0.0025	-0.0028	-0.0032	-0.0038	-0.0042	-0.0044

Notes: Out-of-sample mean QLIKE loss difference across Monte Carlo simulations for the correctly specified MEM-GARCH(1,1) model, considering forecasting horizons $h \in \{5, 10, 22, 44, 66\}$. Mean loss is constructed as the loss from the QML estimator minus the loss from the HM estimator. The estimation horizon for the HM estimator corresponds to $h^e \in \{3, 5, 8, 10, 14, 18, 22, 28, 32, 38, 44, 52, 60, 66\}$. For a given forecasting horizon, the estimation horizon considered is $h^e \leq h$.

5.2. Power properties of a misspecification test

For the power analysis, we consider nine realistic DGPs by varying the 'degree of memory' under the alternative hypothesis of misspecification. The parameter values for these DGPs are obtained by fitting the realised measure for the IBM individual stock to the Gaussian FIGARCH in the parameterisation of Chung (1999)[4]. Estimation is done in OxMetrics using the G@RCH package[5]. We fix the memory parameter d at various levels, $d \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$, and estimate the remaining parameters. This produces a grid of values $\beta \in \{0.65, 0.69, 0.73, 0.77, 0.8, 0.84, 0.87, 0.89, 0.92\}$ and $\phi = 0.95$[6]. As for the Gamma distribution shape parameter a, it is not obtained directly due to the Gaussian

[4]Both Chung (1999) and Karanasos et al. (2004) apply the fractional differencing operator to the demeaned process, as was originally suggested in the ARFIMA model for the mean process.

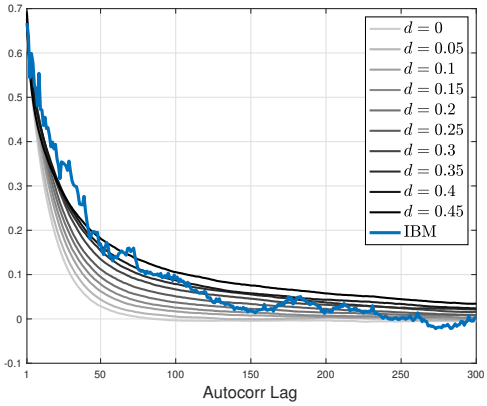
[5]Package link: https://sites.google.com/iae-aix.com/slaurent/grch?authuser=0.

[6]We smoothed and adjusted estimated parameters to ensure that the realised measure is non-negative for all simulations. However, not all alternative DGPs satisfy necessary and sufficient conditions from Conrad and Haag, 2006 to guarantee non-negative conditional variance and, consequently, realised measure.

errors in the estimated model. Instead, we compute the variance of the standardised error, x_t/σ_t^2, which is approximately 0.5 for all the DGPs, corresponding to $a = 2$ in $\Gamma(a, 1/a)$. Therefore, a is fixed at this level for all DGPs. Interestingly, while most estimated parameters show modest variation with d, parameter β, on the contrary, increases monotonically. The length of the simulated series and the number of simulations are the same as those in the size analysis ($N = 1000, T = 5000$).

Figure 4 illustrates mean sample autocorrelations for realised measure series across ten different DGPs (including the one for the size). First, the first-order autocorrelation structure appears to be similar across simulated DGPs. Discrepancies become more pronounced at larger lags, with the highest autocorrelation values across lags characterising the DGP with $d = 0.45$. This illustrates that the autocorrelation function of order j, denoted by ρ_j, for the LMGARCH model follows $\rho_j \sim j^{2d-1}$, implying that the autoregressive parameter β primarily influences only first-order autocorrelations. Second, the simulated DGP with $d = 0.45$ closely resembles real data (in particular, up to 100 lags) which we take to be the realised kernel for IBM stock. This is not surprising given that the parameter values used to generate data under this DGP correspond to parameter estimates obtained from modeling IBM stock using the LMGARCH model. Third, the figure indicates that the MEM-GARCH(1,1) model corresponding to the DGP with $d = 0$ might be too restrictive for the realised measure series because its autocorrelation function decays too quickly to zero, indicating that the MEM-GARCH(1,1) model is dynamically misspecified.

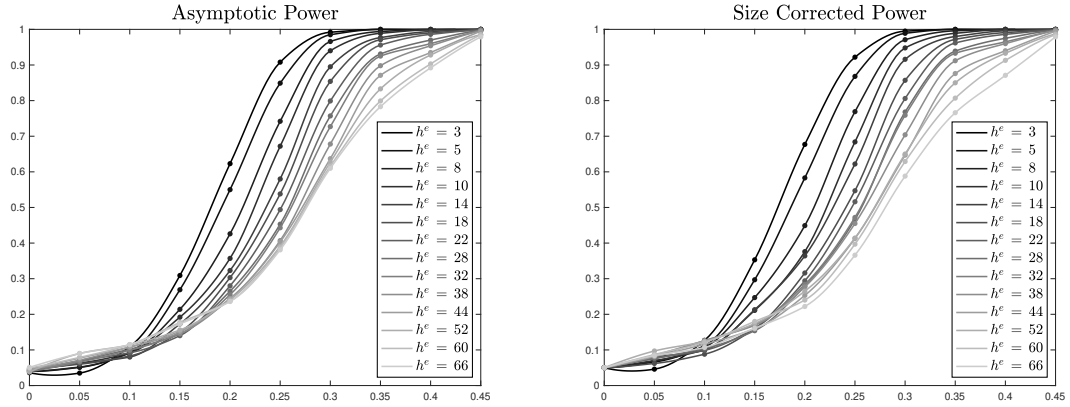
Figure 4. Mean Autocorrelation Functions of Simulated Data for Power Analysis.



Notes: Mean autocorrelation functions across Monte Carlo simulations for simulated realised measures are plotted for the DGP with $d = 0$ (size analysis) and nine alternative DGPs with $d \neq 0$ (power analysis). In addition, the autocorrelation function for the realised measure of IBM stock is included.

Figure 5 reveals that the power of the misspecification test rises with an increase in the memory parameter d, demonstrating consistency of the test. As d increases, indicating larger deviations from the null hypothesis of correct specification ($d = 0$), so does the power of the test across all values of the estimation horizon h^e in the HM estimator. Interestingly, the highest power is observed for $h^e = 3$, while the lowest power is observed for $h^e = 66$. We explain this observation as follows: although a smaller difference between $\hat{\theta}_T^d$ and $\hat{\theta}_T^c$ is expected for smaller values of h^e compared to larger values of h^e for a given d, elements of the variance-covariance matrix estimate may be smaller for smaller values of h^e, thereby amplifying the test statistic, and thus the power. Table 2 indeed shows that the Euclidean distance between the two estimators increases both with an increment in d for a fixed value of h^e and with an increment in h^e for a fixed value of d.

Figure 5. Empirical Power of Misspecification Test.



Notes: Rejection frequencies of the null hypothesis of correct specification for a misspecification test at 5% significance level. Results are shown for ten DGPs from the LMGARCH(1,d,1) model for different values of $d \in \{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$, shown in the horizontal axis, and for different values of estimation horizons (h^e) for the HM estimator. Considered estimation horizons are $h^e \in \{3, 5, 8, 10, 14, 18, 22, 28, 32, 38, 44, 52, 60, 66\}$. The model used in the estimation is the MEM-GARCH(1,1). Hence, results corresponding to $d = 0$ are the size results. Results are based on the West (1997) estimator for the variance-covariance matrix of the scaled scores. Left plot corresponds to the actual power as a function of d calculated using asymptotic critical values of the chi-squared distribution. Right plot corresponds to the actual power as a function of d calculated using size-corrected critical values. Size-corrected critical values are computed as the $100 \times (1 - 0.05)$ percentile of the Monte Carlo distribution of the Hausman test statistic under the null hypothesis of correct specification.

Table 2. Mean Euclidean Distance between the QML and the HM estimators.

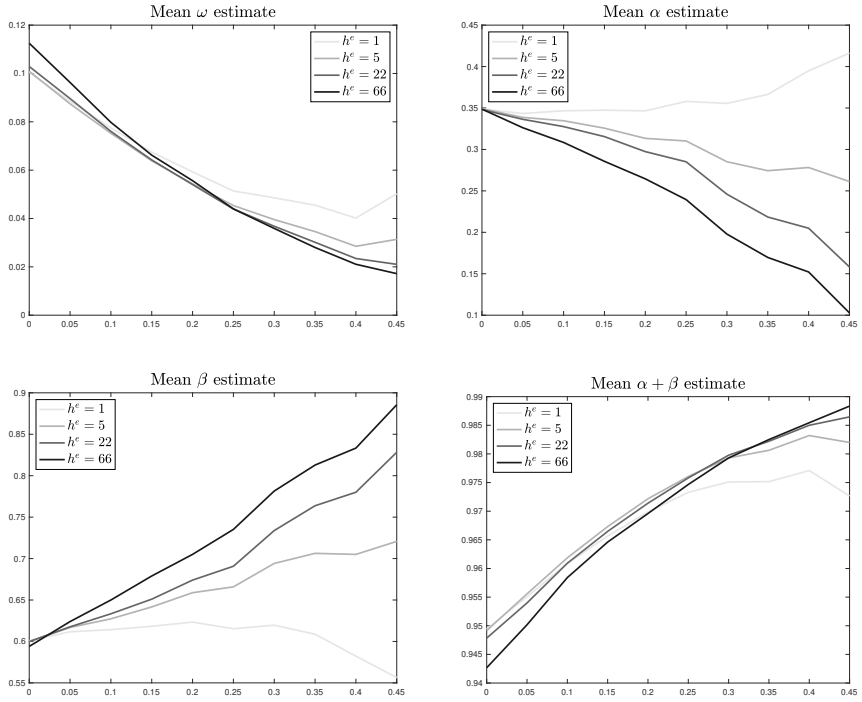
	$h^e = 3$	$h^e = 5$	$h^e = 8$	$h^e = 10$	$h^e = 14$	$h^e = 18$	$h^e = 22$	$h^e = 28$	$h^e = 32$	$h^e = 38$	$h^e = 44$	$h^e = 52$	$h^e = 60$	$h^e = 66$
$d = 0$	0.01	0.02	0.02	0.03	0.04	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.10	0.11
$d = 0.05$	0.01	0.02	0.03	0.03	0.04	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.11
$d = 0.1$	0.02	0.02	0.03	0.03	0.04	0.05	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12
$d = 0.15$	0.02	0.03	0.04	0.04	0.05	0.06	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13
$d = 0.2$	0.04	0.05	0.06	0.06	0.07	0.07	0.08	0.09	0.09	0.10	0.11	0.12	0.13	0.14
$d = 0.25$	0.05	0.07	0.08	0.09	0.10	0.10	0.11	0.12	0.13	0.14	0.15	0.16	0.18	0.19
$d = 0.3$	0.07	0.10	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20	0.21	0.22	0.23
$d = 0.35$	0.10	0.13	0.16	0.18	0.19	0.20	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29
$d = 0.4$	0.12	0.17	0.21	0.23	0.25	0.26	0.28	0.29	0.30	0.31	0.32	0.33	0.34	0.35
$d = 0.45$	0.16	0.23	0.29	0.31	0.34	0.36	0.38	0.40	0.41	0.42	0.43	0.44	0.45	0.46

Notes: Mean Euclidean distance between the QML and the HM estimator for the MEM-GARCH(1,1) model. Results are shown for ten DGPs from the LMGARCH(1,d,1) model for different values of $d \in \{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$ and for different values of estimation horizon (h^e) for the HM estimator. Considered estimation horizons are $h^e \in \{3, 5, 8, 10, 14, 18, 22, 28, 32, 38, 44, 52, 60, 66\}$.

To better understand the difference between the two estimators, Figure (6) plots the mean parameter estimates across Monte Carlo simulations, along with the volatility persistence estimate, measured by the sum of α and β. An interesting observation is that as d increases, the HM estimates of both α and β diverge from the QML ones. In particular, the larger the estimation horizon, the greater the divergence. Specifically, for the HM estimator, the mean estimate of α decreases as d increases, while β rises, which is consistent with $\alpha + \beta$ being bounded by one. In contrast, for the QML estimator, β decreases (non-monotonically), while α increases. Looking at volatility persistence captured by $\alpha + \beta$, the HM estimator consistently shows higher persistence than the QML estimator, though, as expected, the magnitude of the difference is relatively small. For the largest degree of misspecification, the highest

persistence is observed for the HM estimator with the largest estimation horizon.

Figure 6. Mean Parameter Estimates based on the QML and HM Estimators.



