Bootstrap based probability forecasting in multiplicative error models

Indeewara Perera

Department of Economics, The University of Sheffield, Sheffield S1 4DT, UK

i.perera@sheffield.ac.uk

Mervyn J. Silvapulle

Department of Econometrics and Business Statistics, Monash University, Australia

mervyn.silvapulle@monash.edu.au

Abstract

As evidenced by an extensive empirical literature, multiplicative error models (MEM) show good performance in capturing the stylized facts of nonnegative time series; examples include, trading volume, financial durations, and volatility. This paper develops a bootstrap based method for producing multi-step-ahead probability forecasts for a nonnegative valued time-series obeying a parametric MEM. In order to test the adequacy of the underlying parametric model, a class of bootstrap specification tests is also developed. Rigorous proofs are provided for establishing the validity of the proposed bootstrap methods. The paper also establishes the validity of a bootstrap based method for producing probability forecasts in a class of semiparametric MEMs. Monte Carlo simulations suggest that our methods perform well in finite samples. A real data example illustrates the methods.

JEL Classifications: C12, C52, C53.

Keywords: Multiplicative error model; Bootstrap; Probability forecast; Goodness-of-fit; Multi-step forecast.

Corresponding author: Indeewara Perera, Department of Economics, The University of Sheffield, Sheffield S1 4DT, UK; e-mail: i.perera@sheffield.ac.uk

1 INTRODUCTION AND MOTIVATION

Statistical models for non-negative random variables have been used in many areas, including finance, economics, health sciences, and engineering. In finance, the family of multiplicative error models plays a key role in modelling non-negative valued time series processes (Engle, 2002; Pacurar, 2008). For example, they have been used for modelling financial durations (Engle and Russell, 1998; Allen *et al.*, 2008; Gao *et al.*, 2015), trading volume of orders (Manganelli, 2005), high-low range of asset prices (Chou, 2005), spikes in electricity price (Christensen *et al.*, 2012), absolute value of daily returns (Engle and Gallo, 2006), and realized volatility (Brownlees *et al.*, 2012). This paper develops new methodology for producing multi-step ahead probability forecasts for a nonnegative valued time-series obeying a multiplicative error model. To complement the proposed methods a class of specification tests is also proposed.

Let $\{Z_i : i \in \mathbb{Z}\}$ denote a series of nonnegative random variables, for example realized volatility, where $\mathbb{Z} := \{0, \pm 1, \pm 2, \cdots\}$. A *multiplicative error model* [MEM] takes the form,

$$Z_i = \Psi_i \varepsilon_i, \tag{1}$$

where Ψ_i is a function of the past information set at time i, denoted $\mathcal{H}_{i-1}$, $\{\varepsilon_i\}$ are independent and identically distributed [iid] with unit mean and finite variance, and ε_i is independent of $\mathcal{H}_{i-1}$ $(i \in \mathbb{Z})$; hence Ψ_i is identified as the conditional mean of Z_i given $\mathcal{H}_{i-1}$ $(i \in \mathbb{Z})$. Let F_0 denote the distribution function of ε_i. Broadly, the objective of this paper is to develop new methods for forecasting in MEMs. More specifically, we are interested in constructing probability forecasts for future values of Z_{n+k} and Ψ_{n+k}, for some positive integer k; for example, multi-step-ahead interval forecasts for Z_{n+k} or prediction bounds for Ψ_{n+k}, assuming that the observations $\{Z_1, \ldots, Z_n\}$ up to the current time n are available.

The existing methods for forecasting in MEMs are mainly based on point forecasts (see Engle and Russell, 1997; Dufour and Engle, 2000; Bauwens and Giot, 2001; Hautsch, 2011; Luca *et al.*, 2017) and evaluating one-step-ahead density forecasts (Bauwens *et al.*, 2004; Corsi *et al.*, 2008; Hautsch *et al.*, 2014). Although such methods have been widely used in empirical studies (see Bauwens and Giot, 2000; Fernandes and Grammig, 2005; Engle and Gallo, 2006; Corsi *et al.*, 2008; Gao *et al.*, 2015 and references there in), the literature is scant on methods for interval forecasts or multi-step-ahead density/distribution forecasts. However, in risk management, for example in managing financial investments, one needs to take into account of the forecast of not only the very next observation, but also of those observations that are several steps ahead. Hence, our focus on multi-step-ahead forecasting in MEMs is of interest.

This paper develops bootstrap based methodology for producing probability forecasts in MEMs by considering two distinct approaches under different assumptions: (A) *Semiparametric approach:* we assume a parametric model for Ψ_i but not for F_0, and develop a semiparametric method for forecasting, which includes a semiparametric component for estimating F_0, (B) *Parametric approach:* in addition to the parametric model for Ψ_i in the previous approach, we also assume a parametric model for F_0, and develop a fully parametric method for probability forecasting. To complete the methodology, we also develop a method for testing the adequacy of the parametric specifications of Ψ_i and F_0. Depending on the nature of the application of interest, each of the aforementioned parametric and semiparametric approaches have advantages and disadvantages, but none would be uniformly better than the other. For example, under the correct specification of the MEM, the parametric approach is expected to be more efficient than the semiparametric approach. However, in some empirical applications, it is of interest to produce probability forecasts that rely only on the specification of the conditional mean but do not require any parametric assumptions on the error distribution. Hence, the semi parametric approach is also of interest. Thus, it is of interest to develop the methodology for both these approaches. Several bootstrap based methods that use a semiparametric approach similar to (A) have been considered for producing probability forecasts in additive and conditionally heteroscedastic time series models; see Christoffersen and Goncalves (2005), Pascual *et al.* (2006), Chen *et al.* (2011), Mancini and Trojani (2011), and Mazzeu *et al.* (2017), amongst others. However, the validity of bootstrap based probability forecasting, in the context of MEMs, has not been discussed so far in the literature.

In summary, this paper makes the following main methodological contributions: (1) First, we propose and establish the validity of a bootstrap method in parametric MEMs for constructing multi-step ahead probability forecasts, including distributional forecasts, for Z_{n+k} and Ψ_{n+k}, conditional on $\{Z_1, \ldots, Z_n\}$ ($k = 2, 3, \ldots$). (2) We extend the aforementioned parametric method to a semiparametric MEM that does not specify a parametric form for F_0. (3) To support the parametric method, we develop a bootstrap based testing procedure for fitting a parametric MEM. We state the regularity conditions used in the proof of each method, and provide rigorous proofs for the validity of each of the aforementioned three methods. To demonstrate that our regularity conditions are reasonable, we also show that they are satisfied by MEM(p_1, p_2), which is perhaps the most widely used MEM.

One attractive feature of the proposed methods is that they have intuitively appealing simple structure, and are easy to program. The methods also perform well in an extensive simulation study. Further, in an illustrative real data example, the probability forecasts produced by the semiparametric method perform reasonably well, and those produced by the parametric method perform better. These results indicate that the methods we develop for probability forecasting are of practical interest.

The rest of this paper is structured as follows. Section 2 formulates the problem, defines the probability forecasts and test statistics, and provides several results relating to the asymptotic validity of the proposed methods. Section 3 describes a simulation study. A real data example involving daily annualized realized volatility constructed from intraday spot price data for the S&P500 index, is provided in Section 4. Section 5 concludes the paper. Appendix A contains some of the main proofs. The details of the simulation results, additional figures for the empirical example, and some of the omitted proofs are provided in an online supplementary material.

2 MAIN RESULTS

Let the class of MEMs and the definitions of $Z_i, \varepsilon_i, \mathcal{H}_i$, and F_0 be as in the Introduction. Let p and q be known positive integers. Let the transpose of any vector or matrix be denoted by the superscript "$\top$". We say that a sequence of random variables $\{X_i : i = 1, 2, \cdots\}$ converge *exponentially almost surely* to zero, denoted $X_i \overset{e.a.s.}{\rightarrow} 0$, if there exist a $\gamma > 1$ such that $\gamma^i X_i \to 0$ almost surely (a.s.) as $i \to \infty$. Let $\mathcal{F} = \{F_\theta : \theta \in \Theta \subset \mathbb{R}^q\}$ be a given family of distribution functions, where F_θ has mean 1, variance $0 < \sigma_\theta^2 < \infty$, and almost everywhere (a.e.) positive density f_θ. Let $\mathbb{R}^+ := [0, \infty)$. Let $\mathcal{A} = \{\Psi_i(\phi) : \phi \in \Phi \subset \mathbb{R}^p\}$ be a given parametric family defined by

$$\Psi_i(\phi) := g_\phi(Z_{i-1}, \cdots, Z_{i-p_1}, \Psi_{i-1}, \cdots, \Psi_{i-p_2}), \quad \phi \in \Phi, \tag{2}$$

where $\{g_\phi, \phi \in \Phi\}$ denotes a parametric family of nonnegative functions on $(\mathbb{R}^+)^{p_1+p_2}$ with $p_1, p_2 \geq 0$ being known integers. In many parametric MEMs, the conditional mean $\Psi_i := \mathbb{E}(Z_i \mid \mathcal{H}_{i-1})$ can be written in the general form in (2). For example, the linear MEM of Engle and Russell (1998), denoted MEM(p_1, p_2), is given by $\Psi_i(\phi) = \alpha + \sum_{j=1}^{p_1} \beta_j Z_{i-j} + \sum_{j=1}^{p_2} \gamma_j \Psi_{i-j}(\phi)$, where $\phi = (\alpha, \beta_1, \ldots, \beta_{p_1}, \gamma_1, \ldots, \gamma_{p_2})^\top$.

Let us assume that $\Phi(\subset \mathbb{R}^p)$ and $\Theta(\subset \mathbb{R}^q)$ are compact subsets, and

$$(\Psi_i, F_0) \in \mathcal{A} \times \mathcal{F} \tag{3}$$

with the true parameter vector, denoted $(\phi_0^\top, \theta_0^\top)^\top$, being an interior point of $\Phi \times \Theta$.

Let $\{Z_1, \ldots, Z_n\}$ be a sequence of n observations generated by the MEM given by (1), (2) and (3); this defines a parametric form for the entire conditional distribution of Z_i given $\mathcal{H}_{i-1}$. Our objective is to obtain asymptotically valid probability forecasts for Z_{n+k} and Ψ_{n+k}, for a given integer $k > 1$, conditional on $\{Z_1, \ldots, Z_n\}$.

Let us first outline the method of estimating the model defined by (1), (2) and (3). Note that, the function $\Psi_i(\phi)$ in (2) depends on the unobserved part of the process $\{\dots, Z_{-1}, Z_0\}$ extending back to the infinite past. For example, in the MEM(1,1) model, we have $\Psi_i(\phi) = \alpha(1-\gamma)^{-1} + \beta\sum_{j=1}^{\infty}\gamma^{j-1}Z_{i-j}$. Hence, to approximate $\Psi_i(\phi)$ based on $\{Z_1, \dots, Z_n\}$, we use a sample estimate, denoted $\widetilde{\Psi}_i(\phi)$, obtained by assuming $(Z_0, \cdots, Z_{1-p_1}, \Psi_0, \cdots, \Psi_{1-p_2})^\top = \varsigma_0$, where $\varsigma_0 = (z_0, \cdots, z_{1-p_1}, s_0, \cdots, s_{1-p_2})^\top$ is a set of suitable starting values in $(\mathbb{R}^+)^{p_1+p_2}$; it will be shown later that the effect due to the choice of the starting values is asymptotically negligible.

In view of Proposition 3.12 of Straumann and Mikosch (2006), if the family $\{g_\phi\}$ satisfies certain random coefficient Lipschitz conditions, then, irrespective of the starting value ς_0,

$$\sup_{\phi\in\Phi} |\widetilde{\Psi}_i(\phi) - \Psi_i(\phi)| \overset{e.a.s.}{\to} 0, \quad i \to \infty, \tag{4}$$

where $\overset{e.a.s.}{\to} 0$ denotes the *exponential almost sure convergence* to 0, as defined earlier. In the case of the MEM(1,1), a sufficient condition for (4) is, $\mathbb{E}\sup_{\phi\in\Phi}[\log(\beta\varepsilon_1 + \gamma)] < 0$. A number of other MEMs, including the MEM(p_1, p_2) of Engle and Russell (1998), also satisfy (4), under similar conditions.

Since (1), (2) and (3) specify a fully parametric MEM, it is appropriate to estimate the model by maximum likelihood (ML). To this end, let $\zeta = (\phi^\top, \theta^\top)^\top$ denote an arbitrary point in $\Phi \times \Theta$, and let $\zeta_0 = (\phi_0^\top, \theta_0^\top)^\top$ be the true value in (3). Let

$$\widetilde{\chi}_i(\zeta) := \log\widetilde{\Psi}_i(\phi) - \log\left[f_\theta\{Z_i/\widetilde{\Psi}_i(\phi)\}\right] \quad \text{and} \quad \widetilde{\Upsilon}_n(\zeta) := \sum_{i=1}^{n} -\widetilde{\chi}_i(\zeta),$$

where f_θ denotes the probability density function [pdf] of the distribution function F_θ. Consider the estimator $\hat{\zeta} = (\hat{\phi}^\top, \hat{\theta}^\top)^\top$ of ζ_0 defined by

$$\hat{\zeta} = \arg\max_{\zeta\in\Phi\times\Theta}\widetilde{\Upsilon}_n(\zeta). \tag{5}$$

This is not the true ML estimator, because the objective function $\widetilde{\Upsilon}_n$ is based on $\widetilde{\Psi}_i$ instead of the true conditional mean function Ψ_i. Therefore, we refer to the estimator $\hat{\zeta}$ as the *approximate maximum likelihood estimator (AMLE)*. The true likelihood function that corresponds to Ψ_i is defined by

$$\Upsilon_n(\zeta) := \sum_{i=1}^{n} -\chi_i(\zeta), \quad \chi_i(\zeta) := \log\Psi_i(\phi) - \log\left[f_\theta\{Z_i/\Psi_i(\phi)\}\right]. \tag{6}$$

The summands in (6) are stationary and, under (4), $n^{-1}\widetilde{\Upsilon}_n$ approximates $n^{-1}\Upsilon_n$ with an error decaying to zero as $n \to \infty$. By using these properties, and several regularity conditions, we show that $\hat{\zeta}$ is consistent and asymptotically normal; see Appendix A.1.

In the next subsection we introduce the bootstrap method that we propose for producing probability forecasts for MEMs by using $\Psi_i(\phi)$, F_θ, and the estimator $\hat{\zeta}$.

2.1 Bootstrap probability forecasts: parametric approach

In this subsection we continue to assume the parametric MEM specified by (1), (2) and (3). In order to produce probability forecasts for Z_{n+k} and Ψ_{n+k} conditional on the realized values $\{Z_1, \ldots, Z_n\}$, we propose the following bootstrap procedure:

Step 1: Compute the estimate $\hat{\zeta} = (\hat{\phi}^\top, \hat{\theta}^\top)^\top$ using the observed sample $\{Z_1, \ldots, Z_n\}$.
Step 2: Generate $m + n + 1$ independent observations, $\varepsilon^*_{-m}, \ldots, \varepsilon^*_n$ from $F_{\hat{\theta}}$.
Step 3: Generate $Z^*_{-m}, \ldots, Z^*_n$ recursively by

$$Z^*_i = \Psi^*_i(\hat{\phi})\varepsilon^*_i, \quad \Psi^*_i(\phi) := g_\phi(Z^*_{i-1}, \cdots, Z^*_{i-p_1}, \Psi^*_{i-1}(\phi), \cdots, \Psi^*_{i-p_2}(\phi)), \quad \phi \in \Phi,$$

$i = -m, \ldots, n$, with $(Z^*_{-m-1}, \cdots, Z^*_{-m-p_1}, \Psi^*_{-m-1}, \cdots, \Psi^*_{-m-p_2})^\top = \varsigma_0$.
Step 4: Discard $\{Z^*_{-m}, \ldots, Z^*_0\}$, and use $\{Z^*_1, \ldots, Z^*_n\}$ as the bootstrap sample.
Step 5: Compute the bootstrap analogue $\hat{\zeta}^* = (\hat{\phi}^*, \hat{\theta}^*)^\top$ of $\hat{\zeta}$ based on $\{Z^*_1, \ldots, Z^*_n\}$, and generate k independent observations, $\varepsilon^*_{n+1}, \ldots, \varepsilon^*_{n+k}$ from $F_{\hat{\theta}}$.
Step 6: Generate $\check{Z}^*_{n+k}$ and $\check{\Psi}^*_{n+k}(\hat{\phi}^*)$ recursively, by

$$\begin{aligned}
&\check{Z}^*_{n+i} = \check{\Psi}^*_{n+i}(\hat{\phi}^*)\varepsilon^*_{n+i}, \quad i = 1, \ldots, k, \\
&\check{\Psi}^*_{n+i}(\phi) := g_\phi(\check{Z}^*_{n+i-1}, \cdots, \check{Z}^*_{n+i-p_1}, \check{\Psi}^*_{n+i-1}(\phi), \cdots, \check{\Psi}^*_{n+i-p_2}(\phi)), \quad \phi \in \Phi,
\end{aligned}$$

with $(\check{Z}^*_n, \cdots, \check{Z}^*_{n+1-p_1}, \check{\Psi}^*_n, \cdots, \check{\Psi}^*_{n+1-p_2})^\top = (Z_n, \cdots, Z_{n+1-p_1}, \widetilde{\Psi}_n, \cdots, \widetilde{\Psi}_{n+1-p_2})^\top$.
Step 7: Repeat Steps 2–6, B number of times and obtain the bootstrap replicates, $\{\check{Z}^{*(1)}_{n+k}, \ldots, \check{Z}^{*(B)}_{n+k}\}$ for $\check{Z}^*_{n+k}$, and $\{\check{\Psi}^{*(1)}_{n+k}(\hat{\phi}^*), \ldots, \check{\Psi}^{*(B)}_{n+k}(\hat{\phi}^*)\}$ for $\check{\Psi}^*_{n+k}$.
Step 8: Produce the required probability forecasts by using the empirical distributions of $\{\check{Z}^{*(1)}_{n+k}, \ldots, \check{Z}^{*(B)}_{n+k}\}$ and $\{\check{\Psi}^{*(1)}_{n+k}(\hat{\phi}^*), \ldots, \check{\Psi}^{*(B)}_{n+k}(\hat{\phi}^*)\}$; for example, with

$$G_{B,\check{Z}^*,k}(x) := \frac{1}{B}\sum_{b=1}^{B} I\big\{\check{Z}^{*(b)}_{n+k} \le x\big\} \quad \text{and} \quad G_{B,\check{\Psi}^*,k}(x) := \frac{1}{B}\sum_{b=1}^{B} I\big\{\check{\Psi}^{*(b)}_{n+k}(\hat{\phi}^*) \le x\big\}, \tag{7}$$

where $I(\cdot)$ denotes the indicator function, the $100(1-\omega)\%$ prediction intervals for Z_{n+k} and Ψ_{n+k}, for $0 < \omega < 1$, can be obtained as

$$[G^{-1}_{B,\check{Z}^*,k}(\omega/2), G^{-1}_{B,\check{Z}^*,k}(1-\omega/2)] \quad \text{and} \quad [G^{-1}_{B,\check{\Psi}^*,k}(\omega/2), G^{-1}_{B,\check{\Psi}^*,k}(1-\omega/2)], \tag{8}$$

respectively, and $100(1-\omega)\%$ upper prediction bounds for Z_{n+k} and Ψ_{n+k} are given by

$$[0, G^{-1}_{B,\check{Z}^*,k}(1-\omega)] \quad \text{and} \quad [0, G^{-1}_{B,\check{\Psi}^*,k}(1-\omega)], \quad 0 < \omega < 1. \tag{9}$$

Later we show that the bootstrap method outlined in Steps 1–8 is asymptotically valid, and that $G_{B,\check{Z}^*,k}(x)$ and $G_{B,\check{\Psi}^*,k}(x)$ in (7) mimic the conditional distributions of Z_{n+k} and Ψ_{n+k} in large samples. In view of this, one may use $\{\check{Z}^{*(1)}_{n+k}, \ldots, \check{Z}^{*(B)}_{n+k}\}$

and $\{\breve{\Psi}^{*(1)}_{n+k}(\hat{\phi}^*),\ldots,\breve{\Psi}^{*(B)}_{n+k}(\hat{\phi}^*)\}$, to produce various probability forecasts, and to approximate the pdfs (i.e. to produce density forecasts) of Z_{n+k} and Ψ_{n+k}, conditional on $\{Z_1,\ldots,Z_n\}$. In this study, we illustrate the details for producing interval forecasts for Z_{n+k} and Ψ_{n+k} as in (8) and (9). Similarly, other measures of uncertainty related to Z_{n+k} and Ψ_{n+k}, conditional on $\{Z_1,\ldots,Z_n\}$, may also be produced; useful examples include value-at-risk, expected utility, quantile forecasts, and sharpe-ratio.

The validity of the bootstrap method we propose by Steps 1–8 depends on the conditional distribution of Z_i specified by $\Psi_i(\phi)$ and F_θ. Therefore, prior to applying the proposed method it is important to test the goodness-of-fit of (1), (2) and (3). For this purpose, we develop a bootstrap testing procedure in Section 2.4. In the next subsection, we propose a method for producing probability forecasts by using a semiparametric approach which does not assume a parametric form for F_0.

2.2 Bootstrap probability forecasts: semiparametric approach (based on the empirical distribution of the residuals)

The probability forecasts in Section 2.1 require parametric specifications of both the conditional mean Ψ_i and the error distribution F_0. However, in some empirical applications, it is of interest to produce probability forecasts that rely only on the specification of the conditional mean but do not require any parametric assumptions on the error distribution. In fact, there are several tests available in the literature for fitting a parametric model for Ψ_i without making any parametric assumptions on F_0 (see Hidalgo and Zaffaroni, 2007; Koul *et al.*, 2012; Perera and Koul, 2017). Therefore, in this section, we propose a method to produce probability forecasts by using bootstrap replicates obtained from the empirical distribution of the standardized residuals instead of using the parametric distribution $F_{\hat{\theta}}$. This method can be viewed as a variant of the procedure previously proposed by Pascual *et al.* (2006) in the context of GARCH models and implemented by Mazzeu *et al.* (2017). The key idea of this bootstrap method is to use an empirical distribution instead of a parametric one, and hence not rely on any parametric assumptions on the error distribution.

For the rest of this subsection, we assume the semiparametric MEM specified by (1) and (2); therefore, F_0 may not be of the form F_θ. We continue to denote the true value of ϕ by ϕ_0. To estimate ϕ_0 we use the *quasi maximum likelihood estimator (QMLE)* based on the standard exponential distribution, defined by

$$\hat{\phi}_{qml} = \arg\min_{\phi\in\Phi}\sum_{i=1}^{n}\ell_i(\phi), \quad \ell_i(\phi) = \log\widetilde{\Psi}_i(\phi) + Z_i/\widetilde{\Psi}_i(\phi). \tag{10}$$

The residuals are defined as $\tilde{\varepsilon}_i^{(qml)} = Z_i/\widetilde{\Psi}_i(\hat{\phi}_{qml}),\ i = 1,\ldots,n$. We chose the QMLE in (10) because it provides consistent estimates, without relying on any parametric assumptions on F_0 (see Drost and Werker, 2004). Further, $\hat{\phi}_{qml}$ is usually asymptotically linear for many models (see Lee and Hansen, 1994; Berkes and Horváth, 2004; Francq and Zakoïan, 2004; Straumann and Mikosch, 2006; Ling, 2007), and thus, there exist some constant $c \neq 0$ and a function ϱ such that $\mathbb{E}\varrho(\varepsilon_1) = 0$, $\mathbb{E}\varrho^2(\varepsilon_1) < \infty$ and

$$n^{1/2}(\hat{\phi}_{qml} - \phi_0) = c\Sigma^{-1} n^{-1/2} \sum_{i=1}^{n} \lambda_i(\phi_0)\varrho(\varepsilon_i) + o_p(1), \tag{11}$$

where $\lambda_i(\phi) := \dot{\Psi}_i(\phi)/\Psi_i(\phi)$ and $\Sigma := \mathbb{E}\{\lambda_1(\phi_0)\lambda_1^\top(\phi_0)\}$. The asymptotic linearity of $\hat{\phi}_{qml}$ in (11) is used for establishing the validity of the probability forecasts.

QMLE versus AMLE

The quasi maximum likelihood (QML) estimation based on distributions belonging to the standard gamma family (with two parameters), such as the exponential, provide consistent estimates under the correct specification of the conditional mean, without relying on any parametric assumptions on F_0 (see Drost and Werker, 2004). Therefore, if F_θ is misspecified then the QMLE is likely to perform better than the AMLE. However, under the correct specification of F_θ, the AMLE is more efficient than the QMLE (see Propositions 1 and 2 in Appendix A.1). Further, in some cases, the QMLE based estimation of MEMs may perform quite poorly, even with quite large samples (see Grammig and Maurer, 2000). Therefore, if the parametric model is 'close' to the true model then it is desirable to use the AMLE instead of the QMLE.

Bootstrap using the empirical distribution of the standardized residuals

First, we define the standardized residuals as

$$\tilde{\varepsilon}_t^{(std)} = \left\{ n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^{(qml)} \right\}^{-1} \hat{\varepsilon}_t^{(qml)}, \quad t = 1, \ldots, n, \tag{12}$$

so that the empirical distribution of $\{\tilde{\varepsilon}_t^{(std)}\}_{t=1}^n$ has mean 1. This is important because one of the model assumptions is that $\mathbb{E}(\varepsilon_0) = 1$. The effect of (12) is similar to that of centering the residuals before resampling in additive regression models.

Step 1: Compute the QMLE $\hat{\phi}_{qml}$ using the observed sample $\{Z_1, \ldots, Z_n\}$.
Step 2: Draw a random sample (with replacement) of size $m+n+1$, say $\varepsilon_{-m}^{q*}, \ldots, \varepsilon_n^{q*}$, from $\{\tilde{\varepsilon}_t^{(std)}; 1 \leq t \leq n\}$.
Step 3: Generate $Z_{-m}^{q*}, \ldots, Z_n^{q*}$ recursively by

$$Z_i^{q*} = \Psi_i^{q*}(\hat{\phi}_{qml})\varepsilon_i^{q*}, \quad \Psi_i^{q*}(\hat{\phi}_{qml}) = g_{\hat{\phi}_{qml}}(Z_{i-1}^{q*}, \cdots, Z_{i-p_1}^{q*}, \Psi_{i-1}^{q*}(\hat{\phi}_{qml}), \cdots, \Psi_{i-p_2}^{q*}(\hat{\phi}_{qml})),$$

$i = -m, \ldots, n$, with $(Z^{q*}_{-m-1}, \cdots, Z^{q*}_{-m-p_1}, \Psi^{q*}_{-m-1}(\hat{\phi}_{qml}), \cdots, \Psi^{q*}_{-m-p_2}(\hat{\phi}_{qml}))^\top = \varsigma_0$.
Step 4: Discard $\{Z^{q*}_{-m}, \ldots, Z^{q*}_0\}$, and use $\{Z^{q*}_1, \ldots, Z^{q*}_n\}$ as the bootstrap sample.
Step 5: Compute the bootstrap analogue $\hat{\phi}^*_{qml}$ of $\hat{\phi}_{qml}$ based on $\{Z^{q*}_1, \ldots, Z^{q*}_n\}$, and draw a random sample of size k, say $\varepsilon^{q*}_{n+1}, \ldots, \varepsilon^{q*}_{n+k}$, from $\{\tilde{\varepsilon}^{(std)}_t; 1 \le t \le n\}$.
Step 6: Repeat the Steps 6–8 of the bootstrap procedure in Section 2.1 by using $\{Z^{q*}_1, \ldots, Z^{q*}_n\}$ and $\{\varepsilon^{q*}_{n+1}, \ldots, \varepsilon^{q*}_{n+k}\}$, instead of $\{Z^*_1, \ldots, Z^*_n\}$ and $\{\varepsilon^*_{n+1}, \ldots, \varepsilon^*_{n+k}\}$, and obtain the analogues of the probability forecasts in (7), (8) and (9) for the above setup.

This procedure can be viewed as a variant of the bootstrap method previously proposed by Pascual *et al.* (2006) in the context of GARCH models. In Section 2.3, we show that the bootstrap method outlined in Steps 1–6 above is asymptotically valid.

2.3 Asymptotic validity of the probability forecasts

In this section, we first show that the probability forecasts of Z_{n+k} and Ψ_{n+k} given by (8) and (9) are asymptotically valid. To this end, we assume that there exists a compact neighbourhood Λ of ϕ_0 and a stationary process $\{\Psi^{**}_i\}_{i\in\mathbb{Z}}$, such that for every integer $r \ge 0$,

$$\sup_{\phi\in\Lambda} |\check{\Psi}^*_{n+r}(\phi) - \Psi^{**}_{n+r}(\phi)| \overset{e.a.s.}{\to} 0, \quad \text{as } n \to \infty, \tag{13}$$

in probability. Note that, due to the effect of initial conditions and the use of $\widetilde{\Psi}_i$ instead of the unobservable Ψ_i, the bootstrap processes $\{\check{Z}^*_{n+i}\}_{i\in\mathbb{N}}$ and $\{\check{\Psi}^*_{n+i}\}_{i\in\mathbb{N}}$ used in (8) and (9) are not stationary. Heuristically speaking, condition (13) implies that the non-stationary process $\{\check{\Psi}^*_{n+i}\}$ can be approximated by an stationary $\{\Psi^{**}_i\}$ for large samples. In view of Proposition 3.12 of Straumann and Mikosch (2006), condition (13) would follow if the parametric family $\{g_\phi\}$ satisfies certain random coefficient Lipschitz conditions, uniformly over the compact space Λ.

Because $\{(\Psi_i, \Psi^{**}_i)\}$ is stationary, with $\mathbb{E}^*$ denoting the bootstrap expectation,

$$\begin{aligned}
&\sup_{\phi\in\Lambda} \mathbb{E}\{\mathbb{E}^*[|\check{\Psi}^*_{n+k}(\phi) - \Psi_{n+k}(\phi)|]\} \\
&\quad \le \sup_{\phi\in\Lambda} \mathbb{E}\{\mathbb{E}^*[|\Psi^{**}_{n+k}(\phi) - \Psi_{n+k}(\phi)|]\} + \sup_{\phi\in\Lambda} \mathbb{E}\{\mathbb{E}^*[|\check{\Psi}^*_{n+k}(\phi) - \Psi^{**}_{n+k}(\phi)|]\} \\
&\quad \le \mathbb{E}\{\mathbb{E}^*[\sup_{\phi\in\Lambda} |\Psi^{**}_n(\phi) - \Psi_n(\phi)|]\} + o(1),
\end{aligned} \tag{14}$$

where the last bound follows from (13) and the Jensen inequality.

The validity of the probability forecasts for the parametric approach

To introduce the regularity conditions for the validity of the bootstrap, we need to define the data generating model for a given point $\zeta = (\phi, \theta)^\top \in \Phi \times \Theta$. To this end, let $\mathcal{U} = \{U_i, i \in \mathbb{Z}\}$ be a sequence of iid random variables from the uniform(0,1) distribution. Let $\varepsilon_i(\theta) = F_\theta^{-1}(U_i) := \inf\{x \geq 0 : U_i \leq F_\theta(x)\}$, for $i \in \mathbb{Z}$. The data generating model for a given point $\zeta = (\phi, \theta)^\top \in \Phi \times \Theta$ is defined by the following:

$$\begin{aligned} Z_i^{(\zeta)} &= \Psi_i^{(\zeta)}(\phi)\varepsilon_i(\theta), \\ \Psi_i^{(\zeta)}(\bar{\phi}) &= g_{\bar{\phi}}\{Z_{i-1}^{(\zeta)}, \cdots, Z_{i-p_1}^{(\zeta)}, \Psi_{i-1}^{(\zeta)}(\bar{\phi}), \cdots, \Psi_{i-p_2}^{(\zeta)}(\bar{\phi})\}, \quad \bar{\phi} \in \Phi,\ i \in \mathbb{Z}; \end{aligned} \tag{15}$$

$\{\varepsilon_i(\theta)\}$ are iid with the common cumulative distribution function [cdf] F_θ.

We let functions of $\{Z_i^{(\zeta)}, \Psi_i^{(\zeta)}\}$ in (15) be denoted with a superscript "(ζ)". For example, $\lambda_i^{(\zeta)}(\cdot) := \dot{\Psi}_i^{(\zeta)}(\cdot)/\Psi_i^{(\zeta)}(\cdot)$ is the analogue of $\lambda_i(\cdot) = \dot{\Psi}_i(\cdot)/\Psi_i(\cdot)$ for the observable process $\{Z_i : i \in \mathbb{Z}\}$. If data are generated from (15) for only $i \geq -m$ (conditional on the starting values $\varsigma_0 = (z_0, \cdots, z_{1-p_1}, s_0, \cdots, s_{1-p_2})^\top$), then we use the superscript "(m,ζ)" instead of "(ζ)". For example, $\Psi_i^{(m,\zeta)}(\cdot)$ and $\lambda_i^{(m,\zeta)}(\cdot) := \dot{\Psi}_i^{(m,\zeta)}(\cdot)/\Psi_i^{(m,\zeta)}(\cdot)$ are the analogues of $\Psi_i^{(\zeta)}(\cdot)$ and $\lambda_i^{(\zeta)}(\cdot)$, respectively, when the data generating model obeys (15) for $i \geq -m$ (conditional on the starting values ς_0). Note that, under (1), (2) and (3), the probability laws of $\Psi_i^{(\zeta_0)}(\cdot)$, $\dot{\Psi}_i^{(\zeta_0)}(\cdot)$ and $\lambda_i^{(\zeta_0)}(\cdot)$ are identical to those of $\Psi_i(\cdot)$, $\dot{\Psi}_i(\cdot)$ and $\lambda_i(\cdot)$, respectively.