Notes: Mean parameter estimates from the QML and HM estimation methods for the MEM-GARCH(1,1) model. Thirst three plots show mean results for separate parameters, while the fourth plot shows mean estimate of volatility persistence ($\alpha + \beta$). Mean estimates are shown for ten DGPs from the LMGARCH(1,d,1) model for different values of $d \in \{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$, shown in the horizontal axis, and for different values of estimation horizons for the HB estimator $h^e = \{5, 22, 66\}$. Estimation horizon $h^e = 1$ corresponds to the QML estimator.

Having analysed the power of the test, we aim to recover forecasts under different degrees of misspecifications to evaluate the effectiveness of the HM estimator for forecasting. Table 3 presents the out-of-sample mean QLIKE loss difference across Monte Carlo simulation when the MEM-GARCH(1,1) model is misspecified. Similar to the forecasting exercise when the model was correctly specified, we use first $T = 2500$ observations to estimate model parameters, followed by the rolling window forecasting for horisons $h \in \{5, 10, 22, 44, 66\}$, with a re-estimation window of 50 observations. We estimate parameters using both the QML and HM estimation methods, where for the latter we allow the estimation horizon to be smaller than the forecasting horizon. There are a couple of key takeaways from this Table. First, when the misspecification degree is small (e.g., $d = 0.05$, $d = 0.1$), the mean loss difference is negative for most forecasting horizons, indicating that the QML estimator outperforms the HM estimator, except for $h = 44$ with $h^e = 3$. This occurs because, for small degree of misspecification, the pseudo-true values are close, resulting in small loss differences evaluated at the pseudo-true values (see (13)). However, the bias and variance terms increase with the estimation horizon, eventually dominating the out-of-sample mean loss difference. Second, as the degree of misspecification grows (larger d), the mean loss difference turns positive, indicating the superior performance of the HM estimator. More importantly, as h increases, it becomes less optimal to match the estimation and forecasting horizons. In particular, for the smallest forecasting horizon ($h = 5$), it is optimal to match horizons starting from $d = 0.3$, which corresponds to the medium degree of misspecification. However, for the largest forecasting horizon ($h = 66$), the optimal estimation horizon increases from $h^e = 8$ only up to $h^e = 32$, which is slightly more than half of

the forecasting horizon of interest. This is because the growing bias and variance terms with an increase in the estimation horizon start to outweigh the benefits of targeting the forecasting horizons.

6. FORECASTING EXERCISE

We analyse the forecasting performance of the two estimators for the GARCH(1,1) and cGARCH models specified for returns and realised measures. We do not attempt to conduct a horserace of volatility models. Rather, we illustrate how our method improves upon the widely-used GARCH models. By considering these two models, we investigate for which type of models (in terms of complexity) our estimator adds more forecasting value. In addition to forecasting, we perform a misspecification test on real data.

6.1. Data

To assess the forecasting performance of the two parameter estimation methods outlined in the sections above we use the data set consisting of 10 individual stocks from the paper by Gorgi et al. (2019). Specifically, these are the components of the Dow Jones Industrial Average index with ticker symbols AA, AXP, BA, CAT, GE, HD, HON, IBM, JPM, and KO. Data are available for ten years, from January 2, 2001, to December 31, 2010 and contain $T = 2515$ trading days for all stocks. The sample contains both low volatility and high volatility periods, where the latter is due to the 'financial crisis'.

Daily individual stock returns are taken as open-to-close. As a realised measure estimator, we have at our disposal the realised kernel of Barndorff-Nielsen et al. (2009). For details on the construction of the realised kernel, we refer the reader to Gorgi et al. (2019). Having open-to-close returns ensures one-to-one correspondence between the conditional variance and the realised kernel that excludes the overnight return.

6.2. In-sample results

Before presenting the forecasting results, we examine the in-sample parameter estimation results for the GARCH(1,1) and the cGARCH models. Table 4 and Table 5 display in-sample parameter estimates for the GARCH(1,1) and the MEM-GARCH(1,1) models, respectively. Similarly, Table 6 and Table 7 show in-sample parameters estimates for the cGARCH and the MEM-cGARCH models, respectively. Parameter estimates are shown for the QML estimation and for the HM estimation method across selected estimation horizons.

Overall, the MEM-GARCH(1,1) model reveals more pronounced parameter differences across estimation horizons. In particular, as the estimation horizon in the HM method increases, the estimate of β increases while α correspondingly decreases. In contrast, the standard GARCH(1,1) model, where β estimate is already near its upper bound for the QML estimation, shows only minor parameter shifts. The cGARCH model similarly shows small differences in parameter estimates, but in the MEM-cGARCH model, β and α change. Initially, β increases and α declines as the estimation horizon in the HM estimator increases, but starting around horizon $h = 22$, β estimate begins to decrease. The parameter ρ, which captures persistence of long-term volatility, remains close to 1 both for the QML and HM estimators.

We also perform a misspecification test for both types of models. Table 8 depicts the number of stocks (out of 10) for which the null hypothesis of correct specification is rejected at the 5% significance level. The test is applied across different horizons for the HM estimator, particularly those used in the forecasting exercise below. Comparing the GARCH(1,1) with the MEM-GARCH(1,1), we find that for the GARCH(1,1), the null hypothesis is rejected for a maximum of three stocks at larger estimation horizons. In contrast, for the MEM-GARCH(1,1) model, the null hypothesis is rejected for all ten stocks

Table 3. Forecasting Results for a Misspecified Model.

	$d = 0.05$	$d = 0.1$	$d = 0.15$	$d = 0.2$	$d = 0.25$	$d = 0.3$	$d = 0.35$	$d = 0.4$	$d = 0.45$
h^e					$h = 5$				
3	-0.05	-0.03	**0.05**	**0.19**	**0.50**	1.09	1.89	2.73	4.38
5	-0.13	-0.10	-0.01	0.12	0.48	**1.15**	**2.06**	**3.01**	**4.99**
					$h = 10$				
3	-0.05	-0.01	**0.07**	**0.28**	0.71	1.67	2.85	4.32	7.17
5	-0.12	-0.08	0.03	0.26	**0.77**	1.99	3.48	5.36	9.20
8	-0.26	-0.23	-0.08	0.14	0.68	**2.01**	**3.61**	**5.67**	10.03
10	-0.36	-0.34	-0.17	0.04	0.59	1.94	3.52	5.62	**10.07**
					$h = 22$				
3	-0.04	0.00	**0.13**	0.45	1.06	2.61	4.40	7.02	11.85
5	-0.11	-0.07	0.11	**0.49**	**1.22**	3.27	5.59	9.03	15.71
8	-0.25	-0.25	0.00	0.42	1.17	3.53	6.11	10.09	18.03
10	-0.36	-0.38	-0.09	0.33	1.11	**3.56**	**6.19**	10.37	18.75
14	-0.62	-0.64	-0.32	0.10	0.84	3.49	6.11	**10.55**	**19.31**
18	-0.92	-0.96	-0.59	-0.20	0.52	3.26	5.88	10.38	19.27
22	-1.28	-1.35	-0.90	-0.55	0.04	2.81	5.43	9.96	18.93
					$h = 44$				
3	-0.04	**0.01**	0.20	0.64	1.48	3.85	6.44	10.62	17.30
5	-0.13	-0.02	**0.21**	**0.78**	**1.73**	4.98	8.44	13.98	23.43
8	-0.26	-0.25	0.14	0.77	1.73	5.61	9.44	16.01	27.48
10	-0.39	-0.39	0.06	0.70	1.70	5.81	9.73	16.67	28.90
14	-0.68	-0.64	-0.17	0.48	1.44	**5.97**	**9.93**	17.44	30.40
18	-0.99	-0.94	-0.41	0.23	1.10	5.94	9.88	**17.63**	31.01
22	-1.32	-1.31	-0.69	-0.06	0.54	5.55	9.53	17.40	31.19
28	-1.74	-2.02	-1.34	-0.71	-0.20	4.87	9.11	16.92	**31.30**
32	-2.13	-2.31	-1.76	-1.09	-0.57	4.61	8.84	16.48	31.08
38	-2.70	-2.87	-2.43	-1.68	-1.30	4.02	8.16	15.75	30.57
44	-3.26	-3.62	-3.34	-2.47	-2.36	3.22	7.28	14.80	29.89
					$h = 66$				
3	-0.06	0.00	0.25	0.68	1.72	4.62	7.60	12.81	20.18
5	-0.18	-0.04	0.32	0.90	2.06	6.16	10.15	17.21	27.79
8	-0.31	-0.28	**0.31**	**0.93**	2.12	7.14	11.55	19.96	33.11
10	-0.44	-0.43	0.26	0.87	**2.13**	7.50	12.02	20.97	35.06
14	-0.75	-0.66	0.07	0.68	1.89	7.90	12.47	22.35	37.29
18	-1.08	-0.93	-0.14	0.49	1.58	**8.03**	**12.59**	**22.95**	38.35
22	-1.37	-1.26	-0.36	0.29	1.01	7.72	12.33	22.91	38.82
28	-1.69	-1.91	-0.92	-0.24	0.25	7.23	12.05	22.67	39.42
32	-2.00	-2.09	-1.28	-0.48	0.01	7.09	11.95	22.35	**39.43**
38	-2.41	-2.58	-1.75	-0.88	-0.47	6.76	11.45	21.95	39.42
44	-2.83	-3.23	-2.47	-1.48	-1.40	6.20	10.71	21.25	39.13
52	-3.43	-4.24	-3.42	-2.45	-2.50	5.44	9.53	20.25	38.57
60	-4.06	-5.18	-4.31	-3.59	-3.66	4.18	8.11	19.08	37.89
66	-4.49	-5.72	-4.93	-4.57	-4.75	3.34	7.17	18.03	37.17

Notes: Out-of-sample mean QLIKE loss difference (scaled by 10^3) across Monte Carlo simulations for the misspecified MEM-GARCH(1,1) model, considering forecasting horizons $h \in \{5, 10, 22, 44, 66\}$. The results are shown for nine DGPs from the LMGARCH(1,1) model for different values of $d \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$. Mean loss is constructed as the loss from the QML estimator minus the loss from the HM estimator. The estimation horizon for the HM estimator corresponds to $h^e \in \{3, 5, 8, 10, 14, 18, 22, 28, 32, 38, 44, 52, 60, 66\}$. For a given forecasting horizon, the estimation horizon considered is $h^e \leq h$. In bold highlighted the largest positive mean loss difference for each DGP and estimation horizon h^e.

Table 4. In-Sample Results for GARCH(1,1) Model.

	$h^e=1$			$h^e=5$			$h^e=10$			$h^e=22$			$h^e=44$			$h^e=66$		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
AA	0.02	0.05	0.95	0.02	0.05	0.95	0.02	0.05	0.95	0.01	0.04	0.95	0.01	0.03	0.97	0.01	0.03	0.97
AXP	0.01	0.08	0.91	0.01	0.07	0.93	0.01	0.07	0.93	0.01	0.05	0.95	0.01	0.04	0.96	0.00	0.04	0.96
BA	0.03	0.06	0.93	0.03	0.06	0.93	0.02	0.05	0.94	0.01	0.04	0.95	0.01	0.04	0.96	0.01	0.04	0.96
CAT	0.04	0.06	0.93	0.04	0.06	0.92	0.02	0.04	0.95	0.01	0.03	0.96	0.01	0.03	0.96	0.02	0.04	0.96
GE	0.01	0.05	0.95	0.00	0.04	0.96	0.00	0.04	0.96	0.00	0.04	0.96	0.00	0.04	0.96	0.00	0.05	0.95
HD	0.03	0.06	0.93	0.03	0.06	0.93	0.03	0.07	0.92	0.02	0.06	0.93	0.02	0.06	0.93	0.01	0.05	0.94
HON	0.05	0.10	0.89	0.04	0.10	0.89	0.01	0.04	0.96	0.01	0.03	0.97	0.01	0.03	0.97	0.01	0.03	0.97
IBM	0.01	0.07	0.92	0.01	0.07	0.92	0.01	0.06	0.94	0.01	0.04	0.95	0.01	0.04	0.95	0.01	0.05	0.95
JPM	0.02	0.11	0.89	0.01	0.08	0.92	0.01	0.07	0.93	0.01	0.06	0.94	0.01	0.05	0.95	0.01	0.05	0.95
KO	0.01	0.06	0.94	0.01	0.06	0.93	0.01	0.07	0.93	0.00	0.05	0.95	0.00	0.04	0.96	0.00	0.04	0.96

Notes: In-sample parameter estimates for the GARCH(1,1) model for 10 individual stocks. Results are shown for the QML estimation method ($h^e = 1$) and for the HM estimation method ($h^e \in \{5, 10, 22, 44, 66\}$).

at the largest horizons. When comparing the two cGARCH models, we observe a similar pattern: the MEM-cGARCH model generally has more rejections across stocks than the standard cGARCH model. However, unlike the GARCH models, the MEM-cGARCH model does not show a monotonic increase in rejections as the estimation horizon grows.