Let $\zeta_n = (\phi_n^\top, \theta_n^\top)^\top$ denote a generic nonrandom sequence in $\Phi \times \Theta$. Let P_n and $\mathbb{E}_n$ denote the probability and expectation corresponding to $\zeta = \zeta_n$ under model (15). Let

$$\chi_{ni}(\zeta) := \log \Psi_i^{(\zeta_n)}(\phi) - \log\left[f_\theta\{Z_i^{(\zeta_n)}/\Psi_i^{(\zeta_n)}(\phi)\}\right], \quad \varphi_{ni}(\zeta) := -\Sigma_0^{-1}\dot{\chi}_{ni}(\zeta),$$

where Σ_0 is as defined by (A.1) in Appendix A. The norm $\|\cdot\|_\Lambda$ for a continuous $r_2 \times r_3$ matrix-valued function H on a compact set $\Lambda \subset \mathbb{R}^{r_1}$, that is $H \in \mathbb{C}[\Lambda, \mathbb{R}^{r_2 \times r_3}]$, is defined by $\|H\|_\Lambda := \sup_{s \in \Lambda}\|H(s)\|$, where r_1, r_2, r_3 are known positive integers. If H is real valued, then $\|H\|_\Lambda = \sup_{s \in \Lambda}|H(s)|$.

In order to establish the asymptotic validity of the bootstrap probability forecasts based on the AMLE, we need to introduce the following additional assumptions.

Condition C.

C.1. *For every $\zeta = (\phi^\top, \theta^\top)^\top \in \Phi \times \Theta$, the model (15) has a unique stationary ergodic solution $\{Z_i^{(\zeta)} : i \in \mathbb{Z}\}$ with $\mathbb{E}(\{Z_0^{(\zeta)}\}^{2+d})$, $\mathbb{E}[\{\Psi_0^{(\zeta)}(\phi)\}^{2+d}]$ and $\mathbb{E}[\|\lambda_0^{(\zeta)}(\phi)\|^{2+d}]$ being finite for some $d > 0$.*

C.2. *For all nonrandom sequences $\zeta_n = (\phi_n^\top, \theta_n^\top)^\top$ for which $\zeta_n \to \zeta_0$, we have $\|\sqrt{n}(\hat{\zeta}_n - \zeta_n) - [n^{-1/2}\sum_{i=1}^n \varphi_{ni}(\zeta_n)]\| = o_{p_n}(1)$, where $\hat{\zeta}_n$ denotes the analogue of the AMLE (5) when the true parameter is ζ_n. Further, $\mathbb{E}_n[\lambda_{n1}(\phi_n)] \to \mathbb{E}[\lambda_1(\phi_0)]$ and $\mathbb{E}_n[\varphi_{n1}(\zeta_n)\varphi_{n1}(\zeta_n)^\top] \to \mathbb{E}[\varphi_1(\zeta_0)\varphi_1(\zeta_0)^\top]$ as $n \to \infty$, with $\lambda_{ni}(\phi) := \dot{\Psi}_{ni}(\phi)/\Psi_{ni}(\phi)$ where $\Psi_{ni}(\phi) = \Psi_i^{(\zeta_n)}(\phi)$.*

C.3. *There exist compact neighbourhoods K_1 of ϕ_0 and K_2 of θ_0, such that the following hold with $K = K_1 \times K_2$: (a) conditional on $\mathcal{U} = \{U_i, i \in \mathbb{Z}\}$, $\sup_{\zeta \in K} \|\Psi_i^{(m,\zeta)} - \Psi_i^{(\zeta)}\|_{K_1}$, $\sup_{\zeta \in K} \|\dot{\Psi}_i^{(m,\zeta)} - \dot{\Psi}_i^{(\zeta)}\|_{K_1}$, $\sup_{\zeta \in K} \|\ddot{\Psi}_i^{(m,\zeta)} - \ddot{\Psi}_i^{(\zeta)}\|_{K_1}$, $\overset{e.a.s.}{\to} 0$ as $i \to \infty$; (b) $\sup_{\zeta \in K} \mathbb{E}\|\ddot{\Psi}_0^{(\zeta)}\|_{K_1}^{2+d} < \infty$ and $\sup_{\zeta \in K} \mathbb{E}\|\lambda_0^{(\zeta)}\|_{K_1}^{2+d} < \infty$ for some $d > 0$, where $\lambda_i^{(\zeta)} := \dot{\Psi}_i^{(\zeta)}/\Psi_i^{(\zeta)}$; (c) $\sup_{\theta \in K_2} \sup_{x \geq 0} f_\theta(x) < \infty$, $\sup_{\theta \in K_2} \int_{x \geq 0} x f_\theta(x)\, dx < \infty$.*

Condition C.1 is satisfied by most MEMs. For example, for the linear MEM(p_1, p_2) given by (15) with

$$\Psi_i^{(\zeta)}(\phi) = \alpha + \sum_{j=1}^{p_1} \beta_j Z_{i-j}^{(\zeta)} + \sum_{j=1}^{p_2} \gamma_j \Psi_{i-j}^{(\zeta)}(\phi), \quad \phi = (\alpha, \beta_1, \ldots, \beta_{p_1}, \gamma_1, \ldots, \gamma_{p_2})^\top,$$

the relation $\mathbb{E}[\sum_{j=1}^{p_1} \beta_j Z_0^{(\zeta)} + \sum_{j=1}^{p_2} \gamma_j] < 1$ is sufficient for the validity of C.1. Condition C.2 is similar to the Assumption E2 in Andrews (1997). As mentioned in Andrews (1997), the proof of Proposition 2 for the asymptotic normality of the AMLE can be altered to obtain the triangular array linear expansion in C.2. Condition C.3 is also expected to be satisfied by a large class of MEMs. For example, the validity of C.3 for the linear MEM(p_1, p_2) model of Engle and Russell (1998) can be verified as follows:

Verification of Condition C.3 for the linear MEM of Engle and Russell (1998)

Let the parametric linear MEM(p_1, p_2) model, denoted M-$\mathcal{F}$, be defined as follows:

$$M\text{-}\mathcal{F}: \quad \begin{cases} Z_i &= \Psi_i \varepsilon_i, \quad \{\varepsilon_i : i \in \mathbb{Z}\} \text{ are iid}, \quad \varepsilon_i \overset{d}{\sim} F_{\theta_0}, \\ \Psi_i &= \alpha_0 + \sum_{j=1}^{p_1} \beta_{j0} Z_{i-j} + \sum_{j=1}^{p_2} \gamma_{j0} \Psi_{i-j} \end{cases} \tag{16}$$

for some (ϕ_0, θ_0) where $\phi_0^\top = (\alpha_0, \beta_{10}, \ldots, \beta_{p_1 0}, \gamma_{10}, \ldots, \gamma_{p_2 0})$. Let $\{U_i\}_{i \in \mathbb{Z}}$ be iid uniform(0,1) random variables. Without loss of generality, let $\varepsilon_i = F_{\theta_0}^{-1}(U_i)$. A typical assumption made in empirical studies involving MEM(p_1, p_2) is that the following constraints hold at the true parameter (ϕ_0, θ_0) (see Engle 2002):

$$\{1 - \sum_j \beta_{j0} - \sum_j \gamma_{j0}\} > 0 \quad \text{and} \quad \mathbb{E}\{|F_{\theta_0}^{-1}(U_i)|^2\} < \infty. \tag{17}$$

Next, let $V_i = \sup_{\bar{\theta} \in K_\theta} |F_{\bar{\theta}}^{-1}(U_i)|$ where K_θ denotes a closed ball in Θ containing θ_0 as an interior point. If $\{1 - \sum_j \beta_{j0} - \sum_j \gamma_{j0}\} > 0$ and $\mathbb{E}(V_i^{2+\delta}) < \infty$ for some $\delta > 0$ and K_θ, then the conditions in C.3 are also satisfied; these verifications require results on *e.a.s.* convergence involving *Stochastic Recurrence Equations* and can be obtained from the authors upon request. The additional requirement $\mathbb{E}(V_i^{2+\delta}) < \infty$ is only slightly stronger than $\mathbb{E}|F_{\theta_0}^{-1}(U_i)|^2 < \infty$ in (17). These conditions can similarly be verified for many other MEMs.

The next theorem shows that the probability forecasts proposed in Subsection 2.1 are asymptotically valid. Here and in the sequel $G_{\check{Z}^*,k}(x)$ and $G_{\check{\Psi}^*,k}(x)$ denote the distribution functions of $\check{Z}^*_{n+k}$ and $\check{\Psi}^*_{n+k}$, respectively, conditional on $\{Z_1,\ldots,Z_n\}$.

Theorem 1. *Let $\{Z_i; i\in\mathbb{Z}\}$ be a strictly stationary and ergodic process obeying the model specified by (1), (2) and (3). Suppose that the assumptions of Proposition 2 in Appendix A.1 and (13) are satisfied. Additionally, assume that conditions C.1, C.2, and C.3 hold, and that there exists a compact neighbourhood $B\,(\subset\Phi)$ of ϕ_0 with $\sup_{\phi\in B}|\widetilde{\Psi}_i(\phi)-\Psi_i(\phi)|$, $\sup_{\phi\in B}\|\dot{\widetilde{\Psi}}_i(\phi)-\dot{\Psi}_i(\phi)\| \overset{e.a.s.}{\to} 0$ as $i\to\infty$. Then, conditional on $\{Z_1,\ldots,Z_n\}$, for every $0<\omega<1$, the following hold:* (a) $P\{Z_{n+k}\le G^{-1}_{\check{Z}^*,k}(\omega)\}\to_p\omega$ *as $n\to\infty$,* (b) $P\{\Psi_{n+k}\le G^{-1}_{\check{\Psi}^*,k}(\omega)\}\to_p\omega$ *as $n\to\infty$.*

The proof of Theorem 1 makes use of (13) and (14) and is given in Appendix A. By the Glivenko-Cantelli theorem we have that, $\sup_x|G_{\check{Z}^*,k}(x)-G_{B,\check{Z}^*,k}(x)| \overset{P_n^*-a.s.}{\to} 0$ and $\sup_x|G_{\check{\Psi}^*,k}(x)-G_{B,\check{\Psi}^*,k}(x)| \overset{P_n^*-a.s.}{\to} 0$, as $B\to\infty$. Hence, one may make $G_{B,\check{Z}^*,k}(x)$ and $G_{B,\check{\Psi}^*,k}(x)$ arbitrarily close to $G_{\check{Z}^*,k}(x)$ and $G_{\check{\Psi}^*,k}(x)$, respectively, by selecting B large enough. Thus, it follows from Theorem 1 that the probability forecasts in (8) and (9) are asymptotically valid.

The validity of the probability forecasts for the semiparametric approach

The next theorem shows that the probability forecasts proposed in Section 2.2 based on bootstrap replicates obtained from the empirical distribution of the standardized residuals, are also asymptotically valid. First, we introduce several regularity conditions on the error distribution F_0 and the functional form $\Psi_i(\phi)$.

L.1. *The support of the error distribution F_0 is $[0,\infty)$, ε_i^2 has a non-degenerate distribution, and $\mathbb{E}[\varepsilon_0^{(2+d)}]<\infty$ for some $d>0$.*

L.2. *The function $(\phi,\boldsymbol{s})\mapsto g_\phi(\boldsymbol{z},\boldsymbol{s})$ is twice continuously differentiable in its domain, $\mathbb{E}\{|\log\Psi_i(\phi_0)|\}<\infty$, and $\mathbb{E}[\|\lambda_i(\phi_0)\|^2]<\infty$.*

L.3. *The parameter ϕ_0 is an interior point in Φ, and there exists a compact neighbourhood $B\,(\subset\Phi)$ of ϕ_0 such that $\sup_{\phi\in B}|\widetilde{\Psi}_i(\phi)-\Psi_i(\phi)|$, $\sup_{\phi\in B}\|\dot{\widetilde{\Psi}}_i(\phi)-\dot{\Psi}_i(\phi)\| \overset{e.a.s.}{\to} 0$ as $i\to\infty$. Furthermore, $\hat{\phi}_{qml}\to_p\phi_0$, $n^{1/2}(\hat{\phi}_{qml}-\phi_0)=O_p(1)$, and $\hat{\phi}_{qml}$ satisfies (11).*

Condition L.2 follows from the regularity assumptions B.5, D.2, and D.4 stated in Appendix A.1. The convergence properties and the root-n consistency of the QMLE in L.3 are typically valid for many parametric models of the form (2); see Bauwens and Giot, 2001; Francq and Zakoïan, 2010; Hautsch, 2011.

Theorem 2. *Let $\{Z_i; i \in \mathbb{Z}\}$ be a strictly stationary and ergodic process obeying (1) and (2) with $\Psi_i = \Psi_i(\phi_0)$. Let F_n denote a sequence of non-random cdf's. Suppose that L.1, L.2, L.3, B.6, and B.7 are satisfied, and that the analogue of (13) holds for the bootstrap process in Section 2.2. Additionally, assume that the analogues of C.1, C.2, and C.3 hold if F_θ is replaced by a non-parametric cdf F, F_{θ_n} is replaced by F_n with $(\phi_n, F_n) \to (\phi_0, F_0)$, and AMLE is replaced by QMLE. Let $G_{\check{Z}^{*qml},k}(x)$ and $G_{\check{\Psi}^{*qml},k}(x)$ be the analogues of $G_{\check{Z}^*,k}(x)$ and $G_{\check{\Psi}^*,k}(x)$ for the bootstrap algorithm in Section 2.2. Then, conditional on $\{Z_1, \ldots, Z_n\}$, for every $0 < \omega < 1$ we have the following:* (a) $P\{Z_{n+k} \le G^{-1}_{\check{Z}^{*qml},k}(\omega)\} \to_p \omega$, (b) $P\{\Psi_{n+k} \le G^{-1}_{\check{\Psi}^{*qml},k}(\omega)\} \to_p \omega$, *as $n \to \infty$.*

The proof of Theorem 2 is given in Appendix A. Note that, the analogue of C.2 for the setting in Theorem 2 is that, for all non-random sequences $(\phi_n, F_n) \to (\phi_0, F_0)$, as $n \to \infty$, we have $\|\sqrt{n}(\hat{\phi}_{n,qml} - \phi_n) - [c_n \Sigma_n^{-1} n^{-1/2} \sum_{i=1}^n \lambda^\dagger_{ni}(\phi_n) \varrho(\varepsilon^\dagger_{ni})]\| = o_{p_n}(1)$, with $c_n \Sigma_n^{-1} \to c\Sigma^{-1}$, $\varepsilon^\dagger_{ni} \sim F_n$, and $\lambda^\dagger_{ni}(\phi) := \dot{\Psi}^\dagger_{ni}(\phi)/\Psi^\dagger_{ni}(\phi)$, where $\hat{\phi}_{n,qml}$ is the analogue of the QMLE (11) when the true parameter is ϕ_n, and $\Psi^\dagger_{ni}$ denotes the conditional mean process for the model with the true parameter ϕ_n and the error distribution F_n as defined by (A.6) in Appendix A.2.

In the next subsection we propose a class of specification tests to test the adequacy of the parametric specifications of the conditional mean and the error distribution.

2.4 Specification tests for the conditional distribution

In the bootstrap method proposed in Section 2.1, we assume that $\{Z_i : i \in \mathbb{Z}\}$ obeys a parametric MEM of the form (1), (2) and (3). Because such a parametric model specifies the entire conditional distribution of Z_i, it is prudent to first test for the adequacy of this parametric model. To this end, we need to test the null hypothesis

$$H_0 : Pr(Z_i \le z | \mathcal{H}_{i-1}) = F_{\theta_0}(z/\Psi_i(\phi_0)),\ z \ge 0,\ \text{for some } (\phi_0, \theta_0) \in \Phi \times \Theta, \qquad (18)$$

against the alternative '$H_1 :$ Not H_0'.

There are several tests in the literature that we can use for this purpose; see, for example Fernandes and Grammig (2005), Gao *et al.* (2015), and Perera and Silvapulle (2017). However, these tests rely on QMLE for estimating the null model and hence sacrifice efficiency in favour of achieving robustness against distributional misspecifications. Since AMLE is more efficient than QMLE under H_0, in this paper, we develop a testing procedure based on AMLE for testing H_0. The bootstrap tests of Perera *et al.* (2016) are based on QMLE and applicable only for testing the conditional mean function but not (18). The tests proposed in Hidalgo and Zaffaroni (2007) are based on a Gaussian QMLE and can be applied for fitting an ARCH(∞)

model which includes several GARCH-type models; these tests cannot be used for testing the adequacy of a distributional model such as (18).

To describe the testing procedure that we develop, let $\hat{\zeta} = (\hat{\phi}^\top, \hat{\theta}^\top)^\top$ and $\widetilde{\Psi}_i(\cdot)$ be as in the previous section. Let 'plim' denote the *probability limit operator*. Define $\zeta_0 = (\phi_0, \theta_0)^\top := \operatorname{plim} \hat{\zeta}$, so that ζ_0 is the true value under H_0, and it denotes a pseudo true value under H_1. Let $\widetilde{F}_n(x) = n^{-1} \sum_{i=1}^{n} I(\widetilde{\varepsilon}_i \leq x)$ be the empirical distribution function of the estimated residuals $\{\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_n\}$, where $\widetilde{\varepsilon}_i := Z_i / \widetilde{\Psi}_i(\hat{\phi})$, $i = 1, \cdots, n$. Let the residual empirical process $\widetilde{W}_n$ estimated under the null hypothesis be

$$\widetilde{W}_n(x) := \sqrt{n} \left\{ \widetilde{F}_n(x) - F_{\hat{\theta}}(x) \right\}, \quad x \geq 0. \tag{19}$$

Let $D[0, \infty)$ and $D[0, 1]$ be the spaces of *càdlàg* functions on $[0, \infty)$ and $[0, 1]$, respectively, equipped with the uniform metric. Certain functionals of $\widetilde{W}_n$ defined on $D[0, \infty)$ may be used as possible test statistics for testing H_0 against H_1. We consider a test statistic $T_{\mathfrak{h}}$, of the following general form, that satisfies

$$T_{\mathfrak{h}} := \mathfrak{h}(\widetilde{W}_n \circ F_{\hat{\theta}}^{-1}) = \mathfrak{h}(\widetilde{W}_n \circ F_{\theta_0}^{-1}) + o_p(1), \tag{20}$$

under H_0, where $\mathfrak{h}$ is a known continuous functional on $D[0, 1]$.

Possible candidates for $\mathfrak{h}$ include: (a) the *Kolmogorov-Smirnov* functional $\mathfrak{h}(\widetilde{W}_n \circ F_{\hat{\theta}}^{-1}) := \sup_{t \in [0,1]} |\widetilde{W}_n \circ F_{\hat{\theta}}^{-1}(t)|$, and (b) the *Cramér-von Mises* functional $\mathfrak{h}(\widetilde{W}_n \circ F_{\hat{\theta}}^{-1}) := \int_0^1 \{\widetilde{W}_n \circ F_{\hat{\theta}}^{-1}(t)\}^2 dt$, among others (D'Agostino and Stephens, 1986). Similar functionals have also been considered in Meintanis *et al.* (2017) in a simulation study for testing the goodness-of-fit of the error distribution in ACD models. Guo and Li (2018) propose specification tests for MEMs when the conditional mean admits a Markov structure, such that $\Psi_i = \mathbb{E}[Z_i \mid Z_{i-1}]$ $(i \in \mathbb{Z})$; in this paper, we do not make this assumption.

Next, we introduce some notation. The composition of two functions f_1 and f_2 is denoted by $f_1 \circ f_2$. The empirical distribution function $F_n(x)$ and the corresponding empirical process $W_n(x)$ of the unobserved errors $\{\varepsilon_1, \ldots, \varepsilon_n\}$ are defined by

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I(\varepsilon_i \leq x) \quad \text{and} \quad W_n(x) = \sqrt{n}\{F_n(x) - F_{\theta_0}(x)\}, \quad x \geq 0.$$

Under Conditions D.1–D.4, there exist martingale difference sequences $\{\varphi_i^{(1)}(\zeta)\}$ and $\{\varphi_i^{(2)}(\zeta)\}$, with $\|n^{1/2}(\hat{\phi} - \phi_0) - [n^{-1/2} \sum_{i=1}^{n} \varphi_i^{(1)}(\zeta_0)]\| = o_p(1)$ and $\|n^{1/2}(\hat{\theta} - \theta_0) - [n^{-1/2} \sum_{i=1}^{n} \varphi_i^{(2)}(\zeta_0)]\| = o_p(1)$, where $[(\varphi_i^{(1)}(\zeta))^\top (\varphi_i^{(2)}(\zeta))^\top]^\top = \varphi_i(\zeta) := \Sigma_0^{-1} \dot{\chi}_i(\zeta)$, $\|\mathbb{E}\varphi_i^{(1)}(\zeta_0)\varphi_i^{(1)}(\zeta_0)^\top\| < \infty$ and $\|\mathbb{E}\varphi_i^{(2)}(\zeta_0)\varphi_i^{(2)}(\zeta_0)^\top\| < \infty$ [see (A.2) in the proof of Proposition 2 in Appendix A]. By using these asymptotic representations, and Conditions D.1–D.4, the next theorem establishes the weak convergence of $\widetilde{W}_n \circ F_{\theta_0}^{-1}$.

Theorem 3. *Let $\{Z_i; i \in \mathbb{Z}\}$ be a stationary and ergodic process that obeys the model described by (1), (2) and (3) under H_0. Suppose that the assumptions of Theorem 1 are satisfied. Additionally, assume that $\dot{F}_{\theta_0}(y)$ and $yf_{\theta_0}(y)$ are uniformly continuous on $\mathbb{R}^+$. Then, $\widetilde{W}_n \circ F_{\theta_0}^{-1}$ converges weakly in $D[0,1]$ to a centred Gaussian process R, defined by the covariance kernel $Cov\{R(s), R(t)\} := \min\{s,t\} - st + \mathcal{R}(s,t,\zeta_0)$, where*

$$
\begin{aligned}
\mathcal{R}(s,t,\zeta) &= F_\theta^{-1}(s) f_\theta(F_\theta^{-1}(s))\mathbb{E}[\lambda_1(\phi)]^\top \mathbb{E}[\varphi_1^{(1)}(\zeta) I(\varepsilon_i \leq F_\theta^{-1}(t))] \\
&\quad - [\dot{F}_\theta(F_\theta^{-1}(s))]^\top \mathbb{E}[\varphi_1^{(2)}(\zeta) I(\varepsilon_i \leq F_\theta^{-1}(t))] - [\dot{F}_\theta(F_\theta^{-1}(t))]^\top \mathbb{E}[\varphi_1^{(2)}(\zeta) I(\varepsilon_i \leq F_\theta^{-1}(s))] \\
&\quad - \mathbb{E}\left\{ \left[\dot{F}_\theta(F_\theta^{-1}(s))\right]^\top \varphi_1^{(2)}(\zeta)\mathbb{E}[\lambda_1(\phi)]^\top \varphi_1^{(1)}(\zeta) \right\} F_\theta^{-1}(s) f_\theta(F_\theta^{-1}(s)) \\
&\quad + F_\theta^{-1}(t) f_\theta(F_\theta^{-1}(t))\mathbb{E}[\lambda_1(\phi)]^\top \mathbb{E}[\varphi_1^{(1)}(\zeta) I(\varepsilon_i \leq F_\theta^{-1}(s))] \\
&\quad - F_\theta^{-1}(t) f_\theta(F_\theta^{-1}(t))\mathbb{E}\left\{ \mathbb{E}[\lambda_1(\phi)]^\top \varphi_1^{(1)}(\zeta)\varphi_1^{(2)}(\zeta)^\top \dot{F}_\theta(F_\theta^{-1}(s)) \right\} \\
&\quad + F_\theta^{-1}(t) f_\theta(F_\theta^{-1}(t))\mathbb{E}\left\{ \mathbb{E}[\lambda_1(\phi)]^\top \varphi_1^{(1)}(\zeta)\varphi_1^{(1)}(\zeta)^\top \mathbb{E}[\lambda_1(\phi)] \right\} F_\theta^{-1}(s) f_\theta(F_\theta^{-1}(s)),
\end{aligned}
$$

and $\lambda_i(\phi) := \dot{\Psi}_i(\phi)/\Psi_i(\phi)$ for $\phi \in \Phi$.

The next corollary yields the limiting distributions of the test statistic $T_{\mathfrak{h}}$ under the null and under a fixed alternative. Here, and in the sequel, '$\xrightarrow{d}$' and '$\xrightarrow{p}$' denote convergence in distribution and convergence in probability, respectively.

Corollary 1. *Suppose that the assumptions of Theorem 3 are satisfied. Let $\mathfrak{h}$ be a continuous functional on $D[0,1]$. Then, under H_0, $\mathfrak{h}(\widetilde{W}_n \circ F_{\theta_0}^{-1}) \xrightarrow{d} \mathfrak{h}(R)$ as $n \to \infty$, and under any fixed alternative, with probability (w.p.) 1, $\mathfrak{h}(\widetilde{W}_n \circ F_{\theta_0}^{-1}) \longrightarrow \infty$ as $n \to \infty$.*

Since $\mathfrak{h}$ is a continuous functional on $D[0,1]$, and $T_{\mathfrak{h}} = \mathfrak{h}(\widetilde{W}_n \circ F_{\theta_0}^{-1}) + o_p(1)$ under H_0, it follows from Corollary 1 that the limiting null distribution of $T_{\mathfrak{h}}$ is $\mathfrak{h}(R)$. Therefore, an asymptotic test based on $T_{\mathfrak{h}}$ would reject H_0 if $T_{\mathfrak{h}} > c_{\mathfrak{h}}$, where $c_{\mathfrak{h}}$ is the $(1-\alpha)$th quantile of $\mathfrak{h}(R)$. The distribution of $\mathfrak{h}(R)$ is model-dependent and is not free from the nuisance parameters ζ and F_θ. Hence, asymptotic critical values cannot be computed for general use. Therefore, to implement the tests, in this paper, we adopt a parametric bootstrap approach that is commonly used in the literature, for example, as considered in Horváth *et al.* (2004), Fernandes and Grammig (2005), and Meintanis *et al.* (2017), amongst others. In implementing this approach, in the bootstrap data generation, we use the AMLE $\hat{\zeta}$ in (5) instead of using the QMLE; this approach has its roots in the parametric bootstrap approach originally proposed by Andrews (1997). We expect that this method would complement the existing methods for specification testing in MEMs, including Hidalgo and Zaffaroni (2007); Koul *et al.* (2012); Perera and Koul (2017); Meintanis *et al.* (2017); Guo and Li (2018), amongst others. The bootstrap testing procedure is described by the following steps:

Bootstrap procedure

Step 1: Compute the estimate $\hat{\zeta} = (\hat{\phi}^\top, \hat{\theta}^\top)^\top$ and the test statistic $T_{\mathfrak{h}}$ based on the observed sample $\{Z_1, \ldots, Z_n\}$.

Step 2: Generate a bootstrap sample $\{Z_1^*, \ldots, Z_n^*\}$ by repeating Steps 2–4 of the bootstrap procedure in Section 2.1.

Step 3: Based on the bootstrap sample $\{Z_1^*, \ldots, Z_n^*\}$, compute the bootstrap counterparts of $\hat{\zeta}$, $\widetilde{\Psi}_i$, $\{\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_n\}$, $\widetilde{F}_n$ and $\widetilde{W}_n$, which we denote by $\hat{\zeta}^* = (\{\hat{\phi}^*\}^\top, \{\hat{\theta}^*\}^\top)^\top$, $\widetilde{\Psi}_i^*$, $\{\widetilde{\varepsilon}_1^*, \ldots, \widetilde{\varepsilon}_n^*\}$, $\widetilde{F}_n^*$ and $\widetilde{W}_n^*$, respectively. Then, compute $T_{\mathfrak{h}}^* = \mathfrak{h}(\widetilde{W}_n^* \circ F_{\hat{\theta}^*}^{-1})$, the bootstrap analogue of $T_{\mathfrak{h}} = \mathfrak{h}(\widetilde{W}_n \circ F_{\hat{\theta}}^{-1})$.

Step 4: Estimate the distribution of $T_{\mathfrak{h}}^*$ by repeating Steps 2 and 3 many times and compute $c_{\mathfrak{h}}^*$, the $(1-\alpha)$th quantile of the sampled values of $T_{\mathfrak{h}}^*$. Now, reject H_0 at level α if $T_{\mathfrak{h}} > c_{\mathfrak{h}}^*$.

Theorem 4 and Corollary 2 below show that the foregoing bootstrap test is asymptotically valid. Theorem 4 establishes the weak convergence of the estimated bootstrap empirical process $\widetilde{W}_n^* \circ F_{\hat{\theta}}^{-1}$. Corollary 2 shows that, conditional on $\{Z_1, \ldots, Z_n\}$, $T_{\mathfrak{h}}^*$ converges in distribution (in the bootstrap sense). All the convergence results relating to bootstrapped processes such as $\widetilde{W}_n^* \circ F_{\hat{\theta}}^{-1}$ hold almost surely.

Theorem 4. *Suppose that the assumptions of Theorem 3, except the null hypothesis H_0, are satisfied. Additionally, assume that the Conditions C.1, C.2 and C.3 are also satisfied. Then, conditional on $\{Z_1, \ldots, Z_n\}$, w.p. 1, the process $\widetilde{W}_n^* \circ F_{\hat{\theta}}^{-1}$ converges weakly in $D[0,1]$ to a centred Gaussian process $\mathcal{S}$ with covariance kernel Cov$\{\mathcal{S}(s), \mathcal{S}(t)\} := \min\{s,t\} - st + \mathcal{R}(s,t,\zeta_0)$, where $\mathcal{R}(s,t,\zeta)$ is as in Theorem 3.*

Let $O_{p_n^*}$, $o_{p_n^*}$, and E^* denote the usual stochastic orders of magnitude and expectation, respectively, with respect to the bootstrap law, P_n^*, conditional on $\{Z_1, \ldots, Z_n\}$. The convergence in distribution of bootstrap statistics is denoted by '$\xrightarrow{d^*}$'. The following corollary establishes the asymptotic validity of the bootstrap tests.

Corollary 2. *Suppose that the assumptions of Theorem 4 are satisfied. Let $\mathcal{S}$ be as in Theorem 4. Then, conditional on $\{Z_1, \ldots, Z_n\}$, w.p. 1, $T_{\mathfrak{h}}^* \xrightarrow{d^*} \mathfrak{h}\{\mathcal{S}\}$. Further, conditional on $\{Z_1, \ldots, Z_n\}$, w.p. 1, the bootstrap implementation of $T_{\mathfrak{h}}$ has asymptotic power one against any fixed alternative.*

Suppose that H_0 is true. Then $\zeta_0 = (\phi_0^\top, \theta_0^\top)^\top$ is the true value satisfying $[\Psi_i, F_0] = [\Psi_i(\phi_0), F_{\theta_0}]$, and hence, the process $\mathcal{S}$ in Theorem 4 is the same as the centred Gaussian process R in Theorem 3. Thus, under H_0, the distribution of $\mathfrak{h}\{\mathcal{S}\}$ is the same as that of $\mathfrak{h}\{R\}$, the asymptotic null distribution of $T_{\mathfrak{h}}$. Therefore, it follows from Corollary 2 that the bootstrap test based on $T_{\mathfrak{h}}$ is a valid level α asymptotic test.

3 SIMULATION STUDIES

We conducted two separate simulation studies. In the first simulation study, we evaluated the finite sample performance of the prediction intervals and upper prediction bounds proposed in Section 2.1. In the second simulation study we evaluated the specification tests proposed in Section 2.4 in terms of size and power. These two simulation studies involve several models and evaluations with respect to different criteria. Therefore, in this section, we provide a summary of the main observations, and relegate the detailed tables to Appendix S.1 in the supplementary material.

Simulation Study 1: Evaluating predictions for Z_{n+k} and Ψ_{n+k}

Let us first describe the design of the study. For the error distribution, we considered the following families of distribution functions on $\mathbb{R}^+$ with mean 1 and pdf f:
(i) *Exponential,* denoted E: $f(x) = \exp(-x)$.
(ii) *Weibull* $[W(\kappa)]$: $f(x) = (\kappa/c)(x/c)^{\kappa-1}\exp\{-(x/c)^\kappa\}$, $\kappa > 0$, $c = [\Gamma(1+\kappa^{-1})]^{-1}$.
(iii) *Burr* $[B(a,b)]$: $f(x) = (a/\sigma)(x/\sigma)^{a-1}[1 + b(x/\sigma)^a]^{-(1+b^{-1})}$, $a > b > 0$, and $\sigma = \{\Gamma(1+a^{-1})\Gamma(b^{-1}-a^{-1})\}^{-1} b^{(1+a^{-1})}\Gamma(1+b^{-1})$.
(iv) *Generalized gamma* $[GG(a,c)]$:
$f(x) = c\{\sigma\Gamma(a)\}^{-1}(x/\sigma)^{ac-1}\exp\{-(x/\sigma)^c\}$, $a, c > 0$, and $\sigma = \{\Gamma(a + c^{-1})\}^{-1}\Gamma(a)$.

The first two distributions have been identified as having important roles in multiplicative error models (see Engle and Russell, 1998; Drost and Werker, 2004; Engle and Gallo, 2006). The next two have been suggested in various empirical studies (see Lunde, 1999; Grammig and Maurer, 2000; Grammig and Wellner, 2002).