In practice, a forecaster does not need to run a test for all possible estimation horizons for the HM estimator. The results from Table 8 indicate that the test's rejection varies with the estimation horizon. It may be sufficient to test a few estimation horizons to check for the rejection of the null hypothesis of correct specification. If the test consistently rejects the null hypothesis, the forecaster then need to find an appropriate estimation horizon for the HM estimator to construct forecasts. In the following subsection, we explore the optimal estimation horizon for the considered models and quantify the accuracy gains from using the proposed estimator.

6.3. Out-of-sample results

When a forecaster is interested in the forecasting horizon of $h > 1$, there are two options for estimating parameters using the HM estimation approach. A forecaster directly targets the horizon h or estimates parameters at a shorter, intermediate horizon. It is possible that estimating parameters at a shorter horizon yields smaller out-of-sample (OOS) losses due to reduced variance despite an increase in bias in parameters, and thus forecasts. This situation indicates the bias-variance trade-off where a reduction in variance might outweigh the increased bias, leading to an overall smaller out-of-sample loss. Below, we empirically investigate this trade-off by comparing the OOS performance of the HM estimator with respect to the QML estimator for different intermediate estimation horizons.

We consider h-step-ahead forecasts with $h \in \{5, 10, 22, 44, 66\}$, where h corresponds to a forecasting horizon of interest, corresponding to the one or two trading weeks, the one or two trading months, and lastly, for the next quarter, respectively. Let h^e denote the estimation horizon for the HM approach. For each h, we consider a grid of values for h^e, where $h^e \leq h$. For $h^e \in [2, 10]$, we consider a grid with a step size of one; for $h^e \in [12, 44]$, we consider a grid with a step size of two, and for $h^e \in [48, 66]$ we consider a grid with a step size of fourth. This leads to $\{4, 9, 15, 26, 31\}$ estimators $\hat{\theta}^c_T$ for $h = \{5, 10, 22, 44, 66\}$, respectively. To produce forecasts, we use a 'rolling window' estimation scheme. Denoting the total sample size by T, we let m be the fixed size of the estimation window. We formulate the first h-step-

Table 5. In-Sample Results for MEM-GARCH(1,1) Model.

	$h^e=1$			$h^e=5$			$h^e=10$			$h^e=22$			$h^e=44$			$h^e=66$		
	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β	ω	α	β
AA	0.12	0.33	0.65	0.07	0.25	0.74	0.05	0.20	0.79	0.04	0.16	0.83	0.02	0.07	0.93	0.01	0.07	0.93
AXP	0.03	0.42	0.58	0.02	0.31	0.69	0.02	0.28	0.72	0.01	0.20	0.80	0.00	0.05	0.95	0.00	0.05	0.95
BA	0.06	0.33	0.65	0.04	0.24	0.75	0.04	0.22	0.77	0.03	0.16	0.83	0.01	0.06	0.93	0.01	0.07	0.93
CAT	0.10	0.38	0.59	0.07	0.30	0.68	0.06	0.28	0.70	0.05	0.25	0.74	0.02	0.07	0.93	0.02	0.08	0.92
GE	0.03	0.41	0.59	0.02	0.28	0.72	0.01	0.25	0.75	0.00	0.08	0.92	0.00	0.06	0.94	0.00	0.07	0.93
HD	0.07	0.37	0.61	0.04	0.26	0.73	0.03	0.24	0.75	0.02	0.15	0.84	0.01	0.08	0.92	0.01	0.09	0.91
HON	0.07	0.41	0.58	0.07	0.34	0.64	0.06	0.33	0.66	0.05	0.30	0.69	0.01	0.04	0.96	0.01	0.05	0.95
IBM	0.03	0.36	0.63	0.03	0.30	0.69	0.02	0.29	0.70	0.02	0.27	0.72	0.01	0.09	0.90	0.01	0.10	0.90
JPM	0.06	0.51	0.49	0.03	0.37	0.63	0.03	0.31	0.69	0.01	0.08	0.92	0.01	0.05	0.95	0.01	0.05	0.95
KO	0.03	0.38	0.60	0.02	0.27	0.72	0.01	0.23	0.76	0.01	0.22	0.78	0.01	0.17	0.83	0.01	0.15	0.85

Notes: In-sample parameter estimates for the MEM-GARCH(1,1) model for 10 individual stocks. Results are shown for the QML estimation method ($h^e = 1$) and for the HM estimation method ($h^e \in \{5, 10, 22, 44, 66\}$).

Table 6. In-Sample Results for cGARCH model.

	$h^e=1$					$h^e=10$					$h^e=22$					$h^e=44$				
	ω	α	β	φ	ρ	ω	α	β	ϕ	ρ	ω	α	β	ϕ	ρ	ω	α	β	φ	ρ
AA	0.01	0.03	0.80	0.04	1.00	0.01	0.05	0.86	0.04	1.00	0.01	0.06	0.85	0.05	1.00	0.01	0.11	0.70	0.05	1.00
AXP	0.01	0.07	0.84	0.05	1.00	0.01	0.09	0.82	0.06	1.00	0.01	0.09	0.80	0.06	1.00	0.00	0.11	0.72	0.06	1.00
BA	0.01	0.04	0.89	0.03	1.00	0.01	0.05	0.88	0.03	1.00	0.01	0.04	0.91	0.03	1.00	0.01	0.02	0.95	0.03	1.00
CAT	0.02	0.04	0.88	0.03	0.99	0.01	0.08	0.79	0.04	1.00	0.01	0.08	0.76	0.04	1.00	0.01	0.08	0.67	0.04	1.00
GE	0.00	0.07	0.73	0.04	1.00	0.00	0.09	0.65	0.04	1.00	0.00	0.12	0.46	0.05	1.00	0.00	0.00	0.39	0.04	1.00
HD	0.02	0.03	0.91	0.04	0.99	0.02	0.05	0.88	0.05	0.99	0.02	0.03	0.89	0.06	0.99	0.01	0.04	0.91	0.06	1.00
HON	0.01	0.09	0.80	0.03	0.99	0.01	0.09	0.81	0.04	1.00	0.01	0.12	0.73	0.04	1.00	0.01	0.17	0.53	0.04	1.00
IBM	0.01	0.04	0.87	0.04	1.00	0.01	0.08	0.81	0.05	0.99	0.01	0.08	0.78	0.05	0.99	0.01	0.08	0.64	0.05	1.00
JPM	0.01	0.08	0.79	0.06	1.00	0.01	0.07	0.81	0.07	1.00	0.01	0.08	0.73	0.07	1.00	0.01	0.07	0.70	0.07	1.00
KO	0.00	0.05	0.84	0.03	1.00	0.00	0.09	0.78	0.05	1.00	0.00	0.12	0.73	0.06	1.00	0.00	0.14	0.63	0.06	1.00

Notes: In-sample parameter estimates for the cGARCH(1,1) model for 10 individual stocks. Results are shown for the QML estimation method ($h^e = 1$) and for the HM estimation method ($h^e \in \{5, 10, 22, 44, 66\}$).

Table 7. In-Sample Results for MEM-cGARCH model.

	$h^e=1$					$h^e=10$					$h^e=22$					$h^e=44$				
	ω	α	β	φ	ρ	ω	α	β	φ	ρ	ω	α	β	φ	ρ	ω	α	β	φ	ρ
AA	0.03	0.27	0.57	0.09	0.99	0.02	0.17	0.74	0.12	1.00	0.02	0.17	0.74	0.14	1.00	0.02	0.33	0.51	0.16	1.00
AXP	0.01	0.27	0.54	0.19	1.00	0.01	0.21	0.71	0.17	1.00	0.00	0.25	0.67	0.18	1.00	0.00	0.33	0.55	0.16	1.00
BA	0.01	0.28	0.57	0.09	1.00	0.01	0.22	0.71	0.08	1.00	0.01	0.22	0.71	0.08	1.00	0.01	0.23	0.68	0.11	1.00
CAT	0.06	0.21	0.25	0.25	0.98	0.02	0.23	0.70	0.10	0.99	0.01	0.26	0.67	0.11	1.00	0.01	0.29	0.59	0.16	1.00
GE	0.01	0.31	0.53	0.13	1.00	0.00	0.24	0.65	0.15	1.00	0.00	0.32	0.53	0.16	1.00	0.00	0.40	0.34	0.14	1.00
HD	0.03	0.25	0.33	0.22	0.99	0.02	0.17	0.75	0.14	1.00	0.01	0.22	0.68	0.15	1.00	0.01	0.26	0.55	0.16	1.00
HON	0.01	0.33	0.55	0.10	1.00	0.01	0.29	0.62	0.10	1.00	0.01	0.36	0.53	0.11	1.00	0.01	0.49	0.30	0.11	1.00
IBM	0.03	0.16	0.30	0.30	0.99	0.01	0.17	0.76	0.17	0.99	0.01	0.20	0.72	0.19	0.99	0.01	0.28	0.58	0.20	1.00
JPM	0.02	0.37	0.42	0.19	1.00	0.01	0.28	0.61	0.19	1.00	0.01	0.32	0.55	0.18	1.00	0.01	0.37	0.46	0.16	1.00
KO	0.01	0.22	0.52	0.20	0.99	0.01	0.12	0.81	0.16	1.00	0.01	0.11	0.83	0.18	1.00	0.01	0.13	0.77	0.20	1.00

Notes: In-sample parameter estimates for the MEM-cGARCH(1,1) model for 10 individual stocks. Results are shown for the QML estimation method ($h^e = 1$) and for the HM estimation method ($h^e \in \{5, 10, 22, 44, 66\}$).

Table 8. Empirical Rejection Frequency of a Misspecification Test.

	Rejection Frequency								
h^e	GARCH(1,1)	MEM-GARCH(1,1)	cGARCH	MEM-cGARCH	h^e	GARCH(1,1)	MEM-GARCH(1,1)	cGARCH	MEM-cGARCH
2	0	2	0	2	26	2	5	0	2
3	1	6	0	3	28	2	5	0	2
4	1	6	1	5	30	2	6	0	2
5	0	6	2	7	32	2	8	1	2
6	0	6	1	7	34	2	8	1	2
7	0	7	1	7	36	3	9	1	2
8	0	7	2	7	38	3	9	1	2
9	0	7	3	7	40	3	9	1	2
10	0	7	3	7	42	3	10	1	2
12	1	7	2	5	44	3	10	1	2
14	3	5	2	5	48	2	10	2	2
16	3	6	5	4	52	2	10	2	2
18	3	5	5	4	56	2	10	1	2
20	3	6	5	4	60	2	10	2	2
22	2	5	6	4	66	2	10	2	2
24	2	5	0	2					

Notes: Empirical rejection frequency of a misspecification for GARCH and cGARCH models for different estimation horizons h^e used to obtain the HM estimator. Significance level is 5%. For the GARCH(1,1) and the MEM-GARCH(1,1) we use West (1997) estimator to construct the test statistic. For the cGARCH and the MEM-cGARCH models, we use the Newey and West (1987) estimator with the number of lags determined by the following rule: $k \times h^e$, where $k = 30$ for $h^e \leq 5$, $k = 10$ for $5 < h^e \leq 10$, $k = 5$ for $10 < h^e \leq 22$, and $k = 1/2$ for $h^e > 22$. The scaling factors were determined empirically to ensure an accurate size of a test.

ahead cumulative return variance forecast at time m for the period $m+1 : m+h$ using a sample of observations indexed $1, \ldots, m$, and compare this forecast with cumulative realised measure (realised kernel) defined as $x_{m+1,h} = x_{m+1} + x_{m+2} + \ldots + x_{m+h}$. The second forecast is formulated at time $m+1$ using a sample of previous observations indexed by $2, \ldots, m+1$, ensuring that the in-sample size remains the same. The second forecast is compared with $x_{(m+1)+1,h} = x_{(m+1)+1} + x_{(m+1)+2} + \ldots + x_{(m+1)+h}$. Iterating this procedure forward yields $n = T - m - h + 1$ forecasts for a given h. We repeatedly re-estimate parameters each 25 observations, which roughly corresponds to monthly updating.