Let AM and M denote the following two models, with unknown parameters:

AM [Asymmetric MEM(1,1); Fernandes and Grammig (2006)]:

$$\begin{aligned} \Psi_i(\phi) &= \phi_1 + \phi_2\Psi_{i-1}(\phi)\{|\varepsilon_{i-1} - \phi_4| + \phi_5(\varepsilon_{i-1} - \phi_4)\} + \phi_3\Psi_{i-1}(\phi), \qquad (21)\\ \varepsilon_{i-1} &= \varepsilon_{i-1}(\phi) = Z_{i-1}/\Psi_{i-1}(\phi), \quad \phi = (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)^\top, \end{aligned}$$

M [Linear MEM(1,1); Engle and Russell (1998)]:

$$\Psi_i(\phi) = \phi_1 + \phi_2 Z_{i-1} + \phi_3\Psi_{i-1}(\phi), \quad \phi_1 > 0, \phi_2 \geq 0, \phi_3 \geq 0, \quad \phi_2 + \phi_3 < 1. \qquad (22)$$

The term $\{|\varepsilon_{i-1} - \phi_4| + \phi_5(\varepsilon_{i-1} - \phi_4)\}$ in the asymmetric model (21) allows the conditional mean $\{\Psi_i\}$ to respond in distinct manners to small and large shocks, through the additional shift and rotation parameters ϕ_4 and ϕ_5 (see Fernandes and Grammig, 2006). The linear MEM(1,1) in (22) may be recovered from this general model by setting $\phi_4 = \phi_5 = 0$ in (21).

The data generating processes were based on the following two multiplicative error models for the conditional mean $\mathit{\Psi}_i$, with $\varepsilon_i = Z_i/\mathit{\Psi}_i$.

$$\begin{aligned} M_1: \quad & \mathit{\Psi}_i = 0.20 + 0.10 Z_{i-1} + 0.70\mathit{\Psi}_{i-1}, \\ AM_1: \quad & \mathit{\Psi}_i = 0.1 + 0.2\mathit{\Psi}_{i-1}\{|\varepsilon_{i-1} - 0.1| + 0.5(\varepsilon_{i-1} - 0.1)\} + 0.6\mathit{\Psi}_{i-1}. \end{aligned}$$

Clearly, M_1 belongs to M and AM_1 belongs to AM. We use the notation M-$\mathcal{F}$ to denote the class of MEMs defined by $Z_i = \mathit{\Psi}_i\varepsilon_i$, where $\mathit{\Psi}_i$ is of the form M and $\{\varepsilon_i; i \in \mathbb{Z}\}$ are iid with common distribution $F \in \mathcal{F}$ as in (16). Similarly, AM-$\mathcal{F}$ denotes the class of MEMs, where $\mathit{\Psi}_i$ follows the model AM and $\{\varepsilon_i; i \in \mathbb{Z}\}$ are iid with common distribution $F \in \mathcal{F}$. For example, AM-GG represents the class of MEMs defined by the asymmetric conditional mean function of AM in (21) and a generalized gamma error distribution, denoted GG.

We carried out several sets of simulations to evaluate the accuracy of the conditional probability forecasts for Z_{n+k} and $\mathit{\Psi}_{n+k}$ given by the bootstrap method proposed in Section 2.1. For each generated sample, the prediction intervals and bounds given by (8) and (9) were constructed under several parametric MEMs that nest the DGP. The simulations are based on 1000 Monte Carlo samples. For each Monte Carlo sample, we used 1000 bootstrap replicates to produce the probability forecasts.

The accuracy of the left and right limits of the prediction intervals are also of interest to know about whether the prediction intervals are properly centred and accurate. Therefore, we also considered the following procedure in our simulations.

The accuracy of the left and right limits of the prediction intervals:

Let $G_{z,k}(\cdot)$ denote the cdf of Z_{n+k} conditional on $\{Z_1, \ldots, Z_n\}$. Let q_L and q_U denote the lower and upper quantiles of the distribution $G_{z,k}$ at the nominal level α. Let q_L^* and q_U^* denote the lower and upper limits of the bootstrap prediction intervals, conditional on $\{Z_1, \ldots, Z_n\}$, at level α. In order to measure the accuracy of q_L^* and q_U^* we adopt the following procedure.

Step 1: For the ℓth Monte Carlo sample, say $\{Z_1^\ell, \ldots, Z_n^\ell\}$, generate $M = 100{,}000$ independent observations of Z_{n+k}^ℓ, and obtain approximate conditional quantiles, say q_L^ℓ and q_U^ℓ, by using the generated M values, $\ell = 1, \ldots, N$, $N = 1000$; note that, q_L^ℓ and q_U^ℓ can be made arbitrarily close to their true values by selecting M large enough.
Step 2: Compute the lower and upper limits of the bootstrap prediction intervals, say $q_L^{*\ell}$ and $q_U^{*\ell}$, conditional on $\{Z_1^\ell, \ldots, Z_n^\ell\}$, $\ell = 1, \ldots, N$.
Step 3: Compute the standard deviations of $\{(q_L^{*1} - q_L^1), \ldots, (q_L^{*N} - q_L^N)\}$ and $\{(q_U^{*1} - q_U^1), \ldots, (q_U^{*N} - q_U^N)\}$, denoted sd_L and sd_U, respectively, and also compute

$$R_L = N^{-1}\sum_{\ell=1}^{N}[q_L^{*\ell} - q_L^\ell], \quad R_U = N^{-1}\sum_{\ell=1}^{N}[q_U^{*\ell} - q_U^\ell]. \tag{23}$$

If q_L^* and q_U^* are centred around their corresponding population values q_L and q_U, then we expect R_L and R_U to be close to zero, and sd_L and sd_U to be not (relatively) too large. One may also obtain similar measures for the accuracy of the prediction intervals of Ψ_{n+k} and prediction bounds of Z_{n+k} and Ψ_{n+k}.

A summary of the results for prediction intervals of Z_{n+k}, for $k = 2$ and 4, are given in Tables S.1–S.3 in Appendix S.1 in the supplementary material. Results for upper prediction bounds of Z_{n+k}, for $k = 2$ and 4, and those of Ψ_{n+k}, for $k = 3$ and 5, are summarized in Tables S.4–S.6 and Tables S.7–S.9, respectively; see Appendix S.1 in the supplementary material. As illustrated by these results, the coverage percentages of the interval forecasts were all close to the nominal levels, while the widths of the intervals exhibited some variation among the different parametric models. Since the estimated coverages of the prediction intervals and bounds were all close to the nominal levels our simulation results also indicate good coverage properties on the left and right of the prediction intervals. Thus, the results suggest that our probability forecasts perform well in finite samples. The results for the measures R_L, R_U, sd_L, and sd_U, indicate that even if the variation in the parameter estimation is small, if a prediction is to be made for few steps ahead, then the resulting variations in the probability forecasts can still be significant (see Tables S.1–S.9 in Appendix S.1).

Simulation Study 2: Comparison of tests in terms of size and power

To evaluate the finite sample performance of the specification tests, in addition to *Exponential* $[E]$, *Weibull* $[W(\kappa)]$, *Burr* $[B(a, b)]$, and *Generalized gamma* $[GG(a, c)]$, we also considered the following families of distribution functions on $\mathbb{R}^+$ with mean 1:

(v) *Gamma* $[G(\alpha)]$: the pdf is $f(x) = \alpha^\alpha \Gamma(\alpha)^{-1} x^{\alpha-1} \exp(-\alpha x)$ and $\alpha > 0$.

(vi) *G-mixture* $[GmGG(\delta)]$: $\delta G(2) + (1 - \delta)GG(3, 0.5)$, a mixture of Gamma and Generalized gamma distributions with mixing proportion δ.

(vii) *W-mixture* $[WmGG(\delta)]$: $\delta W(0.6) + (1 - \delta)GG(3, 0.5)$, a mixture of Weibull and Generalized gamma distributions with mixing proportion δ.

(viii) *Generalized Extreme Value Distribution* $[GEV(k, \sigma, \mu)]$: the pdf is

$$f(x) = \sigma^{-1} \exp(-(1 + k(x - \mu)/\sigma)^{-1/k})(1 + k(x - \mu)/\sigma)^{-1-1/k}, \quad k \neq 0,$$

$1 + k(x - \mu)/\sigma > 0$, $\mu = 1 - \sigma\{\Gamma(1 - k) - 1\}/k$.

The mixture families $GmGG(\delta)$ and $WmGG(\delta)$ were used for evaluating the power of the tests near gamma and Weibull distributions. In view of the regularity conditions, the first order validity of our bootstrap tests is established under the assumption $\mathbb{E}[\varepsilon_0^{2+d}] < \infty$ for some $d > 0$. Hence, it is of interest to investigate how tests perform when this assumption breaks down. For this purpose we use the last family of distributions, $GEV(k, \sigma, \mu)$. This family has infinite second moment when the shape parameter $k \geq 1/2$; the condition $\mu = 1 - \sigma\{\Gamma(1-k) - 1\}/k$ ensures that mean is 1.

For the data generating processes, in addition to M_1 and AM_1, we also considered the following models:

$M_1^\dagger$ [Non-stationary MEM(1,1)]: $\Psi_i = 0.20 + 0.20 Z_{i-1} + 0.80 \Psi_{i-1}$,

M_2 [MEM(2,1)]: $\Psi_i = 0.10 + 0.20 Z_{i-1} + 0.10 Z_{i-2} + 0.60 \Psi_{i-1}$,

TM_1 [3-regime Threshold MEM(1,1), see Zhang *et al.* (2001)]:

$$\Psi_i = \begin{cases} 1.05 + 0.09 Z_{i-1} + 0.90 \Psi_{i-1} & \text{for } \ 0 < Z_{i-1} < 0.25, \\ 0.50 + 0.55 Z_{i-1} + 0.10 \Psi_{i-1} & \text{for } \ 0.25 \leq Z_{i-1} < 1.5, \\ 0.05 + 0.05 Z_{i-1} + 0.60 \Psi_{i-1} & \text{for } \ 1.5 \leq Z_{i-1} < \infty, \end{cases}$$

Recall that M_1 belongs to M and AM_1 belongs to AM, where AM and M denote the asymmetric MEM in (21) and linear MEM(1,1) in (22), respectively, with unknown parameters. The data generating processes M_2 and TM_1 do not belong to M or AM. Hence, we use M_2 and TM_1 to evaluate the power of the tests. The DGP $M_1^\dagger$ is also of the form M, but it does not satisfy the condition $\phi_2 + \phi_3 < 1$ in (22), and hence is not in the parameter space where the stationarity holds. Thus, we use $M_1^\dagger$ to investigate the behaviour of the bootstrap tests when the stationarity does not hold.

We considered the following five test statistics:

$KS \ = \sup_{x \geq 0} |\widetilde{W}_n(x)|$ [Kolmogorov-Smirnov].

$Ku \ = \sup_{x \geq 0} \widetilde{W}_n(x) - \inf_{x \geq 0} \widetilde{W}_n(x)$ [Kuiper].

$CvM = \int \widetilde{W}_n^2(x) dF_{\hat\theta}(x)$ [Cramér-von Mises].

$A^2 \ = \int \widetilde{W}_n^2(x)[F_{\hat\theta}(x)\{1 - F_{\hat\theta}(x)\}]^{-1} dF_{\hat\theta}(x)$ [Anderson-Darling].

$U^2 \ = \int \left\{\widetilde{W}_n(x) - \int [\widetilde{W}_n(x)] dF_{\hat\theta}(x)\right\}^2 dF_{\hat\theta}(x)$ [Watson].

Each of the above test statistics is of the form $T_{\mathfrak{h}}$ in (20). These functionals have previously been employed in the literature for goodness-of-fit testing in other settings (see D'Agostino and Stephens, 1986).

For comparison, the FG (Fernandes and Grammig, 2005) and JG (Janssen *et al.*, 2005) tests were also considered. The results are based on 1000 Monte Carlo repetitions. In order to reduce the computational burden, we adopted the "Warp-Speed" Monte Carlo method of Giacomini *et al.* (2013) for evaluating the bootstrap method.

A summary of the results for 5% level bootstrap tests is given in Tables S.10–S.14 in Appendix S.1 in the supplementary material. The patterns of the results at the other levels of significance, for example 10% and 2.5%, were similar to those at the 5% level, and hence those results are not given, but are available from the authors. The results on Type I error rates indicate that all the tests performed well in terms of size (see Table S.10). However, in terms of power (see Tables S.11–S.14), the tests proposed in this paper performed significantly better than JG and FG tests. In particular, the Anderson-Darling type A^2 test, exhibited the best overall performance, followed by the Cramér-von Mises type test CvM.

In a related study, Meintanis *et al.* (2017) provide a simulation study of bootstrap tests for the error distribution in MEM(1,1). They considered several tests, including A^2, CvM, and KS tests, which are similar to the ones in our study, and hence the results in Meintanis *et al.* (2017) could be used for comparison with ours. It is reassuring to note that the general nature of their results agree with ours. In this regard, both studies note that A^2 generally performed the best among the empirical process based tests, and the FG-test exhibited low power. Our tests also performed well under violations of the moment condition $\mathbb{E}[\varepsilon_0^{2+d}] < \infty$ (see Table S.12) and when stationarity does not hold (see Tables S.12 and S.13). Thus, the simulation results indicate that even if some of the regularity conditions required for the first order validity of the tests are not satisfied, the bootstrap tests may still perform well. However, it should be noted that, under violations of the regularity conditions, the results we provide on the asymptotic validity of the bootstrap tests cannot be applied.

4 AN EMPIRICAL ILLUSTRATION

In this section we illustrate an application of the probability forecasts and the testing procedure using a real data example. The variable of interest is a measure of daily annualized realized volatility, $\{Z_i\}$, constructed from intraday spot price data for the S&P500 index.[1] The dataset that we consider spans the period September 12, 2005 to December 17, 2012. The first 1518 observations (from September 12, 2005 to September 20, 2011) are used as the initial dataset for fitting a parametric MEM. The rest of the sample (from September 21, 2011 to December 17, 2012) is reserved to evaluate predictions for Z_{n+k} and Ψ_{n+k} using both parametric and semiparametric bootstrap methods.

Fitting a parametric MEM

In order to fit a parametric model to the entire conditional distribution of Z_i, we consider the Asymmetric MEM(1,1) in (21) for Ψ_i, and proceed to evaluate it in combination with several distributions for the error term; namely, Exponential[E], Weibull[W], Gamma[G], Generalized-gamma[GG] and Burr[B]. The results of applying the specification tests developed in this paper, together with the FG and JG tests

[1] The data for this example were obtained from Gael Martin. The raw index data have been cleaned using methods similar to those of Brownlees and Gallo (2006). For details regarding data handling and the construction of realized volatility, see Maneesoonthorn *et al.* (2012).

discussed in the previous section, are given in Table 1. The p-values for the tests were computed by applying the bootstrap algorithm in Subsection 2.4. These results show that models other than $AM - B$ do not fit well, where $AM - B$ stands for the Asymmetric MEM(1,1) with Burr conditional distribution as in Section 3. The p-value for $AM - B$ ranges from 0.08 to 0.76, depending on the test.

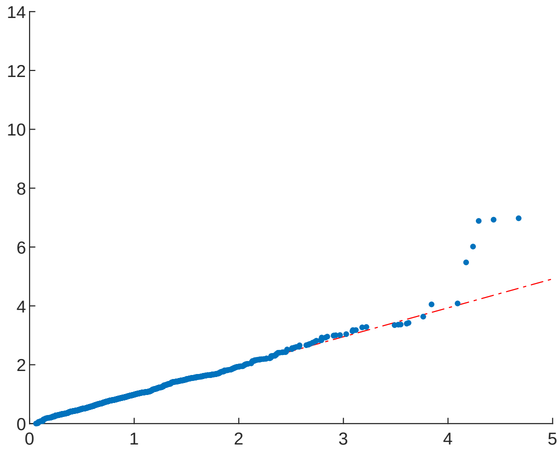
Table 1: The p-values for specification tests of different multiplicative error models for the S&P 500 realized volatility data.