Table 9 presents the mean difference between the OOS loss evaluated at the HM estimator $\hat{\theta}_T^c$ and the QML estimator $\hat{\theta}_T^d$ for each of the 10 assets and each forecasting horizon h. The first Table shows results for a standard GARCH(1,1) model, while the second shows results for a MEM-GARCH(1,1) model. In both Tables the mean loss difference is reported for the 'best' estimation horizon for a horizon-matched approach, where 'best' means the horizon which minimises the OOS loss evaluated at $\hat{\theta}_T^c$ for a forecasting horizon h. The 'best' estimation horizon (reported h^e) is often smaller than h, in particular for larger h. However, for the MEM-GARCH(1,1) model, estimation and forecasting horizons coincide for $h = 5$ and $h = 10$ for many stocks. From the Table, it is evident that in all cases the average OOS loss difference is negative, favouring the use of the HM estimator. Moreover, for both types of models, the absolute value of the loss difference is relatively small for small to medium horizons ($h = 5, 10, 22$). The loss difference increases for larger horizons, particularly for the MEM-GARCH(1,1) model, indicating that the relative performance of the HM estimator improves. For some stocks, our estimator reduces the OOS loss by up to 0.19 compared to the QML estimator.

To determine whether these differences in OOS loss are statistically significant, we conduct the Diebold-Mariano (DM) test in the framework of Giacomini and White (2006), which fits the 'rolling window' estimation scheme. The results are presented in Appendix A7, with t-statistics for the DM test

Table 9. Out-of-Sample Performance for GARCH(1,1) Models.

	Standard GARCH(1,1)					MEM-GARCH(1,1)					
	h = 5	h = 10	h = 22	h = 44	h = 66		h = 5	h = 10	h = 22	h = 44	h = 66
	$h^e = 4$	$h^e = 4$	$h^e = 4$	$h^e = 3$	$h^e = 3$		$h^e = 5$	$h^e = 10$	$h^e = 12$	$h^e = 32$	$h^e = 32$
AA	-0.01	-0.01	-0.01	-0.02	-0.03	AA	0.00	-0.01	-0.02	-0.06	-0.11
	$h^e = 4$	$h^e = 7$	$h^e = 8$	$h^e = 14$	$h^e = 14$		$h^e = 3$	$h^e = 10$	$h^e = 18$	$h^e = 44$	$h^e = 22$
AXP	0.00	-0.01	-0.01	-0.02	-0.04	AXP	0.00	0.00	-0.02	-0.07	-0.11
	$h^e = 2$	$h^e = 2$	$h^e = 2$	$h^e = 10$	$h^e = 10$		$h^e = 5$	$h^e = 8$	$h^e = 7$	$h^e = 14$	$h^e = 14$
BA	-0.01	-0.01	-0.01	-0.01	-0.02	BA	0.00	0.00	0.00	-0.01	-0.02
	$h^e = 4$	$h^e = 4$	$h^e = 5$	$h^e = 5$	$h^e = 5$		$h^e = 5$	$h^e = 10$	$h^e = 14$	$h^e = 20$	$h^e = 20$
CAT	-0.01	-0.01	-0.02	-0.02	-0.03	CAT	0.00	-0.02	-0.04	-0.10	-0.15
	$h^e = 2$	$h^e = 10$	$h^e = 12$	$h^e = 14$	$h^e = 14$		$h^e = 5$	$h^e = 10$	$h^e = 16$	$h^e = 30$	$h^e = 22$
GE	-0.02	-0.02	-0.03	-0.04	-0.07	GE	0.00	-0.01	-0.04	-0.13	-0.19
	$h^e = 5$	$h^e = 5$	$h^e = 5$	$h^e = 5$	$h^e = 5$		$h^e = 3$	$h^e = 10$	$h^e = 16$	$h^e = 32$	$h^e = 34$
HD	-0.01	-0.01	-0.01	-0.01	-0.02	HD	0.00	-0.01	-0.03	-0.08	-0.13
	$h^e = 2$	$h^e = 2$	$h^e = 7$	$h^e = 7$	$h^e = 7$		$h^e = 2$	$h^e = 2$	$h^e = 14$	$h^e = 14$	$h^e = 14$
HON	-0.02	-0.03	-0.03	-0.03	-0.05	HON	0.00	0.00	0.00	-0.02	-0.03
	$h^e = 4$	$h^e = 4$	$h^e = 4$	$h^e = 3$	$h^e = 3$		$h^e = 5$	$h^e = 10$	$h^e = 10$	$h^e = 10$	$h^e = 16$
IBM	-0.01	-0.01	-0.02	-0.02	-0.03	IBM	0.00	0.00	-0.01	-0.02	-0.04
	$h^e = 2$	$h^e = 2$	$h^e = 2$	$h^e = 2$	$h^e = 2$		$h^e = 3$	$h^e = 10$	$h^e = 18$	$h^e = 34$	$h^e = 34$
JPM	-0.01	-0.01	-0.02	-0.03	-0.05	JPM	0.00	-0.01	-0.06	-0.11	-0.18
	$h^e = 2$	$h^e = 2$	$h^e = 2$	$h^e = 2$	$h^e = 2$		$h^e = 2$	$h^e = 8$	$h^e = 5$	$h^e = 8$	$h^e = 9$
KO	-0.01	-0.02	-0.02	-0.02	-0.02	KO	0.00	0.00	-0.01	-0.01	-0.02

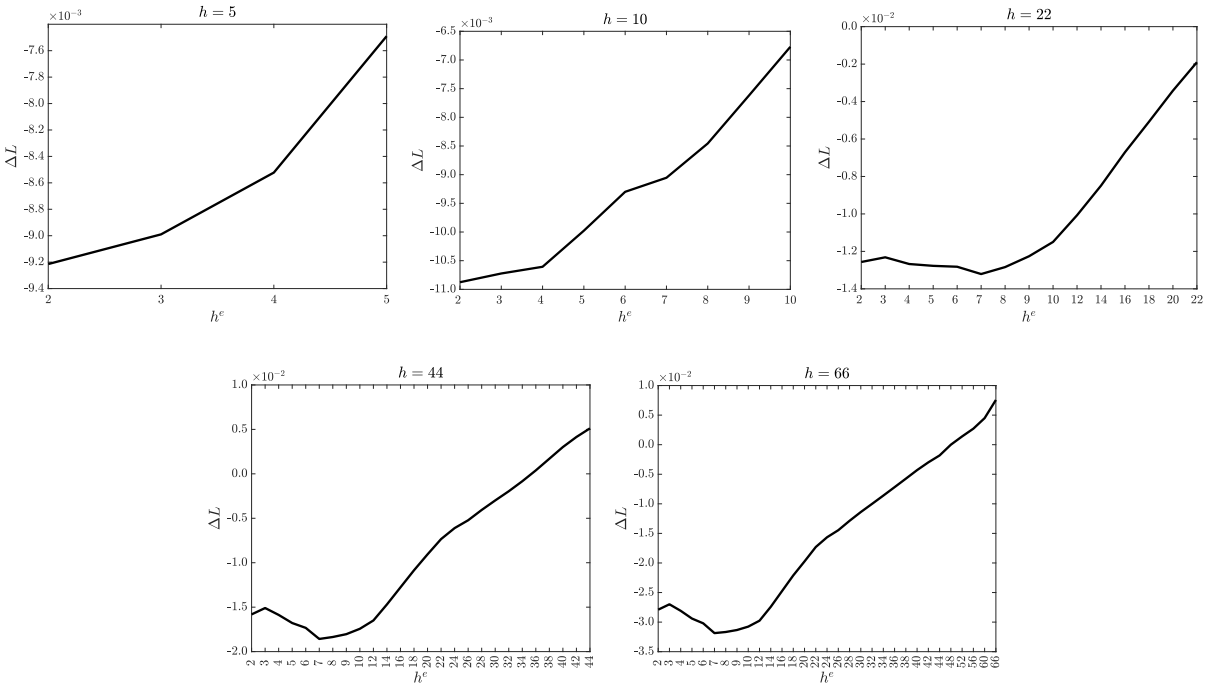
Notes: Out-of-sample average loss difference for QLIKE loss function for 10 individual stocks. Loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$, where L denotes the QLIKE loss evaluated at the QML estimator $\hat{\theta}^d_T$ and at the HM estimator $\hat{\theta}^c_T$. Considered forecasting horizons are $h \in \{5, 10, 22, 44, 66\}$. Horizon at which $\hat{\theta}^c_T$ is obtained does not necessarily match forecasting (evaluation) horizon. For each forecasting horizon and individual stock we report the estimation horizon (h^e) which minimizes the average QLIKE loss evaluated at the HM estimator.

shown in Table 13 for the standard GARCH(1,1) model and Table 11 for the MEM-GARCH(1,1) model. For both models, we use $h-1$ and $0.75n^{1/3}$ lags to estimate the asymptotic variance-covariance matrix of the sample average loss difference, following recommendations of Giacomini and White (2006) and Andrews (1991), respectively. A negative t-statistic indicates that the HM estimator generates lower average OOS loss. We conduct a two-sided test at the 5% significance level. It is important to mention that the significance of results vary across two different lag lengths. In particular, it is not surprising that for smaller lag ($0.75n^{1/3}$) the QML estimator is significantly outperformed by the HM estimator, while for larger lag, the difference is not statistically significant (except for isolated cases).

In addition to Table 9, Figure 7 and Figure 8 plot OOS loss difference averaged across all ten stocks for the GARCH(1,1) model and for the MEM-GARCH(1,1) model, respectively. As visible from the Figures, the optimal estimation horizon for the HM estimator is consistently larger for the MEM-GARCH(1,1) model in contrast to the standard GARCH(1,1). The 'optimal' horizon is defined as the one at which the difference in losses between the HM and QML estimators is most negative, indicating the smallest loss for the HM estimator at that horizon. In particular, the difference in optimal estimation horizons for the two models is most noticeable for larger forecasting horizons. In particular, for the GARCH(1,1) the optimal horizon is around $h^e = 7$ for forecasting horizons $h = 44$ and $h = 66$, while for the MEM-GARCH(1,1) is it around $h^e = 16$. For these models, the reduction in the OOS loss reaches up to 0.035 and 0.092, respectively. Moreover, as the estimation horizon increases, the performance of

the HM estimator deteriorates much more steeply for the GARCH(1,1) model compared to the MEM-GARCH(1,1) model.

Figure 7. Out-of-Sample Performance Averaged Across Stocks for GARCH(1,1) Model.

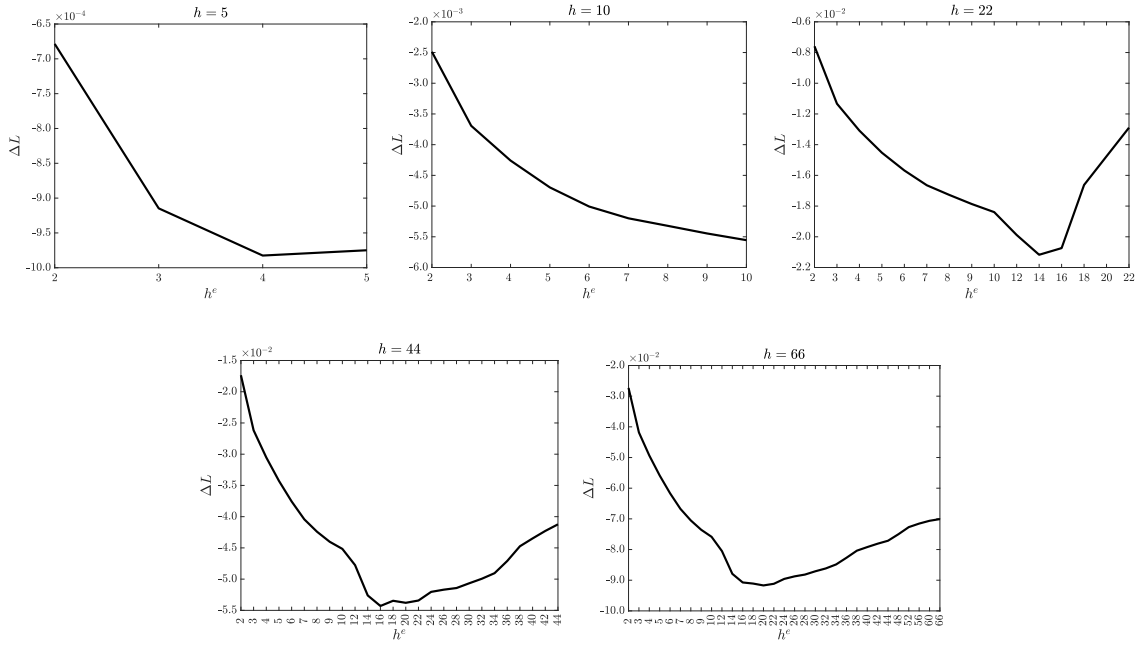


Notes: Out-of-sample average QLIKE loss difference averaged across 10 individual stocks for the MEM-GARCH(1,1) model. Loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$, where L denotes the QLIKE loss function. Each plot corresponds to a specific forecasting horizon $h \in \{5, 10, 22, 44, 66\}$. Horizontal axis ($h^e$) denotes the estimation horizon for the HM estimator, which may be equal to or smaller than the forecasting horizon.

We now analyse the results of the cGARCH model. Figure 9 and Figure 10 plot the OOS loss differences averaged across all ten stocks for the cGARCH model and for the MEM-cGARCH model, respectively. For the standard cGARCH model, the optimal estimation horizon, in particular for forecasting horizons $h = 44$ and $h = 66$, is lower than that for the standard GARCH(1,1) model. Additionally, while there is still the OOS gain in using the HM estimator for the cGARCH model, the amplitude of this gain is smaller in contrast to the standard GARCH(1,1) model: the maximal decrease in the OOS loss for the cGARCH model is roughly 0.005 for the largest forecasting horizon. For the MEM-cGARCH model, the optimal estimation horizon is similar to that of the MEM-GARCH(1,1) model. However, there are two main differences: first, the gains from using the HM estimator are again smaller, and second, the reduction in gains becomes more pronounced once the estimation horizon exceeds the optimal one. This suggest that, overall, the HM estimator provides larger forecasting gains in simpler models, which are more likely to be misspecified.

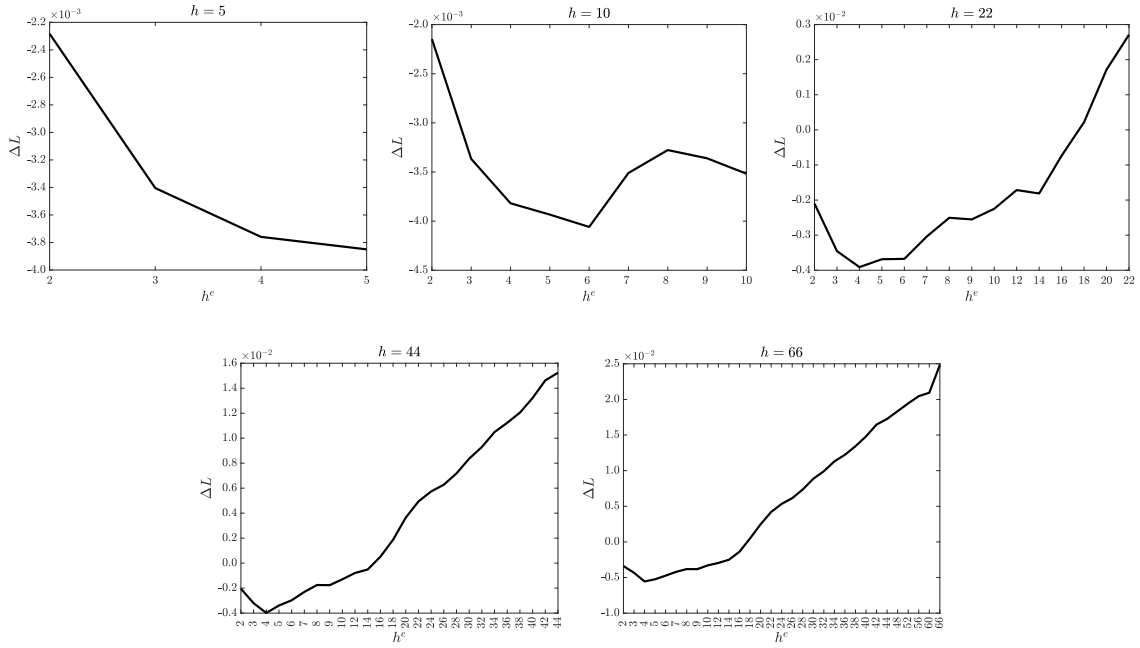
In addition to these aggregated results, Appendix A7 provides results for each stock individually in Table 12. Table 13 and Table 11 present the DM statistic, indicating whether the loss differences are statistically significant.

Figure 8. Out-of-Sample Performance Averaged Across Stocks for MEM-GARCH(1,1) Model.



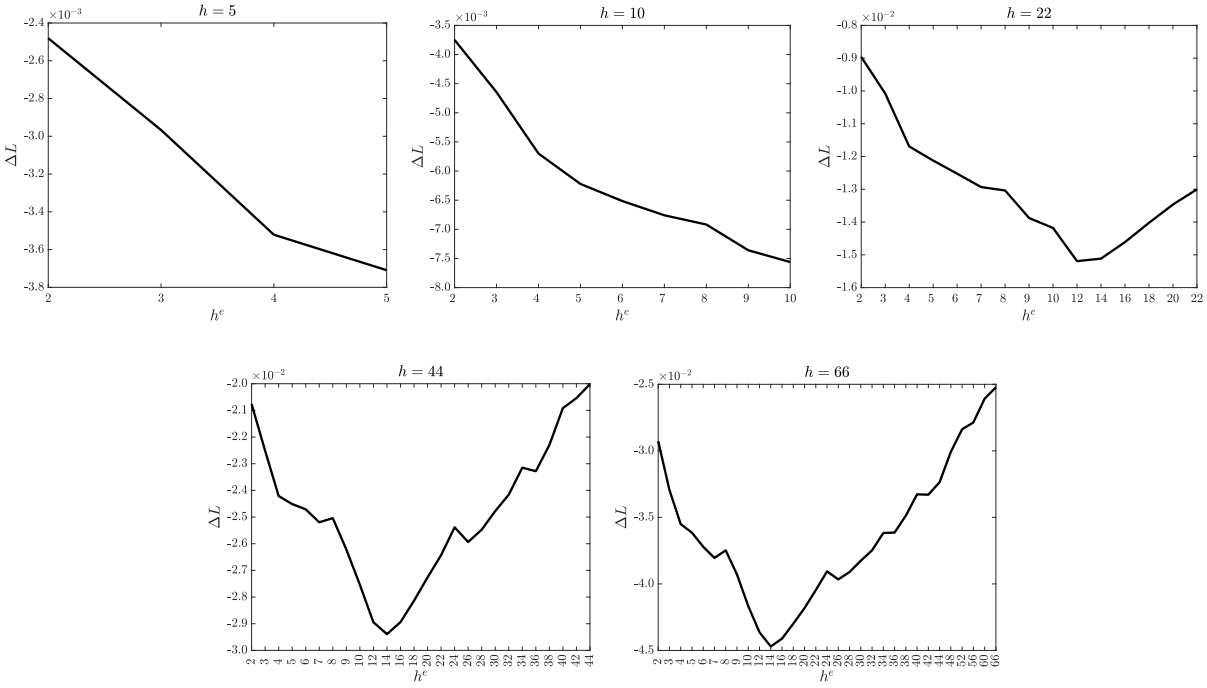
Notes: Out-of-sample average QLIKE loss difference averaged across 10 individual stocks for the MEM-GARCH(1,1) model. Loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$, where L denotes the QLIKE loss function. Each plot corresponds to a specific forecasting horizon $h \in \{5, 10, 22, 44, 66\}$. Horizontal axis ($h^e$) denotes the estimation horizon for the HM estimator, which may be equal to or smaller than the forecasting horizon.

Figure 9. Out-of-Sample Performance Averaged Across Stocks for cGARCH Model.



Notes: Out-of-sample average QLIKE loss difference averaged across 10 individual stocks for the cGARCH model. Loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$, where L denotes the QLIKE loss function. Each plot corresponds to a specific forecasting horizon $h \in \{5, 10, 22, 44, 66\}$. Horizontal axis ($h^e$) denotes the estimation horizon for the HM estimator, which may be equal to or smaller than the forecasting horizon.

Figure 10. Out-of-Sample Performance Averaged Across Stocks for MEM-cGARCH(1,1) Model.



Notes: Out-of-sample average QLIKE loss difference averaged across 10 individual stocks for the MEM-cGARCH model. Loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$, where L denotes the QLIKE loss function. Each plot corresponds to a specific forecasting horizon $h \in \{5, 10, 22, 44, 66\}$. Horizontal axis ($h^e$) denotes the estimation horizon for the HM estimator, which may be equal to or smaller than the forecasting horizon.

7. DISCUSSION

This paper proposes an estimation method to improve multi-period volatility forecasts from misspecified GARCH-type models. In many financial decision-making environments, it may not be possible to substitute the given statistical model, at least in the short term, due to potential institutional barriers that hinder such changes. We overcome this challenge by maintaining the functional form of the candidate model and estimating its parameters by 'matching' the estimation loss function to the specification of the forecast evaluation loss function, which we assume to be given ex-ante by the decision-maker. 'Matching' of loss functions is done to reduce the impact of model misspecification on the accuracy of multi-period volatility forecasts (Hansen & Dumitrescu, 2022).

With the primary purpose of our estimator being forecasting, we propose to use it in a misspecification test based on the Hausman principle (Hausman, 1978). Indeed, while our estimator is by construction consistent for the true parameter under correct model specification, in cases of model misspecification, it converges in probability to the pseudo-true value that differs from the pseudo-true value from the standard estimation method. We explore this result by comparing the two estimators in the misspecification test. In a Monte-Carlo study, we examine the size and power properties of the test using the hard-to-beat GARCH(1,1) model, treating it as misspecified with respect to the empirical phenomenon of long-memory dynamics in volatility. The results demonstrate that the test is reasonably sized and has increasing power with increasing long-memory misspecification. The highest power is achieved when our estimator is derived using the shortest horizon in the estimation loss function. Additionally, we recover multi-period volatility forecasts from simulated data and find that, under correct specification, both estimators perform equivalently. However, our estimator yields more accurate forecasts when the

model is misspecified. In these cases, our findings show that for our estimator, it is optimal to 'match' the estimation loss function to a shorter horizon than the forecasting horizon. The shorter optimal estimation horizon reflects a bias-variance trade-off: the proposed estimator adds additional variance to the forecast to reduce the bias arising from model misspecification.

We apply the proposed estimation method to an out-of-sample forecasting exercise across ten series of returns and realised measures from the paper by Gorgi et al. (2019) over $2001-2010$. Using GARCH(1,1) and cGARCH models specified for returns and realised measures, we show that our estimator consistently produces more accurate out-of-sample forecasts across different forecasting horizons, from short to large. Three main findings emerge. First, the accuracy gains from our estimator are more significant for the realised measures models, suggesting that our estimator performs better in environments with higher signals about the volatility. Second, for both returns and realised measures types of models, the optimal estimation horizon for our estimator is lower than the forecasting horizon of interest, consistent with results from our Monte-Carlo study. However, this discrepancy between the estimation and forecasting horizons is more pronounced for the model based on returns, particularly at longer forecasting horizons. Third, by comparing GARCH(1,1) with cGARCH models, we find that the amplitude of gains from our estimator is smaller for cGARCH models. This indicates that the value of our estimator is more pronounced for underparameterised models, which are more likely to be misspecified.

We conclude with practical guidelines for applying our estimation approach in forecasting contexts. Like other misspecification tests, the test that this paper develops provides only statistical evidence of model misspecification without identifying its specific sources. In this paper, we demonstrate one such source - long-memory dynamics, a common characteristic of volatility - that the test detects. Overall, the rejection of the test's null hypothesis should prompt further adjustments to the baseline model. However, since these adjustments may be limited, we suggest the following approach. Generally, we recommend performing the test for several estimation horizons below the forecasting one to check whether the test consistently rejects the null hypothesis or not. If the test does not reject the null hypothesis at a given significance level for several horizons, we recommend estimating the baseline model using the standard estimation method and constructing forecasts accordingly. If the test rejects the null, we recommend using our estimator instead. For models specified for realised measures, we recommend aligning the estimation horizon exactly to the forecasting horizon for shorter forecasting horizons (e.g., $h = 5, 10$). For medium (e.g., $h = 22, 44$) and large forecasting horizons, we suggest using the estimation horizon around $h^e = 15$. For models specified for returns, we advise using either the standard estimator for short to medium forecasting horizons or our estimator with short estimation horizons close to $h^e = 3$. However, for longer forecasting horizons, we recommend our estimator with an estimation horizon close to ten.

APPENDIX

A1. GARCH(2,2) representation of cGARCH

Using the lag operator L, the first equation in (17) could be written as

$$\begin{aligned}\sigma_t^2 &= q_t + \alpha(r_{t-1}^2 - Lq_t) + \beta(\sigma_{t-1}^2 - Lq_t) \\ &= \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + (1-(\alpha+\beta)L)q_t,\end{aligned} \tag{53}$$

where $Lq_t = q_{t-1}$. In the similar manner, we could re-write the second equation in (17) for long-run component q_t:

$$q_t = \omega + \rho L q_t + \varphi(r_{t-1}^2 - \sigma_{t-1}^2) \tag{54}$$

implying that

$$q_t = \frac{1}{1-\rho L}\left(\omega + \varphi(r_{t-1}^2 - \sigma_{t-1}^2)\right). \tag{55}$$

Plugging (55) into (53) results in

$$\sigma_t^2 = \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \frac{1-(\alpha+\beta)L}{1-\rho L}(\omega + \varphi(r_{t-1}^2 - \sigma_{t-1}^2)). \tag{56}$$

Multiplying both sides by $(1-\rho L)$ gives the GARCH(2,2) structure for σ_t^2.