	Tests						
Null model	JG	FG	CvM	U^2	A^2	KS	Ku
$AM - E$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$AM - W$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$AM - G$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$AM - GG$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$AM - B$	0.50	0.76	0.19	0.26	0.16	0.08	0.14

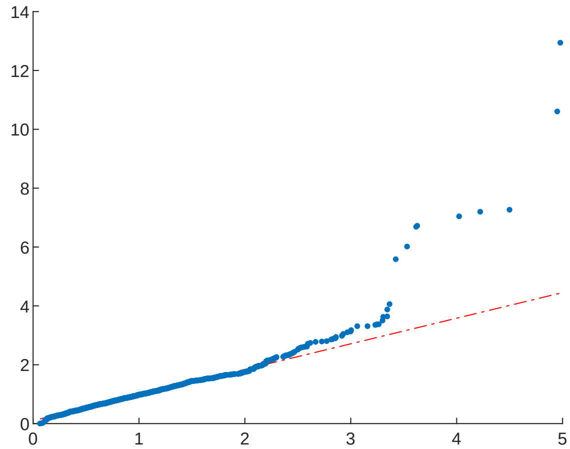
Note: The conditional mean specification $AM : \Psi_i(\phi) = \phi_1 + \phi_2\Psi_{i-1}(\phi)\{|\varepsilon_{i-1} - \phi_4| + \phi_5(\varepsilon_{i-1} - \phi_4)\} + \phi_3\Psi_{i-1}(\phi), \varepsilon_{i-1} = \varepsilon_{i-1}(\phi) = Z_{i-1}/\Psi_{i-1}(\phi),\ \phi = (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)^\top$. The tests CvM, U^2, A^2, KS and Ku are described in Section 3. The JG and FG tests are, Janssen *et al.* (2005) [JG] and the D-test of Fernandes and Grammig (2005) [FG].

Apart from the p-values, it is also of interest to see graphically the extent to which the fitted and the empirical distributions of the residuals differ. To this end we constructed the QQ-plots of the residuals for different distributions. Figure 1 shows the QQ plots for evaluating the goodness-of-fit of Generalized gamma and Burr error distributions, after fitting the Asymmetric MEM$(1, 1)$ model for the conditional mean. The plot for the Burr distribution appears to be significantly closer to a straight line than that for the Generalized gamma which is consistent with the conclusion based on the specification tests. The QQ plots for Exponential, Weibull and Gamma distributions were also constructed (not shown here). They exhibited systematic departures from a straight line, and hence there are indications that the conditional distribution of Z_i is not Exponential, Weibull, Gamma or Generalized gamma. The results of the bootstrap tests given in Table 1 are consistent with this observation.

Figure 1: QQ-plots for evaluating the goodness-of-fit of the error distributions; (a): Burr, and (b): Generalized gamma, after fitting the Asymmetric MEM(1,1) model to the conditional mean process of the realized volatility data.



(a): QQ plot for Burr distribution



(b): QQ plot for generalized Gamma distribution

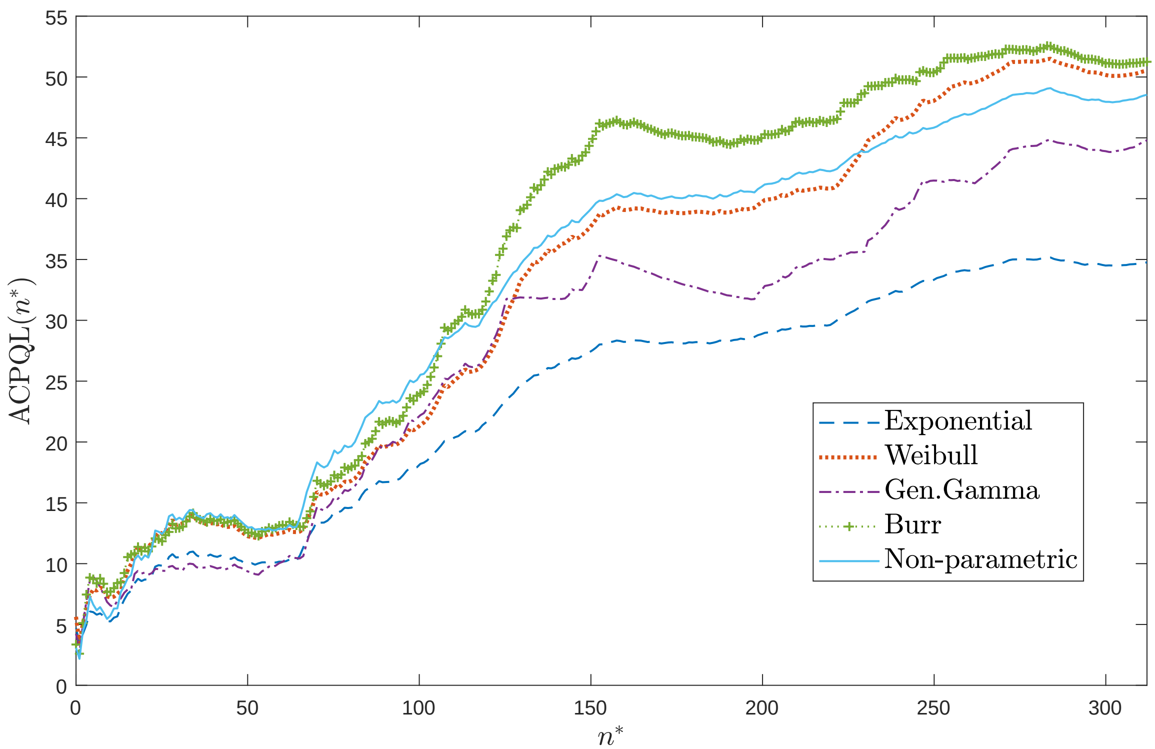
Out of sample forecasting performance: Since our main objective is to choose a suitable model for forecasting, it is also of interest to evaluate the models in terms of out-of-sample forecasting performance. To this end, we compute the *Average Cumulative Predictive Likelihood* [ACPL] for the out-of-sample period. To introduce the definition for ACPL let $\{Z_1, \ldots, Z_n\}$ denote the initial dataset (from Sep 12, 2005 to Sep 20, 2011), with $n = 1518$. Let $N^* = 312$ denote the number of out-of-sample observations. First, we estimate the Asymmetric MEM(1,1) model under each of the error distributions by using the observations $\{Z_1, \ldots, Z_{n+j-1}\}$ for $j = 1, 2, \ldots, N^*$, thus expanding the time period by 1 for each new observation in the out-of-sample period. For $j = 1, 2, \ldots, N^*$, we compute the predictive (conditional) mean $\widehat{\Psi}_{n+j}(\hat{\phi})$, for the next period $(n+j)$ by using the data $\{Z_1, \ldots, Z_{n+j-1}\}$ up to the current time $(n+j-1)$ and the in-sample parameter estimates. Then ACPL is defined as

$$\mathrm{ACPL}(n^*) := (n^*)^{-1} \sum_{j=n+1}^{n+n^*} \{\widehat{\Psi}_j(\hat{\phi})\}^{-1} f_{\hat{\theta}}[Z_j/\widehat{\Psi}_j(\hat{\phi})], \tag{24}$$

for $n^* = 1, \ldots, N^*$; for example, see Chan *et al.* (2014). For each n^*, ACPL(n^*) gives the average cumulative predictive likelihood based on the observed $\{Z_{n+1}, \ldots, Z_{n+n^*}\}$. A large value of ACPL indicates better predictive ability of the underlying model. Figure 2 provides a plot of ACPL(n^*) against n^* for several parametric models, and a nonparametric kernel density estimator obtained by using the asymmetric gamma kernel approach, as proposed by Chen (2000), based on QMLE residuals $(n^* = 1, \ldots, N^*)$.

Ranking of the models in terms of ACPL values turned out to be consistent with the p-values given by the specification tests (see Table 1).

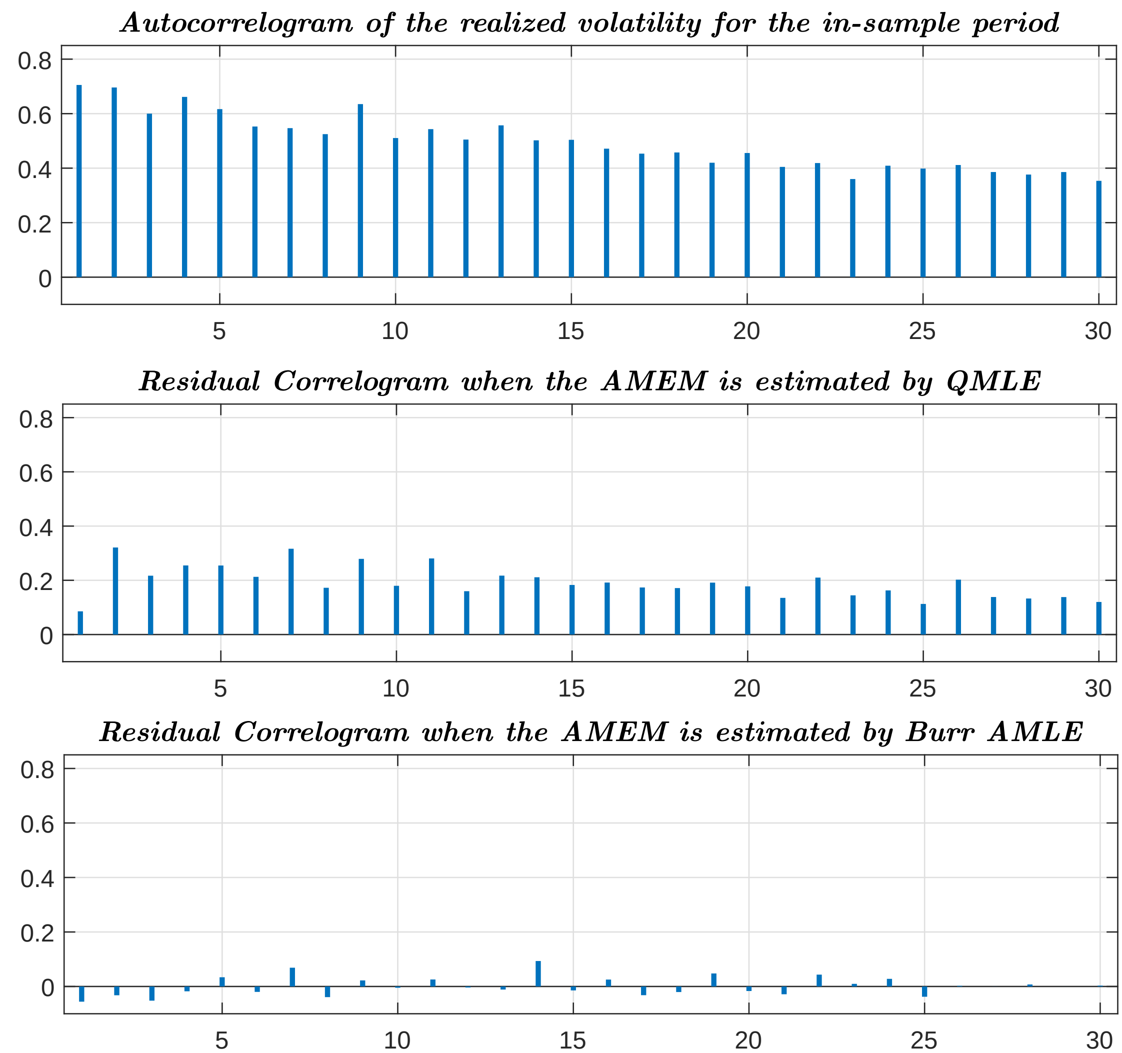
Figure 2: Average Cumulative Predictive Likelihood [ACPL] for the Asymmetric MEM(1,1) model with F: Exponential, Weibull, generalized-Gamma, Burr, and a non-parametric kernel density estimator obtained based on QMLE residuals.



Predictions for Z_{n+k} and Ψ_{n+k}

We evaluate the performance of the parametric bootstrap method in Section 2.1, based on the Burr-AMLE, in comparison with the semiparametric bootstrap method in Section 2.2, over the out-of-sample period. First to evaluate the two estimation methods for the in-sample period, in Figure 3, we compare the autocorrelogram of the realized volatility with the residual correlograms for the Asymmetric MEM(1,1): (a) when the model is estimated by the QMLE based on the exponential distribution, and (b) when the model is estimated by the Burr-AMLE. The correlogram for the QMLE residuals indicates a very significant autocorrelation even up to lag 30. By contrast, the autocorrelations of the residuals obtained from the Burr-AMLE are significantly smaller. This may be related to the fact that AMLE is more efficient than the QMLE if the parametric model used by AMLE is correctly specified. Despite its asymptotic consistency, the QML method has been found to perform unsatisfactorily for some parametric forms of Ψ_i even in quite large samples (see Grammig and Maurer, 2000; Fernandes and Grammig, 2005). Thus, our empirical observation appears to be consistent with the observations in the aforementioned literature.

Figure 3: Autocorrelogram of the realized volatility (first panel), and the Residual Correlograms for AMEM [the Asymmetric MEM(1,1)], when estimated by the QMLE (the middle panel) and when estimated by the Burr-AMLE (bottom panel).



We compute prediction intervals and upper prediction bounds for Z_{n+k} and Ψ_{n+k} for the out of sample period consisting of 312 daily annualized realized volatilities from $n = 1519$ to $n = 1830$ of the full sample, for several lead times k. To have a clearer idea about the adequacy of the interval forecasts, we also compute the *Interval Score* proposed by Gneiting and Raftery (2007) to evaluate the accuracy of the interval forecasts over the out-of-sample period. If the forecaster quotes the $(1-\alpha)\times 100\%$

prediction interval $[l, u]$ and x materializes, then the negatively oriented interval score of Gneiting and Raftery (2007) is defined by

$$S_{\alpha}^{int}(l, u; x) := (u - l) + \frac{2}{\alpha}(l - x)I\{x < l\} + \frac{2}{\alpha}(x - u)I\{x > u\}. \quad (25)$$

This scoring rule has intuitive appeal as the forecaster is rewarded for narrow prediction intervals, and incurs a penalty, based on α, if the observation misses the interval.

A summary of the main results is presented in Tables 2–3 and Figures 4–5. The results for the parametric method are based on the aforementioned Burr-AMLE. For the lead times $k = 2, 3$, and 4, the bootstrap interval forecasts produced by the parametric bootstrap method yield good empirical coverage (EC) properties for both Z_{n+k} and Ψ_{n+k}, at 80%, 95%, and 99% levels (see Tables 2 and 3). The corresponding forecasts produced by the semiparametric bootstrap method also exhibit good coverage properties, but not as good as those from the parametric method. In this regard, we make the following observations: the semiparametric method yields good coverage properties at the 99% level and for some cases at the 95% level, but at the 80% level, the ECs for the semiparametric method are all either significantly below/above the desired nominal rate; for example, (a) in the 80% prediction intervals (PIs) for Z_{n+3} (see Table 2), the EC for the semiparametric method is only 70.3%, and, by contrast, for the parametric method it is 81.3% which is much closer to the desired level of 80%; and (b) in the 80% prediction bounds for Z_{n+2} (see Table 3), the semiparametric method yields an EC of 90.0%, and the parametric method yields an EC of 80.4% which is again comparably much closer to the desired nominal level.

In terms of the average width and the average value of the interval score in (25), the interval forecasts produced by the parametric method perform uniformly better than those produced by the semiparametric method (see Tables 2–3 and Figures 4–5). For example, for the 99% PIs for Z_{n+3}, the average width for the semiparametric PIs is over 0.08 higher than that for the parametric PIs (see Table 2). Similar significant differences in average width can also be observed for most of the other cases. Graphical illustrations of such differences in terms of the prediction bounds for Z_{n+2} are provided in Figures 4 and 5. Additional graphical illustrations for other interval forecasts are provided in Figures S.1–S.6 in Appendix S.2 in the supplementary material.

Overall, in terms of the out-of-sample probability forecasts produced in this empirical example, the semiparametric method performs reasonably well, and the parametric method performs better. However, we emphasize that the better performance of the parametric method is not a general result, but an observation in one empirical example. It is almost certainly the case that there are empirical settings in which the semiparametric method would perform better in view of the weaker assumption that it does not require a parametric form for the error distribution.

Table 2: Empirical coverage (EC) percentage, average width (AW), and average Interval Score (AIS) of k-step-ahead bootstrap Prediction Intervals for Z_{n+k} and Ψ_{n+k} for the out of sample period from Sep 21, 2011 to Dec 17, 2012 (i.e. $n = 1519$ to $n = 1830$) at different Nominal Coverage (NC) rates.

		$k=2$			$k=3$			$k=4$		
	NC(%)	EC(%)	AW	AIS	EC(%)	AW	AIS	EC(%)	AW	AIS
					Parametric bootstrap with Burr-AMLE					
Z_{n+k}	99	98.7	0.097	0.141	99.4	0.116	0.128	99.4	0.135	0.145
	95	95.5	0.055	0.086	94.2	0.062	0.098	96.4	0.071	0.084
	80	78.5	0.027	0.049	81.3	0.030	0.051	85.1	0.033	0.050
Ψ_{n+k}	99	98.7	0.047	0.103	99.4	0.068	0.078	99.0	0.088	0.090
	95	94.5	0.027	0.051	94.8	0.040	0.056	95.5	0.049	0.062
	80	79.7	0.014	0.029	80.3	0.020	0.033	84.8	0.024	0.037
					Semiparametric bootstrap with QMLE					
Z_{n+k}	99	99.0	0.174	0.180	98.7	0.198	0.214	99.0	0.219	0.220
	95	91.6	0.084	0.101	91.9	0.096	0.125	90.3	0.101	0.116
	80	73.0	0.039	0.059	70.3	0.042	0.065	68.3	0.043	0.067
Ψ_{n+k}	99	99.4	0.083	0.148	99.7	0.121	0.124	99.7	0.135	0.147
	95	97.1	0.044	0.078	94.2	0.057	0.069	95.1	0.064	0.085
	80	69.8	0.022	0.040	70.0	0.026	0.041	69.3	0.028	0.044

Note: The values under AIS denotes the average value of the negatively oriented interval score (25) for the prediction intervals over the out-of-sample period.