A2. Autocorrelation function for the GARCH and the cGARCH models

The GARCH(1,1) model can be expressed as the ARMA(1,1) model for the second-order sequence $\{r_t^2\}$ of squared returns (residuals), that is

$$r_t^2 = \omega + (\alpha+\beta)r_{t-1}^2 + v_t - \beta v_{t-1}, \quad \text{equivalently} \quad (1-\alpha L - \beta L)r_t^2 = \omega + (1-\beta L)v_t, \tag{57}$$

with L the lag operator $Lv_t = v_{t-1}$ and $v_t = r_t^2 - \sigma_t^2$ being the martingale difference sequence (MDS), and thus, white-noise (WN) innovation term. The autocovariance function of the ARMA(1,1) model displays the shape of that for an AR(1) model for $k \geq 2$, specifically

$$\gamma_{r^2}(k) = (\alpha+\beta)\gamma_{r^2}(k-1), \quad k \geq 2. \tag{58}$$

Hence, the autocorrelation function for squared returns is equal to

$$\rho_{r^2}(k) = \frac{\gamma_{r^2}(k)}{\gamma_{r^2}(0)} = (\alpha+\beta)^{k-1}\rho_{r^2}(1), \quad k \geq 2, \tag{59}$$

with

$$\rho_{r^2}(1) = \frac{\alpha(1-(\alpha+\beta)\beta)}{1-(\alpha+\beta)^2+\alpha^2}, \tag{60}$$

where $\gamma_{r^2}(k) = \mathbb{E}[(r_t^2 - \overline{\sigma}^2)(r_{t-k}^2 - \overline{\sigma}^2)]$. This shows that the autocorrelation of squared return process decreases towards zero proportionally to $(\alpha+\beta)^{k-1}$.

As discussed above, the cGARCH model has a reduced form of the GARCH(2,2) model. Specifically, (17) can be shown to be

$$\begin{aligned}\sigma_t^2 &= (1-\alpha-\beta)\omega + (\alpha+\varphi)r_{t-1}^2 + (-\varphi(\alpha+\beta) - \rho\alpha)r_{t-2}^2 \\ &\quad + (\rho+\beta-\varphi)\sigma_{t-1}^2 + (\varphi(\alpha+\beta) - \rho\beta)\sigma_{t-2}^2,\end{aligned} \tag{61}$$

which is the GARCH(2,2) model of the form

$$\sigma_t^2 = \omega' + \alpha_1' r_{t-1}^2 + \alpha_2' r_{t-2}^2 + \beta_1'\sigma_{t-1}^2 + \beta_2'\sigma_{t-2}^2 \tag{62}$$

with

$$\begin{aligned}\omega' &= (1-\alpha-\beta)\omega \\ \alpha_1' &= \alpha+\varphi \\ \alpha_2' &= -\varphi(\alpha+\beta)-\rho\alpha \\ \beta_1' &= \rho+\beta-\varphi \\ \beta_2' &= \varphi(\alpha+\beta)-\rho\beta.\end{aligned} \tag{63}$$

Similar to the GARCH(1,1) model, the GARCH(2,2) can be formulated as the ARMA(2,2) process by substituting the variables σ_{t-j}^2 by $r_{t-j}^2 - v_{t-j}$ for $j = 1, 2$:

$$r_t^2 = \omega' + \sum_{i=1}^{2}(\alpha_i' + \beta_i')r_{t-i}^2 + v_t - \sum_{j=1}^{2}\beta_j' v_{t-j}. \tag{64}$$

The autocovariance function of an ARMA(2,2) model displays the shape of that for an AR(2) model for $k \geq 3$, that is

$$\gamma_{r^2}(k) = (\alpha+\beta+\rho)\gamma_{r^2}(k-1) + (-\rho(\alpha+\beta))\gamma_{r^2}(k-2), \quad k \geq 3, \tag{65}$$

where $\gamma_{r^2}(0), \gamma_{r^2}(1)$ and $\gamma_{r^2}(2)$ can be obtained from the set of Yule-Walker equations.

A3. MDS score property

Under correct specification of the first two conditional moments, it follows that

$$\mathbb{E}[s_t^c(\theta_0)|\mathcal{F}_{t-1}] = 0,$$

and hence $\mathbb{E}[s_t^c(\theta_0)] = 0$, where the score $s_t^c(\theta)$ is defined as

$$s_t^c(\theta) = \frac{\partial \text{QLIKE}(\tilde{r}_{t,h}^2, \tilde{\sigma}_{t,h}^2(\theta))}{\partial\theta} = (\tilde{\sigma}_{t,h}^2(\theta) - \tilde{r}_{t,h}^2)\frac{1}{\tilde{\sigma}_{t,h}^4(\theta)}\frac{\partial\tilde{\sigma}_{t,h}^2(\theta)}{\partial\theta}.$$

Proof. The score evaluated at the true parameter vector θ_0 has the martingale difference property:

$$\begin{aligned}\mathbb{E}[\mathbb{E}[s_t^c(\theta_0)|\mathcal{F}_{t-1}]] &= \mathbb{E}\left[\mathbb{E}\left[(\tilde{\sigma}_{t,h}^2(\theta_0) - \tilde{r}_{t,h}^2)|\mathcal{F}_{t-1}\right]\frac{1}{\tilde{\sigma}_{t,h}^4(\theta_0)}\frac{\partial\tilde{\sigma}_{t,h}^2(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\right] \\ &= 0,\end{aligned}$$

since $\mathbb{E}[\tilde{r}_{t,h}^2|\mathcal{F}_{t-1}] = \tilde{\sigma}_{t,h}^2(\theta_0)$, which follows from the assumptions that the conditional mean and variance functions are correctly specified. By the law of iterated expectations,

$$\begin{aligned}\mathbb{E}[s_t^c(\theta_0)] &= \mathbb{E}[\mathbb{E}[s_t^c(\theta_0)|\mathcal{F}_{t-1}]] \\ &= 0.\end{aligned}$$

□

A4. MLE for MEM-GARCH model

GARCH-type model for realised measure is specified as $x_t = \sigma_t^2 u_t$, where x_t is the realised measure, with σ_t^2 following the GARCH structure, where we assume conditional gamma distribution for the error term u_t:

$$f(u_t|\mathcal{F}_{t-1}) = \frac{1}{\Gamma(a)b^a}u_t^{a-1}\exp(-u_t/b). \tag{66}$$

Imposing the unit mean property on the distribution of u_t implies that $b = 1/a$, and thus

$$f(u_t|\mathcal{F}_{t-1}) = \frac{1}{\Gamma(a)}a^a u_t^{a-1}\exp(-au_t). \tag{67}$$

Using a theorem for a transformation of a random variable, the conditional density of x_t is given by

$$f(x_t|\mathcal{F}_{t-1}) = \frac{1}{\sigma_t^2}\frac{1}{\Gamma(a)}a^a\left(\frac{x_t}{\sigma_t^2}\right)^{a-1}\exp\left(-a\frac{x_t}{\sigma_t^2}\right), \tag{68}$$

implying that the log of the conditional density is equal to

$$\log f(x_t|\mathcal{F}_{t-1}) = (a-1)\log x_t + a\log a - \log\Gamma(a) - a\left(\log\sigma_t^2 + \frac{x_t}{\sigma_t^2}\right). \tag{69}$$

A5. Detailed scores for MEM GARCH(1,1) model

We provide a detailed expression for the scores of the MEM-GARCH(1,1) model, which we use to compute the empirical size and power for the misspecification test. This is needed for computing the estimator by West (1997).

Score vectors for the QML and HM estimators are defined as follows:

$$\begin{aligned} s_t^d(\theta) &= -\frac{\partial}{\partial\theta}\text{QLIKE}(x_t,\sigma_t^2(\theta)) = -\frac{\partial}{\partial\theta}\left(\ln\sigma_t^2(\theta) + \frac{x_t}{\sigma_t^2(\theta)}\right) = \left(\frac{x_t-\sigma_t^2(\theta)}{\sigma_t^4(\theta)}\right)\frac{\partial\sigma_t^2(\theta)}{\partial\theta} \\ s_t^c(\theta) &= -\frac{\partial}{\partial\theta}\text{QLIKE}(\tilde{x}_{t,h},\tilde{\sigma}_{t,h}^2(\theta)) \\ &= -\frac{\partial}{\partial\theta}\left(\ln\tilde{\sigma}_{t,h}^2(\theta) + \frac{\tilde{x}_{t,h}}{\ln\tilde{\sigma}_{t,h}^2(\theta)}\right) = \left(\frac{\tilde{x}_{t,h}-\tilde{\sigma}_{t,h}^2(\theta)}{\tilde{\sigma}_{t,h}^4(\theta)}\right)\frac{\partial\tilde{\sigma}_{t,h}^2(\theta)}{\partial\theta}, \end{aligned} \tag{70}$$

where the conditional variance of the cumulative return can be written as

$$\tilde{\sigma}_{t,h}^2(\theta) = \sum_{j=0}^{h-1}\sigma_{t+j|t-1}^2(\theta) = h\overline{\sigma}^2 + \frac{1-(\alpha+\beta)^h}{1-\alpha-\beta}(\sigma_t^2-\overline{\sigma}^2), \tag{71}$$

where for simplicity we write σ_t^2 instead of $\sigma_{t|t-1}^2$.

In (70), we further expand $\partial\sigma_t^2(\theta)/\partial\theta$ and $\partial\tilde{\sigma}_{t,h}^2(\theta)/\partial\theta$ as

$$\begin{aligned} \frac{\partial\tilde{\sigma}_{t,h}^2(\theta)}{\partial\theta} &= (\sigma_t^2(\theta)-\overline{\sigma}^2)\frac{\partial}{\partial\theta}\left(\frac{1-(\alpha+\beta)^h}{1-\alpha-\beta}\right) + \left(h - \frac{1-(\alpha+\beta)^h}{1-\alpha-\beta}\right)\frac{\partial\overline{\sigma}^2}{\partial\theta} \\ &\quad + \frac{1-(\alpha+\beta)^h}{1-\alpha-\beta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}, \end{aligned} \tag{72}$$

with

$$\begin{aligned} \frac{\partial}{\partial\theta}\left(\frac{1-(\alpha+\beta)^h}{1-\alpha-\beta}\right) &= \frac{1-(\alpha+\beta)^h - h(\alpha+\beta)^{h-1} + h(\alpha+\beta)^h}{(1-\alpha-\beta)^2}\begin{pmatrix}0\\1\\1\end{pmatrix}, \\ \frac{\partial\overline{\sigma}^2}{\partial\theta} &= \frac{1}{1-\alpha-\beta}\begin{pmatrix}1\\\overline{\sigma}^2\\\overline{\sigma}^2\end{pmatrix}, \end{aligned} \tag{73}$$

and

$$\frac{\partial\sigma_t^2(\theta)}{\partial\theta} = \begin{pmatrix}1\\x_{t-1}\\\sigma_{t-1}^2\end{pmatrix} + \beta\frac{\sigma_{t-1}^2(\theta)}{\partial\theta} = \sum_{j=0}^{t-1}\beta^j\begin{pmatrix}1\\x_{t-j-1}\\\sigma_{t-j-1}^2\end{pmatrix}, \tag{74}$$

taking $\partial\sigma_1^2/\partial\theta = 0$ when treating σ_1^2 as a fixed constant, not depending on parameters θ.

Note that when estimating the parameters of the MEM-GARCH(1,1) using the QML method, we use the MFE Toolbox for Matlab developed by Kevin Sheppard. In this Toolbox, the likelihood function

maximised is defined as $-\frac{1}{2}\sum_{t=1}^{T}\left(\ln\sigma_t^2(\theta)+\frac{x_t}{\sigma_t^2(\theta)}\right)$, which is different from the estimation criterion defined above by a scale of 1/2. To construct the estimator of the asymptotic variance-covariance matrix of the joint estimator, we use the estimated Hessian output from the Matlab estimation function. Consequently, we divide x_t for the 'cumulative' score and z_t for the 'daily' score by 2 to obtain the correctness of results

Consider:

$$\begin{aligned}
C_{cd} &= \sum_{j=0}^{h-1}\mathbb{E}[s_t^c {s_{t+j}^d}'] \\
&= \sum_{j=0}^{h-1}\mathbb{E}[z_t u_t \varepsilon_{t+j-h+1} {x_{t+j}}'] \\
&= \sum_{j=0}^{h-1}\mathbb{E}[z_t(\varepsilon_t+\psi_1\varepsilon_{t-1}+\ldots+\psi_{h-1}\varepsilon_{t-h+1})\varepsilon_{t+j-h+1}] \\
&= \sum_{j=0}^{h-1}\sum_{i=0}^{h-1}\mathbb{E}[z_t\psi_i\varepsilon_{t-i}\varepsilon_{t+j-h+1}x_{t+j}'] \\
&= \sum_{j=0}^{h-1}\mathbb{E}[\varepsilon_{t+j-h+1}^2 z_t\psi_{h-1-j}x_{t+j}'] \\
&= \mathbb{E}\left[\varepsilon_t^2\left(\sum_{j=0}^{h-1}\psi_j z_{t+j}\right){x_{t+h-1}}'\right],
\end{aligned} \tag{75}$$

where the fifth equality, where the expression is simplified from a double sum to a single sum, uses the fact that $\mathbb{E}[\varepsilon_t\varepsilon_s]=0$ for $t\neq s$; the sixth equality uses covariance-stationarity of $\mathbb{E}[\varepsilon_{t-h+1}^2 z_t x_t']$ to express $\psi_{h-1}\mathbb{E}[\varepsilon_{t-h+1}^2 z_t x_t'] + \psi_{h-2}\mathbb{E}[\varepsilon_{t-h+2}^2 z_t x_{t+1}'] + \ldots + \psi_0\mathbb{E}[\varepsilon_t^2 z_t x_{t+h-1}']$ as $\psi_{h-1}\mathbb{E}[\varepsilon_t^2 z_{t+h-1}x_{t+h-1}'] + \psi_{h-2}\mathbb{E}[\varepsilon_t^2 z_{t+h-2}x_{t+h-1}'] + \ldots + \psi_0\mathbb{E}[\varepsilon_t^2 z_t x_{t+h-1}']$.

Consider:

$$\begin{aligned}
C_{cc} &= \sum_{j=-h+1}^{h-1}\mathbb{E}[s_t^c s_{t+j}^{c\prime}] \\
&= \mathbb{E}[u_t^2 z_t z_t'] + \sum_{j=1}^{h-1}\mathbb{E}[u_t u_{t+j}(z_t z_{t+j}' + z_{t+j}z_t')] \\
&= \mathbb{E}[(\varepsilon_t^2+\psi_1\varepsilon_{t-1}^2+\ldots+\psi_{h-1}^2\varepsilon_{t-h+1}^2)z_t z_t'] \\
&\quad + \sum_{j=1}^{h-1}\mathbb{E}[(\psi_0\psi_j\varepsilon_t^2+\psi_1\psi_{j+1}\varepsilon_{t-1}^2+\ldots+\psi_{h-j-1}\psi_{h-1}\varepsilon_{t-h+j+1}^2)(z_t z_{t+j}'+z_{t+j}z_t')] \\
&= \mathbb{E}\left[\varepsilon_t^2\sum_{j=0}^{h-1}\psi_j^2 z_{t+j}z_{t+j}'\right] + \sum_{j=1}^{h-1}\mathbb{E}\left[\varepsilon_t^2\sum_{i=0}^{h-j-1}\psi_i\psi_{i+j}(z_{t+i}z_{t+i+j}'+z_{t+i+j}z_{t+i}')\right] \\
&= \mathbb{E}\left[\varepsilon_t^2\left(\sum_{j=0}^{h-1}\psi_j z_{t+j}\right)\left(\sum_{j=0}^{h-1}\psi_j z_{t+j}\right)'\right],
\end{aligned} \tag{76}$$

where the third inequality uses the fact that $\mathbb{E}[\varepsilon_t\varepsilon_s]=0$ for $t\neq s$, the fourth inequality uses covariance-stationarity of $\varepsilon_{t-1}^2 z_t z_t'$, which would imply, for instance, $\mathbb{E}[\varepsilon_{t-1}^2 z_t z_t'] = \mathbb{E}[\varepsilon_t^2 z_{t+1}z_{t+1}']$.

Using structure of a GARCH(1,1) model, we can express ψ_t in terms of GARCH(1,1) parameters:

$$u_t = \sum_{j=0}^{h-1}(r_t^2 - \sigma_{t+j|t-1}^2) = \sum_{j=0}^{h-1}(r_{t+j}^2 - \sigma_{t+j}^2) + \sum_{j=1}^{h-1}(\sigma_{t+j}^2 - \sigma_{t+j|t-1}^2) \tag{77}$$

$$= (r_{t+h-1}^2 - \sigma_{t+h-1}^2) + \sum_{i=0}^{h-2}\left(1 + \alpha\frac{1-(\alpha+\beta)^{h-i-1}}{1-\alpha-\beta}(r_{t+i}^2 - \sigma_{t+i}^2)\right) = \sum_{j=0}^{h-1}\psi_j\varepsilon_{t-j}, \tag{78}$$

A6. Plots: admissible parameter values for LMGARCH(1,d,1)

To guarantee that the conditional variance is non-negative a.s. for all t, we need to show that in the ARCH(∞) it is satisfied that $\psi_i \geq 0$. Conrad and Haag (2006) derive necessary and sufficient conditions for that, which greatly enlarge the set of sufficient conditions by Baillie et al. (1996), Bollerslev and Mikkelsen (1996) and Chung (1999), which have been shown to be quite restrictive. From Corollary 1 of Conrad and Haag (2006), it follows that the conditional variance of the LMGARCH(1,d,1) for the case with $0 < \beta < 1$ is non-negative a.s. iff

$$\text{either } \psi_1 \geq 0 \text{ and } \phi \leq f_2, \text{ or for } k > 2 \text{ with } f_{k-1} < \phi \leq f_k, \text{it holds that } \psi_{k-1} \geq 0, \tag{79}$$

where

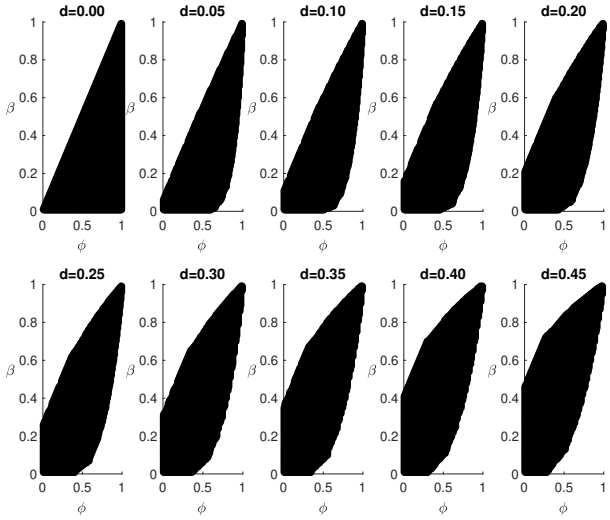
$$\begin{aligned} \psi_1 &= d + \phi - \beta \\ \psi_i &= \beta\psi_{i-1} + (f_i - \phi)(-g_{i-1}), \text{ for all } i \geq 2, \end{aligned} \tag{80}$$

with the coefficients g_i and f_i defined from $(1-L)^d = \sum_{i=0}^{\infty} g_i L^i$:

$$g_0 = 1, \; g_i = f_i g_{i-1}, \quad f_i = \frac{i-1-d}{i}, \text{ for } i = 1, 2, \ldots \tag{81}$$

Given quite high value of ϕ, we check the non-negativity of the conditional variance by ensuring that there exists a $k > 2$ such that the condition $f_{k-1} < \phi \leq f_k$ and $\psi_{k-1} \geq 0$ hold. Figure 11 depicts the admissible set of values for ϕ and β given the value of d such that the constraints are satisfied.

Figure 11. Necessary and sufficient parameter set for the LMGARCH(1,d,1) model for different combinations of (ϕ, β) with $d \in \{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$.



Notes: Necessary and sufficient conditions are taken from Corollary 1 from Conrad and Haag (2006).

A7. Out-of-sample forecasting

Table 10. Out-of-sample performance: DM t-statistic for standard GARCH(1,1) model.

	# of lags : $h-1$					# of lags: $0.75n^{1/3}$					
	h = 5	h = 10	h = 22	h = 44	h = 66		h = 5	h = 10	h = 22	h = 44	h = 66
AA	-3.10	-2.48	-1.89	-1.74	-1.76	AA	-2.71	-2.63	-2.66	-3.15	-3.83
AXP	-1.97	-1.92	-1.51	-1.08	-1.28	AXP	-1.71	-2.06	-2.14	-1.74	-2.47
BA	-2.48	-1.91	-1.44	-1.05	-1.08	BA	-2.06	-2.09	-2.16	-2.07	-2.53
CAT	-3.08	-1.88	-1.29	-1.20	-1.38	CAT	-2.60	-2.04	-1.88	-2.15	-2.74
GE	-2.52	-2.16	-1.92	-1.64	-1.48	GE	-2.18	-2.31	-2.60	-2.78	-3.11
HD	-2.68	-2.25	-1.92	-1.46	-1.34	HD	-2.22	-2.45	-2.76	-2.76	-3.01
HON	-2.14	-1.62	-2.05	-2.15	-2.05	HON	-1.80	-1.73	-2.52	-3.42	-3.98
IBM	-2.33	-1.91	-1.55	-1.33	-1.34	IBM	-1.97	-2.07	-2.23	-2.51	-3.09
JPM	-2.41	-2.16	-2.06	-1.93	-1.56	JPM	-2.05	-2.34	-2.85	-3.73	-3.21
KO	-1.76	-1.51	-1.46	-1.46	-1.54	KO	-1.51	-1.63	-2.04	-2.43	-3.17

Notes: DM t-statistic for the mean difference between the OOS loss, where the loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$. Two tables use different number of lags in the estimator of the asymptotic variance-covariance matrix of sample average of loss difference. Given the number of out-of-sample periods n for each forecasting horizon h, the number of lags is $0.75n^{1/3} = 7$ for all h. This value is exceeded by the number of lags $h-1$, except when $h = 5$. Considering a two-sided test, the null hypothesis of equal predictive performance is rejected at 5% significance level if absolute value of t-statistic exceeds 1.96.

Table 11. Out-of-sample performance: DM t-statistic for MEM-GARCH(1,1) model.

	# of lags: $h-1$					# of lags: $0.75n^{1/3}$					
	h = 5	h = 10	h = 22	h = 44	h = 66		h = 5	h = 10	h = 22	h = 44	h = 66
AA	-0.93	-1.56	-1.80	-1.31	-1.32	AA	-0.93	-1.58	-2.04	-2.07	-2.76
AXP	-0.38	-1.13	-1.51	-1.03	-1.34	AXP	-0.37	-1.15	-1.76	-1.47	-2.45
BA	-0.98	-0.59	-0.56	-0.72	-1.11	BA	-0.92	-0.60	-0.63	-1.15	-2.08
CAT	-2.76	-2.08	-1.72	-1.63	-1.66	CAT	-2.58	-2.18	-2.34	-3.10	-3.89
GE	-0.76	-1.58	-1.78	-1.43	-1.46	GE	-0.74	-1.63	-2.09	-2.22	-2.67
HD	-0.89	-1.26	-1.44	-1.28	-1.30	HD	-0.91	-1.31	-1.92	-2.24	-2.73
HON	1.47	-0.15	-0.34	-1.14	-1.44	HON	1.45	-0.15	-0.40	-1.64	-2.28
IBM	-1.26	-1.19	-1.11	-1.19	-1.02	IBM	-1.23	-1.23	-1.43	-2.24	-2.21
JPM	-0.70	-1.02	-1.53	-1.14	-1.28	JPM	-0.68	-1.03	-1.86	-1.75	-2.30
KO	-1.03	-0.88	-1.12	-0.78	-1.01	KO	-0.99	-0.90	-1.36	-1.32	-1.97

Notes: DM t-statistic for the mean difference between the OOS loss, where the loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$. Two tables use different number of lags in the estimator of the asymptotic variance-covariance matrix of sample average of loss difference. Given the number of out-of-sample periods n for each forecasting horizon h, the number of lags is $0.75n^{1/3} = 7$ for all h. This value is exceeded by the number of lags $h-1$, except when $h = 5$. Considering a two-sided test, the null hypothesis of equal predictive performance is rejected at 5% significance level if absolute value of t-statistic exceeds 1.96.

Table 12. Out-of-Sample Performance for cGARCH(1,1) Models.