Table 3: Empirical coverage (EC) percentages of k-step-ahead bootstrap Prediction Bounds for Z_{n+k} and Ψ_{n+k} for the out of sample period from Sep 21, 2011 to Dec 17, 2012 (i.e. $n = 1519$ to $n = 1830$) at different Nominal Coverage (NC) rates.

		Parametric Burr-AMLE			Semiparametric QMLE		
	NC(%)	$k=2$	$k=3$	$k=4$	$k=2$	$k=3$	$k=4$
Z_{n+k}	99	98.7	98.4	99.4	99.7	99.0	100.0
	95	95.2	97.4	98.1	97.4	98.4	97.4
	80	80.4	82.3	84.5	90.0	90.3	89.0
Ψ_{n+k}	99	98.7	99.0	99.4	98.7	99.7	99.0
	95	93.6	95.8	96.8	96.1	97.1	97.4
	80	82.6	78.1	81.2	86.5	91.0	90.3

Figure 4: Burr AMLE based bootstrap prediction bounds for S&P500 daily annualized realized volatility (Z_{n+2}) for the out of sample period from Sep 21, 2011 to Dec 17, 2012 (n = 1519 to n = 1830). Empirical coverage rates of the prediction bounds are, *top panel: 80.4%*, *middle panel: 95.2%*, and *bottom panel: 98.7%*.

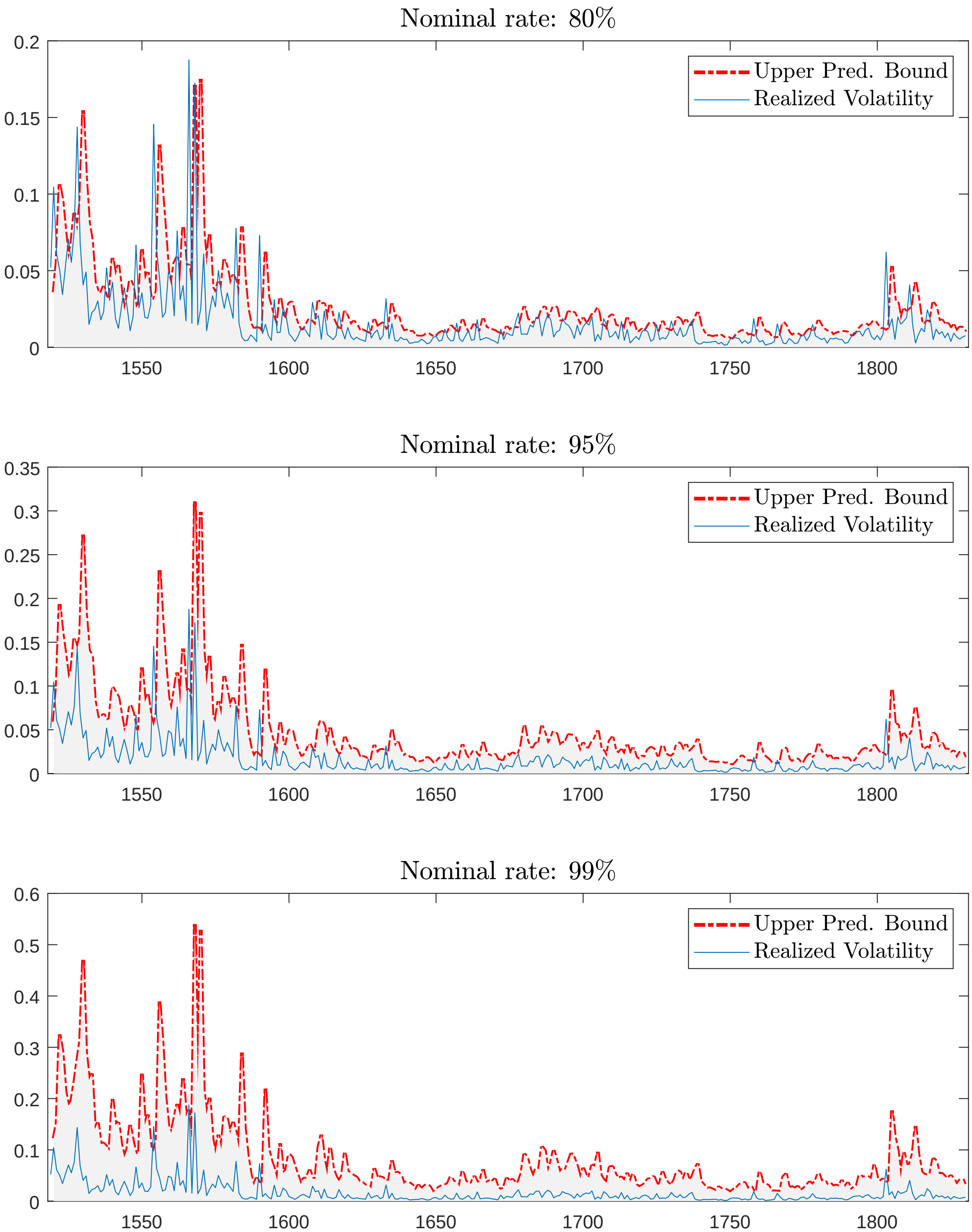
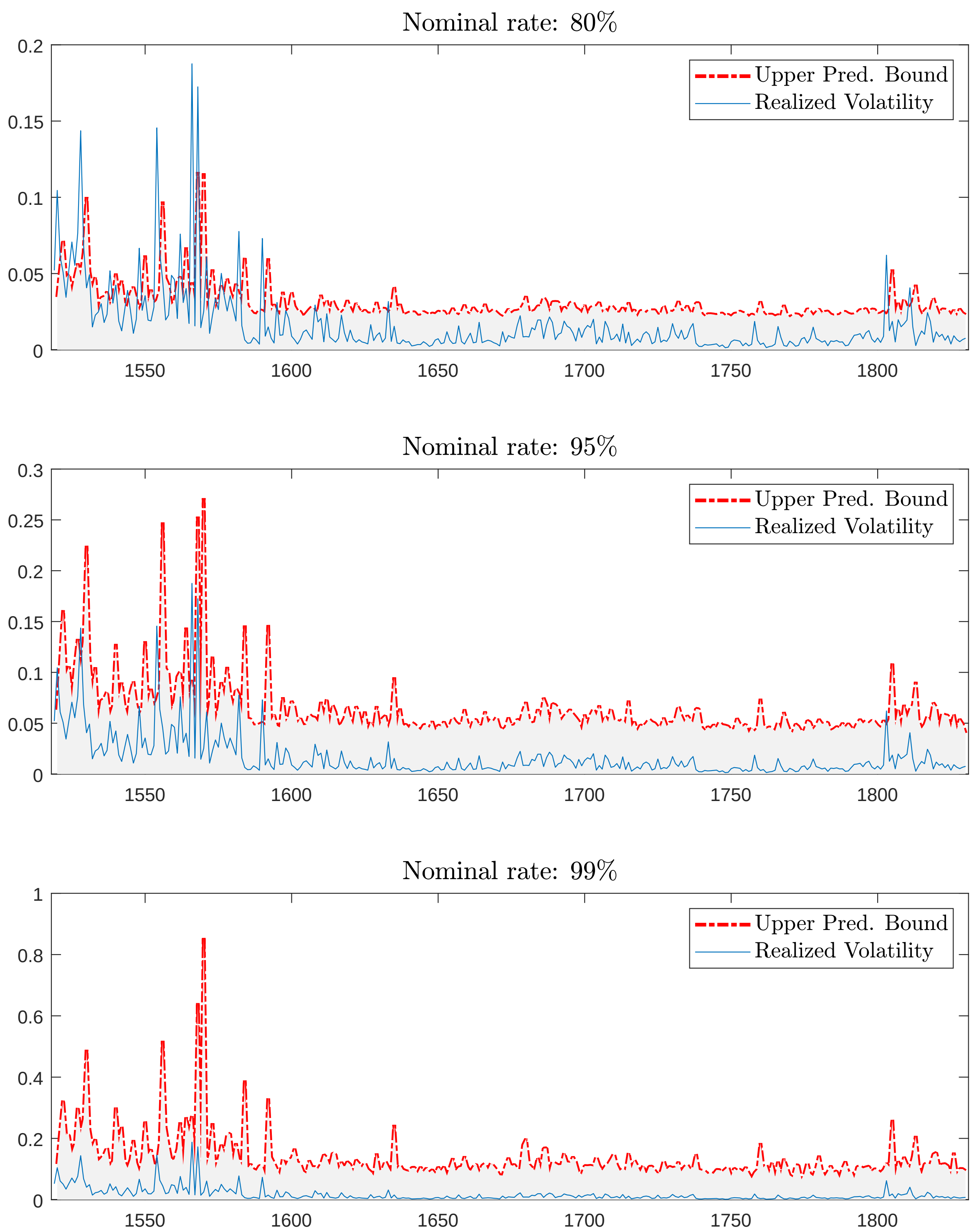


Figure 5: QMLE based semiparametric bootstrap prediction bounds for S&P500 daily annualized realized volatility (Z_{n+2}) for the out of sample period from Sep 21, 2011 to Dec 17, 2012 ($n = 1519$ to $n = 1830$). Empirical coverage rates are, *top panel: 90.0%*, *middle panel: 97.4%*, and *bottom panel: 99.7%*.



5 CONCLUSION

The class of multiplicative error models [MEM] for nonnegative random variables, has been the subject of considerable methodological development. This paper contributes to advance the current state of econometric methodology in MEM for constructing multi-step ahead forecasts of quantities related to a nonnegative valued time series. In particular, we propose a parametric bootstrap procedure to construct probability forecasts for Z_{n+k} and its conditional mean $\varPsi_{n+k}$, given $Z_1, \cdots, Z_n$ $(k = 2, 3, \ldots)$, for a time series $\{Z_i\}$ that obeys a parametric MEM. A variant of the proposed bootstrap method for producing probability forecasts for a semiparametric MEM is also considered. Our bootstrap methods are easy to implement and are flexible enough to be applied to a wide range of MEMs. The proposed bootstrap methods are placed on firm grounds by providing rigorous proofs for their asymptotic validity. The proofs are provided under a set of high-level conditions, which we demonstrate to be realistic by showing that they are satisfied by the well-known family MEM$[p_1, p_2]$. We also consider the problem of fitting a fully parametric MEM, that specifies separate parametric forms for the error distribution and the conditional mean. The test statistics are functionals of an estimated empirical process, and are not asymptotically pivotal. The tests are implemented using a bootstrap method which is shown to be asymptotically valid. Our estimation method is based on an approximate maximum likelihood approach, and hence it is expected to be more efficient than the quasi maximum likelihood estimation under the correct specification. Our bootstrap methods performed well in an extensive simulation study. We illustrate the proposed methods by considering an empirical example based on realized volatility.

A APPENDIX: Regularity conditions and proofs

This appendix provides the proofs for the main results stated in Section 2. We relegate the proofs of some of the results to the supplementary material. First, we state the regularity conditions and obtain several preliminary results.

A.1 Some regularity conditions and preliminary results

We first list several regularity conditions on the family of pdfs $\mathcal{F}' = \{f_\theta : \theta \in \Theta\}$.

B.1. *For each $\theta \in \Theta$ and $x \geq 0$, $f_\theta(x) > 0$.*

B.2. *The map $(\theta, x) \mapsto f_\theta(x)$ is continuous on $\Theta \times [0, \infty)$.*

B.3. *For $\theta_1, \theta_2 \in \Theta$, $f_{\theta_1}(x) = f_{\theta_2}(x)$ for all $x \geq 0$ implies that $\theta_1 = \theta_2$.*

B.4. $\sup_{\zeta \in \Phi \times \Theta} |n^{-1}\widetilde{\Upsilon}_n(\zeta) - n^{-1}\Upsilon_n(\zeta)| \stackrel{a.s.}{\to} 0$ *as* $n \to \infty$.

Conditions B.1 and B.2 are technical assumptions and B.3 is required for the identifiability of the parameter θ. If F_θ is the standard exponential distribution or a member of the standard gamma family (with two parameters), then Condition B.4 follows from (4) (see Straumann and Mikosch, 2006). We expect that this would hold for many other parametric families too. We also make following model assumptions.

B.5. *Model (1) admits a unique stationary ergodic solution $\{(Z_i, \Psi_i)\}$ such that $\Psi_i = \Psi_i(\phi_0)$, $F_0 = F_{\theta_0}$, and $\mathbb{E}\{|\log \Psi_i(\phi_0)|\} < \infty$ with $\phi_0 \in \Phi$ and $\theta_0 \in \Theta$ being interior points.*

B.6. *The family of functions $\{g_\phi, \phi \in \Phi\}$ is bounded from below by some constant $\alpha_L > 0$, such that $g_\phi(\boldsymbol{z}, \boldsymbol{s}) > \alpha_L$, for every $(\boldsymbol{z}, \boldsymbol{s}) \in (\mathbb{R}^+)^{p_1} \times (\mathbb{R}^+)^{p_2}$ and $\phi \in \Phi$. The map $(\phi, \boldsymbol{s}) \mapsto g_\phi(\boldsymbol{z}, \boldsymbol{s})$ is continuous for every $\boldsymbol{z} \in (\mathbb{R}^+)^{p_1}$.*

B.7. *For all $\phi \in \Phi$, $\Psi_0 = \Psi_0(\phi)$ almost surely if and only if $\phi = \phi_0$.*

Conditions B.5 and B.6 are satisfied by most MEMs, including MEM(p_1, p_2) and asymmetric MEM(p_1, p_2), and B.7 is an identifiability assumption. Conditions similar to B.5–B.7 are typically required for the consistency of Gaussian quasi maximum likelihood estimation in closely related GARCH models (see Francq and Zakoïan, 2010).

Proposition 1. *Let $\{Z_i; i \in \mathbb{Z}\}$ be a stationary and ergodic process that obeys the model defined by (1), (2) and (3) with the true value $\zeta_0 = (\phi_0^\top, \theta_0^\top)^\top$. Suppose that B.1–B.7 and (4) are satisfied. Assume additionally that $\mathbb{E}\{|\log f_{\theta_0}(\varepsilon_0)|\} < \infty$. Then, $\hat{\zeta} \stackrel{a.s.}{\to} \zeta_0$ as $n \to \infty$.*

If the almost sure convergence in B.4 is replaced by the weaker assumption

$$\sup_{\zeta \in \Phi \times \Theta} |n^{-1}\widetilde{\Upsilon}_n(\zeta) - n^{-1}\Upsilon_n(\zeta)| \stackrel{p}{\to} 0 \quad \text{as } n \to \infty,$$

then we could obtain that $\hat{\zeta}$ is *weakly* consistent, i.e., $\hat{\zeta} \stackrel{p}{\to} \zeta_0$, instead of $\hat{\zeta} \stackrel{a.s.}{\to} \zeta_0$. This result follows from the proof of Proposition 1 in Appendix A.1.1, once the arguments on the pointwise almost sure convergence $n^{-1}\widetilde{\Upsilon}_n(\zeta) \stackrel{a.s.}{\to} \Upsilon(\zeta)$ are replaced by $n^{-1}\widetilde{\Upsilon}_n(\zeta) \stackrel{p}{\to} \Upsilon(\zeta)$ for $\zeta \in \Phi \times \Theta$, where $\Upsilon(\zeta) = -\mathbb{E}[\log \Psi_0(\phi)] + \mathbb{E}[\log\{f_\theta(Z_0/\Psi_0(\phi))\}]$.

Higher order conditions on $F_{\boldsymbol{\theta}}$, $g_{\boldsymbol{\phi}}$, and $\Psi_i(\boldsymbol{\phi})$

Next we state some higher order conditions that we assume for establishing the validity of bootstrap and the asymptotic normality of the AMLE $\hat{\zeta}$ in (5). Let $\gamma_0(\zeta) = -\chi_0(\zeta)$ and $\upsilon(\phi) = 1/\Psi_0(\phi)$. Then,

$$\gamma_0(\zeta) = \log \upsilon(\phi) + \log\left[f_\theta\{Z_0\upsilon(\phi)\}\right].$$

We assume that $\Psi_0(\phi)$ and $f_\theta(x)$ are twice continuously differentiable. Let us introduce the following notation: For a differentiable function $m(s,x)$ on $\Phi \times \mathbb{R}$ or on $\Theta \times \mathbb{R}$, write $\dot{m}(s,x)$ and $m'(s,x)$ for the derivatives with respect to s and x, respectively. Thus, $\dot{f}_\theta(x) = \partial f_\theta(x)/\partial\theta$, $f_\theta'(x) = \partial f_\theta(x)/\partial x$ and $\dot{\upsilon}(\phi) = \dot{\Psi}_0(\phi)\{\Psi_0(\phi)\}^{-2}$ with $\dot{\Psi}_0(\phi) = [(\partial/\partial\phi_1)\Psi_0(\phi), ..., (\partial/\partial\phi_p)\Psi_0(\phi)]^\top$. $\|\cdot\|$ denotes the Euclidean norm.