Standard cGARCH(1,1)						MEM-cGARCH(1,1)					
	h = 5	h = 10	h = 22	h = 44	h = 66		h = 5	h = 10	h = 22	h = 44	h = 66
	$h^e = 4$	$h^e = 4$	$h^e = 3$	$h^e = 2$	$h^e = 2$		$h^e = 2$	$h^e = 10$	$h^e = 12$	$h^e = 12$	$h^e = 12$
AA	0.00	-0.01	-0.02	-0.04	-0.07	AA	0.00	-0.01	-0.02	-0.04	-0.07
	$h^e = 5$	$h^e = 9$	$h^e = 14$	$h^e = 14$	$h^e = 14$		$h^e = 5$	$h^e = 10$	$h^e = 14$	$h^e = 20$	$h^e = 14$
AXP	0.00	-0.01	-0.02	-0.03	-0.05	AXP	0.00	-0.01	-0.02	-0.03	-0.05
	$h^e = 3$	$h^e = 9$	$h^e = 10$	$h^e = 10$	$h^e = 10$		$h^e = 5$	$h^e = 10$	$h^e = 5$	$h^e = 2$	$h^e = 2$
BA	0.00	0.00	-0.01	-0.02	-0.02	BA	0.00	0.00	-0.01	-0.02	-0.02
	$h^e = 4$	$h^e = 4$	$h^e = 4$	$h^e = 4$	$h^e = 3$		$h^e = 5$	$h^e = 10$	$h^e = 12$	$h^e = 12$	$h^e = 12$
CAT	-0.01	-0.01	-0.03	-0.09	-0.15	CAT	-0.01	-0.01	-0.03	-0.09	-0.15
	$h^e = 5$	$h^e = 10$	$h^e = 20$	$h^e = 24$	$h^e = 24$		$h^e = 5$	$h^e = 7$	$h^e = 20$	$h^e = 18$	$h^e 18$
GE	0.00	0.00	-0.01	-0.02	-0.03	GE	0.00	0.00	-0.01	-0.02	-0.03
	$h^e = 5$	$h^e = 10$	$h^e = 6$	$h^e = 6$	$h^e = 6$		$h^e = 4$	$h^e = 10$	$h^e = 16$	$h^e = 26$	$h^e = 28$
HD	0.00	0.00	-0.01	-0.02	-0.04	HD	0.00	0.00	-0.01	-0.02	-0.04
	$h^e = 3$	$h^e = 4$	$h^e = 4$	$h^e = 3$	$h^e = 4$		$h^e = 4$	$h^e = 9$	$h^e = 3$	$h^e = 2$	$h^e = 2$
HON	-0.01	-0.01	-0.01	-0.02	-0.02	HON	-0.01	-0.01	-0.01	-0.02	-0.02
	$h^e = 5$	$h^e = 9$	$h^e = 7$	$h^e = 5$	$h^e = 5$		$h^e = 5$	$h^e = 7$	$h^e = 7$	$h^e = 7$	$h^e = 14$
IBM	0.00	0.00	0.00	0.01	0.02	IBM	0.00	0.00	0.00	0.01	0.02
	$h^e = 3$	$h^e = 3$	$h^e = 3$	$h^e = 2$	$h^e = 2$		$h^e = 5$	$h^e = 10$	$h^e = 22$	$h^e = 16$	$h^e = 14$
JPM	-0.01	-0.02	-0.05	-0.11	-0.15	JPM	-0.01	-0.02	-0.05	-0.11	-0.15
	$h^e = 4$	$h^e = 4$	$h^e = 4$	$h^e = 2$	$h^e = 2$		$h^e = 4$	$h^e = 10$	$h^e = 4$	$h^e = 2$	$h^e = 2$
KO	0.00	0.00	0.00	0.01	0.02	KO	0.00	0.00	0.00	0.01	0.02

Notes: Out-of-sample average loss difference for QLIKE loss function for 10 individual stocks. Loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$, where L denotes the QLIKE loss evaluated at the QML estimator $\hat{\theta}^d_T$ and at the HM estimator $\hat{\theta}^c_T$. Considered forecasting horizons are $h \in \{5, 10, 22, 44, 66\}$. Horizon at which $\hat{\theta}^c_T$ is obtained does not necessarily match forecasting (evaluation) horizon. For each forecasting horizon and individual stock we report the estimation horizon (h^e) which minimizes the average QLIKE loss evaluated at the HM estimator.

Table 13. Out-of-sample performance: DM t-statistic for standard cGARCH(1,1) model.

# of lags : $h-1$						# of lags: $0.75n^{1/3}$					
	h = 5	h = 10	h = 22	h = 44	h = 66		h = 5	h = 10	h = 22	h = 44	h = 66
AA	-2.14	-1.91	-1.97	-1.81	-1.37	AA	-1.88	-2.01	-2.58	-2.43	-1.73
AXP	-2.81	-2.52	-1.76	-1.74	-2.05	AXP	-2.49	-2.69	-2.37	-2.91	-4.07
BA	-1.85	-1.58	-0.97	-0.96	-0.86	BA	-1.66	-1.71	-1.42	-1.80	-1.83
CAT	-3.30	-2.73	-2.25	-2.49	-2.77	CAT	-2.90	-2.91	-2.96	-3.73	-3.67
GE	-2.88	-1.94	-1.37	-1.24	-1.37	GE	-2.64	-2.08	-1.96	-2.42	-3.03
HD	-2.96	-2.21	-2.32	-0.97	-0.79	HD	-2.53	-2.35	-2.88	-1.78	-1.67
HON	0.62	-0.69	-0.21	0.86	0.98	HON	0.68	-0.64	-0.23	1.42	1.66
IBM	-2.76	-2.50	-2.26	-1.34	-0.62	IBM	-2.66	-2.51	-2.70	-2.01	-1.15
JPM	-1.81	-1.59	-1.62	-1.97	-1.98	JPM	-1.53	-1.73	-2.30	-3.53	-4.13
KO	-1.44	-1.47	-0.99	-1.61	-1.43	KO	-1.27	-1.58	-1.42	-2.58	-2.53

Notes: DM t-statistic for the mean difference between the OOS loss, where the loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$. Two tables use different number of lags in the estimator of the asymptotic variance-covariance matrix of sample average of loss difference. Given the number of out-of-sample periods n for each forecasting horizon h, the number of lags is $0.75n^{1/3} = 7$ for all h. This value is exceeded by the number of lags $h-1$, except when $h = 5$. Considering a two-sided test, the null hypothesis of equal predictive performance is rejected at 5% significance level if absolute value of t-statistic exceeds 1.96.

Table 14. Out-of-sample performance: DM t-statistic for MEM-cGARCH(1,1) model.

	# of lags: $h-1$						# of lags: $0.75n^{1/3}$				
	h = 5	h = 10	h = 22	h = 44	h = 66		h = 5	h = 10	h = 22	h = 44	h = 66
AA	-0.53	-1.69	-1.65	-1.60	-1.54	AA	-0.51	-1.78	-2.36	-2.99	-3.41
AXP	-1.59	-1.98	-1.91	-2.02	-2.12	AXP	-1.53	-2.10	-2.58	-3.52	-4.37
BA	-1.00	-1.48	-1.37	-1.55	-1.53	BA	-0.96	-1.57	-1.90	-2.72	-3.22
CAT	-2.29	-2.11	-1.93	-1.63	-1.57	CAT	-2.12	-2.22	-2.54	-2.71	-3.23
GE	-0.97	-1.00	-1.11	-1.04	-1.13	GE	-0.92	-1.05	-1.53	-1.93	-2.38
HD	-1.25	-1.24	-1.27	-1.09	-1.08	HD	-1.17	-1.32	-1.64	-1.82	-2.19
HON	-1.81	-1.88	-1.39	-1.38	-1.36	HON	-1.70	-1.99	-1.90	-2.33	-2.65
IBM	-0.48	-0.67	-0.17	0.56	0.99	IBM	-0.46	-0.71	-0.24	1.07	2.26
JPM	-2.73	-2.11	-1.95	-2.15	-2.07	JPM	-2.48	-2.23	-2.49	-3.13	-3.46
KO	-0.82	-0.59	0.28	1.54	1.91	KO	-0.75	-0.62	0.39	2.81	3.88

Notes: DM t-statistic for the mean difference between the OOS loss, where the loss difference is defined as $\Delta L = L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^c_T)) - L(\tilde{x}_{t,h}, \tilde{\sigma}^2_{t,h}(\hat{\theta}^d_T))$. Two tables use different number of lags in the estimator of the asymptotic variance-covariance matrix of sample average of loss difference. Given the number of out-of-sample periods n for each forecasting horizon h, the number of lags is $0.75n^{1/3} = 7$ for all h. This value is exceeded by the number of lags $h-1$, except when $h = 5$. Considering a two-sided test, the null hypothesis of equal predictive performance is rejected at 5% significance level if absolute value of t-statistic exceeds 1.96.

REFERENCES

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

Andersen, T. G., Bollerslev, T., & Meddahi, N. (2004). Analytical evaluation of volatility forecasts. *International Economic Review*, *45*(4), 1079–1110.

Andrews, D. W. (1987). Asymptotic results for generalized Wald tests. *Econometric Theory*, *3*(3), 348–358.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*(3), 817–858.

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal*, *12*(3), 1–32.

Bauwens, L., Hafner, C. M., & Laurent, S. (2012). *Handbook of Volatility Models and Their Applications* (Vol. 3). John Wiley & Sons.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T., & Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, *73*(1), 151–184.

Chung, C.-F. (1999). Estimating the fractionally integrated GARCH model. *National Taiwan University Working Paper*.

Conrad, C., & Haag, B. R. (2006). Inequality constraints in the fractionally integrated GARCH model. *Journal of Financial Econometrics*, *4*(3), 413–449.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

De Nicolò, G., & Lucchetta, M. (2017). Forecasting tail risks. *Journal of Applied Econometrics*, *32*(1), 159–170.

Ding, Z., & Granger, C. W. J. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, *73*(1), 185–215.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

Engle, R. F., & Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, *131*(1-2), 3–27.

Engle, R. F., & Lee, G. (1999). A long-run and short-run component model of stock return volatility. In R. F. Engle & H. White (Eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger* (pp. 475–497). Oxford University Press.

Ghysels, E., Plazzi, A., Valkanov, R., Rubia, A., & Dossani, A. (2019). Direct versus iterated multiperiod volatility forecasts. *Annual Review of Financial Economics*, *11*, 173–195.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.

Gorgi, P., Hansen, P., Janus, P., & Koopman, S. J. (2019). Realized Wishart-GARCH: A score-driven multi-asset volatility model. *Journal of Financial Econometrics*, *17*(1), 1–32.

Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operational Research Quarterly*, *20*(2), 199–207.

Granger, C. W. J. (1993). On the limitations of comparing mean square forecast errors: Comment. *Journal of Forecasting*, *12*(8), 651–652.

Granger, C. W. J., & Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, *73*(1), 61–77.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, *20*(7), 873–889.

Hansen, P. R., & Dumitrescu, E.-I. (2022). How should parameter estimation be tailored to the objective? *Journal of Econometrics*, *230*(2), 535–558.

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*(6), 877–906.

Hansen, P. R., & Lunde, A. (2011). Forecasting volatility using high frequency data. In M. P. Clements & D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting* (pp. 525–556.). Oxford University Press.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251–1271.

Karanasos, M., Psaradakis, Z., & Sola, M. (2004). On the autocorrelation properties of long-memory GARCH processes. *Journal of Time Series Analysis*, *25*(2), 265–282.

Kole, E., Markwat, T., Opschoor, A., & Van Dijk, D. (2017). Forecasting Value-at-Risk under temporal and portfolio aggregation. *Journal of Financial Econometrics*, *15*(4), 649–677.

Mancini, L., & Trojani, F. (2011). Robust Value at Risk prediction. *Journal of Financial Econometrics*, *9*(2), 281–313.

Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM review*, *10*(4), 422–437.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1-2), 499–526.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Oh, D. H., & Patton, A. J. (2024). Better the devil you know: Improved forecasts from imperfect models. *Journal of Econometrics*, *242*(1), 105767.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Patton, A. J. (2020). Comparing possibly misspecified forecasts. *Journal of Business & Economic Statistics*, *38*(4), 796–809.

Pesaran, M. H., Pick, A., & Timmermann, A. (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics*, *164*(1), 173–187.

Ruiz, E., & Nieto, M. R. (2023). Direct versus iterated multiperiod Value-at-Risk forecasts. *Journal of Economic Surveys*, *37*(3), 915–949.

Shephard, N., & Sheppard, K. (2010). Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, *25*(2), 197–231.

Sizova, N. (2011). Integrated variance forecasting: Model based vs. reduced form. *Journal of Econometrics*, *162*(2), 294–311.

West, K. D. (1997). Another heteroskedasticity-and autocorrelation-consistent covariance matrix estimator. *Journal of Econometrics*, *76*(1-2), 171–191.

White, H. (1996). *Estimation, Inference and Specification Analysis*. Cambridge University Press.