The first order derivatives of the function $\gamma_0(\zeta)$ are as follows:

$$\frac{\partial\gamma_0}{\partial\phi} = \frac{\dot{\upsilon}(\phi)}{\upsilon(\phi)} + \frac{f_\theta'\{Z_0\upsilon(\phi)\}Z_0\dot{\upsilon}(\phi)}{f_\theta\{Z_0\upsilon(\phi)\}} \quad \text{and} \quad \frac{\partial\gamma_0}{\partial\theta} = \frac{\dot{f}_\theta\{Z_0\upsilon(\phi)\}}{f_\theta\{Z_0\upsilon(\phi)\}}.$$

One may obtain expressions for the matrices $[\partial^2\gamma_0/\partial\phi\partial\theta]$, $[\partial^2\gamma_0/\partial\theta^2]$, and $[\partial^2\gamma_0/\partial\phi^2]$ similarly. We need to make the following additional regularity assumptions.

D.1. *Assumptions B.1–B.7 and (4) are satisfied, and* $\mathbb{E}\{|\log f_{\theta_0}(\varepsilon_0)|\} < \infty$.

D.2. *The functions* $(\phi, \boldsymbol{z}, \boldsymbol{s}) \mapsto g_\phi(\boldsymbol{z}, \boldsymbol{s})$ *and* $(\theta, x) \mapsto f_\theta(x)$ *are twice continuously differentiable in their domains.*

D.3. $n^{-1/2}\sup_{\zeta\in\Phi\times\Theta}\|[\partial\widetilde{\Upsilon}_n/\partial\zeta] - [\partial\Upsilon_n/\partial\zeta]\| \xrightarrow{a.s.} 0$ *as* $n \to \infty$.

D.4.

$$\mathbb{E}\left[\sup_{\zeta\in\Phi\times\Theta}\left\|\frac{\partial^2\gamma_0(\zeta)}{\partial\zeta^2}\right\|\right] < \infty, \qquad \mathbb{E}\left[\left\|\frac{\partial\gamma_0(\zeta_0)}{\partial\zeta}\right\|^2\right] < \infty.$$

Condition D.2 guarantees that $\Psi_0(\phi)$ and $f_\theta(x)$ are twice continuously differentiable. Condition D.3 is used for ensuring that the maximizer of $\Upsilon_n(\zeta)$ is asymptotically equivalent to $\hat{\zeta}$, and D.4 introduces two moment conditions for the random element γ_0. Similar conditions, in the simpler case where F_θ is the standard exponential distribution, have previously been used in the literature (see Straumann and Mikosch, 2006). The next proposition establishes the asymptotic normality of $\hat{\zeta}$.

Proposition 2. *Let* $\{Z_i; i \in \mathbb{Z}\}$ *be a stationary ergodic process that obeys the model (1), (2) and (3) with the true parameter* $\zeta_0 = (\phi_0^\top, \theta_0^\top)^\top$ *being an interior point of* $\Phi \times \Theta$. *Suppose that Conditions D.1–D.4 are satisfied. Further, assume that*

$$\Sigma_0 := \mathbb{E}\left[\frac{\partial^2\gamma_0}{\partial\zeta^2}(\zeta_0)\right] = -\mathbb{E}\left[\left\{\frac{\partial\gamma_0}{\partial\zeta}(\zeta_0)\right\}^\top\left\{\frac{\partial\gamma_0}{\partial\zeta}(\zeta_0)\right\}\right] \tag{A.1}$$

is negative definite. Then, $n^{1/2}(\hat{\zeta} - \zeta_0) \xrightarrow{d} N(0, -\Sigma_0^{-1})$ *as* $n \to \infty$.

A.1.1 Proofs of Propositions 1 and 2

Lemma 1. *Suppose that the assumptions of Proposition 1 are satisfied. Then, the function* $\Upsilon(\zeta) := -\mathbb{E}[\log \Psi_0(\phi)] + \mathbb{E}[\log\{f_\theta(Z_0/\Psi_0(\phi))\}]$ *is uniquely maximized at* $\zeta = \zeta_0$*.*

Lemma 2. *Suppose that the assumptions of Proposition 2 are satisfied. Let* $\tilde{\zeta}$ *denote the maximizer of the function* $\Upsilon_n(\zeta)$ *in (6). Then,* $n^{1/2}(\tilde{\zeta} - \hat{\zeta}) \overset{a.s.}{\to} 0$ *as* $n \to \infty$*.*

We relegate the proofs of Lemmas 1 and 2 to the supplementary material.

Proof of Proposition 1. From Lemma 1 we have that the objective function $\Upsilon(\zeta)$ is uniquely maximized at $\zeta = \zeta_0$. Note that by Conditions B.2 and B.6, the function

$$\zeta \mapsto \chi_i(\zeta) = \log \Psi_i(\phi) - \log\left[f_\theta\{Z_i/\Psi_i(\phi)\}\right]$$

is continuous on $\Phi \times \Theta$ w.p. 1. Hence, by an application of the ergodic theorem, we obtain that $n^{-1}\sum_{i=1}^{n} -\chi_i(\zeta) \overset{a.s.}{\to} \Upsilon(\zeta)$ as $n \to \infty$, for every (fixed) $\zeta \in \Phi \times \Theta$. By arguing as in the proof of Lemma 3.11 of Pfanzagl (1969), one obtains that the function Υ is upper semicontinuous on $\Phi \times \Theta$ and that $\limsup_{n\to\infty} \sup_{\zeta \in K} n^{-1}\Upsilon_n(\zeta) \le \sup_{\zeta\in K} \Upsilon(\zeta)$ w.p. 1 for any compact subset $K \subset \Phi \times \Theta$. By Condition B.4, it also follows that $\limsup_{n\to\infty} \sup_{\zeta\in K} n^{-1}\widetilde{\Upsilon}_n(\zeta) \le \sup_{\zeta \in K} \Upsilon(\zeta)$, w.p. 1. Since, an upper semicontinuous function attains its maximum on compact sets, it follows by the arguments of the proof of Theorem 4.1 of Straumann and Mikosch (2006) that $\hat{\zeta} \overset{a.s.}{\to} \zeta_0$ as $n \to \infty$. □

Proof of Proposition 2. From Lemma 2, we obtain that $n^{1/2}(\tilde{\zeta} - \hat{\zeta}) \overset{a.s.}{\to} 0$. Therefore, it suffices to establish the asymptotic normality of $\tilde{\zeta}$. Since $\dot{\Upsilon}_n(\tilde{\zeta}) = 0$ and $\dot{\Upsilon}_n(\zeta_0) = -\sum_{i=1}^{n}[\partial\chi_i(\zeta_0)/\partial\zeta]$, from Taylor series expansion, $n^{-1}\ddot{\Upsilon}_n(\zeta_n^*)n^{1/2}(\tilde{\zeta} - \zeta_0) = n^{-1/2}\sum_{i=1}^{n}\dot{\chi}_i(\zeta_0)$, for some ζ_n^* satisfying $\|\zeta_n^* - \zeta_0\| \le \|\tilde{\zeta} - \zeta_0\|$, where $\dot{\chi}_i(\zeta) = [\partial\chi_i(\zeta)/\partial\zeta]$. Since $\zeta_n^* \overset{a.s.}{\to} \zeta_0$, we obtain by arguing as before that $n^{-1}\ddot{\Upsilon}_n(\zeta_n^*) \overset{a.s.}{\to} \Sigma_0$. Since Σ_0 is invertible and $n^{-1/2}\sum_{i=1}^{n}\dot{\chi}_i(\zeta_0) = O_p(1)$ then it follows that

$$n^{1/2}(\tilde{\zeta} - \zeta_0) = n^{-1/2}\sum_{i=1}^{n}\varphi_i(\zeta_0) + o_p(1), \quad \varphi_i(\zeta) := \Sigma_0^{-1}\dot{\chi}_i(\zeta). \tag{A.2}$$

From a martingale central limit theorem (Theorem 18.3 in Billingsley, 1999) followed by an application of Slutsky's lemma one obtains that $n^{-1/2}\sum_{i=1}^{n}\varphi_i(\zeta_0) \overset{d}{\to} N(0, -\Sigma_0^{-1})$ as $n \to \infty$, and hence the proof follows. □

A.2 Proofs of Theorems 1 and 2

First, we obtain several preliminary results required for the main proofs. Recall that, the data generating model in (15) for a given point $\zeta = (\phi, \theta)^\top \in \Phi \times \Theta$, is

$$Z_i^{(\zeta)} = \Psi_i^{(\zeta)}(\phi)\varepsilon_i(\theta), \quad \Psi_i^{(\zeta)}(\phi) = g_\phi\left(Z_{i-1}^{(\zeta)}, \cdots, Z_{i-p_1}^{(\zeta)}, \Psi_{i-1}^{(\zeta)}(\phi), \cdots, \Psi_{i-p_2}^{(\zeta)}(\phi)\right), \quad i \in \mathbb{Z},$$

where $\varepsilon_i(\theta) = F_\theta^{-1}(U_i) = \inf\{x \geq 0 : U_i \leq F_\theta(x)\}$. In what follows, $P^{(\zeta)}$ and $\mathbb{E}^{(\zeta)}$ denote the corresponding *Probability* and *Expectation*, respectively. The terms derived from $(Z_i^{(\zeta)}, \Psi_i^{(\zeta)})$, such as $\widetilde{\Psi}_i^{(\zeta)}$ and $\lambda_i^{(\zeta)} = \dot{\Psi}_i^{(\zeta)}/\Psi_i^{(\zeta)}$, are denoted with the superscript "(ζ)", highlighting the fact that the data generating process corresponds to ζ. Further, for brevity, for any given non-random sequence $\zeta_n = (\phi_n^\top, \theta_n^\top)^\top$ where $\zeta_n \to \zeta_0$, we write $(Z_{ni}, \Psi_{ni}, \varepsilon_{ni})$ for $(Z_i^{(\zeta_n)}, \Psi_i^{(\zeta_n)}, \varepsilon_i^{(\zeta_n)})$, with the probability and the expectation operators being denoted by P_n and $\mathbb{E}_n$, respectively.

First, we state two lemmas (Lemmas 3 and 4), and relegate the proofs to Appendix S.3 in the supplementary material.

Lemma 3. *Suppose that B.5, B.6, and B.7 are satisfied, and* $\mathbb{E}[\|\lambda_i(\phi_0)\|^2] < \infty$. *Then, we have that* $\max_{1\leq i\leq n} \|\lambda_i(\phi_0)\| = o_p(n^{1/2})$. *Additionally, assume that Conditions C.1 and C.3 are also satisfied. Then,* $\max_{1\leq i\leq n} \|\lambda_{ni}\|_{K_1} = o_{p_n}(n^{1/2})$.

Lemma 4. *Suppose that B.5, B.6, and B.7 are satisfied, and the function* $(\phi, \boldsymbol{z}, \boldsymbol{s}) \mapsto g_\phi(\boldsymbol{z}, \boldsymbol{s})$ *is twice continuously differentiable. Let* B *be an open neighbourhood of* ϕ_0 *such that* $\mathbb{E}[\|\ddot{\Psi}_0\|_B^{2+d}] < \infty$ *for some* $d > 0$. *Let* $0 < M < \infty$. *Then,*
(a) $n^{1/2} \sup |\Psi_i(t) - \Psi_i(s) - (t-s)^\top \dot{\Psi}_i(s)|/\Psi_i(\phi_0) = o_p(1)$, *where the supremum is taken over* $1 \leq i \leq n$, *and over* $\{(t,s) : t, s \in B, \sqrt{n}\|t-s\| \leq M\}$.
Additionally, assume that Conditions C.1 and C.3 are satisfied and $B \subset K_1$, *where* K_1 *is as in C.3. Then,*
(b) $n^{1/2} \sup |\Psi_{ni}(t) - \Psi_{ni}(s) - (t-s)^\top \dot{\Psi}_{ni}(s)|/\Psi_{ni}(\phi_n) = o_{p_n}(1)$, *with the supremum being taken over the same domain as in (a).*

In what follows, $d_2(F_X, F_Y)$ denotes the Mallows metric for the distance between two probability distributions F_X and F_Y defined by $d_2(F_X, F_Y) = \inf\{\mathbb{E}|X-Y|^2\}^{1/2}$, where the infimum is over all square integrable random variables X and Y with marginal distributions F_X and F_Y. Further, the bootstrap convergence results are "in probability". In particular, the orders $o_{p_n^*}(1)$ and $O_{p_n^*}(1)$ are defined as follows: (a) $X_n^* = o_{p_n^*}(1)$ if $P_n^*\{|X_n^*| > \delta\} \to_p 0$ for all $\delta > 0$, (b) $X_n^* = O_{p_n^*}(1)$ if for any $\delta > 0$, there exists a finite $M > 0$ such that $P\{P_n^*(|X_n^*| > M) < \delta\} \longrightarrow 1$ as $n \to \infty$.

Next, we provide the proof of Theorem 1.

Proof of Theorem 1. Let $G_{Z,k}(x)$ denote the distribution function of Z_{n+k}, conditional on $\{Z_1, \ldots, Z_n\}$. Then, from the definition of the Mallows metric and applying the Cauchy-Schwarz inequality, we obtain that

$$\begin{aligned}
[d_2(G_{\check{Z}^*,k}, G_{Z,k})]^2 &\leq \inf_{\varepsilon^*\sim F_{\hat{\theta}},\ \varepsilon\sim F_{\theta_0}} \mathbb{E}[\mathbb{E}^*\{\check{\Psi}^*_{n+k}(\hat{\phi}^*)\varepsilon^* - \Psi_{n+k}(\phi_0)\varepsilon\}^2] \\
&\leq \mathbb{E}[\mathbb{E}^*\{\check{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)\}^2] + \mathbb{E}\{\Psi^2_{n+k}(\phi_0)\} \inf_{\varepsilon^*\sim F_{\hat{\theta}},\ \varepsilon\sim F_{\theta_0}} \mathbb{E}[\mathbb{E}^*(\varepsilon^* - \varepsilon)^2]
\end{aligned}$$

$$
\begin{aligned}
&\le \mathbb{E}[\mathbb{E}^*\{\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)\}^2] + \mathbb{E}\{\Psi^2_{n+k}(\phi_0)\}[d_2(F_{\hat{\theta}}, F_{\theta_0})]^2 \\
&= \mathbb{E}[\mathbb{E}^*\{\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)\}^2] + o_p(1). \qquad \text{(A.3)}
\end{aligned}
$$

Let Λ be a compact neighbourhood of ϕ_0 that satisfy (13). By using C.1, C.2 and C.3 we obtain that $n^{1/2}(\hat{\phi}^* - \phi_0) = O_{p_n^*}(1)$, and hence, from (4) and Lemma 4, it follows that $|\Psi_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)| = o_{p_n^*}(1)$. Therefore, $\mathbb{E}\{\mathbb{E}^*[|\Psi_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)|]\} = o(1)$, and hence,

$$
\begin{aligned}
&\mathbb{E}\{\mathbb{E}^*[|\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)|]\} \\
&\quad \le \mathbb{E}\{\mathbb{E}^*[|\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\hat{\phi}^*)|]\} + \mathbb{E}\{\mathbb{E}^*[|\Psi_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)|]\} \\
&\quad \le \mathbb{E}\{\mathbb{E}^*[\sup_{\phi\in\Lambda} |\Psi^{**}_n(\phi) - \Psi_n(\phi)|]\} + o(1),
\end{aligned}
$$

where the last inequality follows from (14). Because $\breve{\Psi}^*_n(\cdot) = \widetilde{\Psi}_n(\cdot)$, one obtains from (13) that $\mathbb{E}\{\mathbb{E}^*[\sup_{\phi\in\Lambda} |\Psi^{**}_n(\phi) - \widetilde{\Psi}_n(\phi)|]\} = o(1)$. Further, it follows from (4) that $\sup_{\phi\in\Lambda} |\widetilde{\Psi}_n(\phi) - \Psi_n(\phi)| \overset{e.a.s.}{\to} 0$ as $n \to \infty$. Therefore,

$$
\begin{aligned}
&\mathbb{E}\{\mathbb{E}^*[\sup_{\phi\in\Lambda} |\Psi^{**}_n(\phi) - \Psi_n(\phi)|]\} \\
&\quad \le \mathbb{E}\{\mathbb{E}^*[\sup_{\phi\in\Lambda} |\Psi^{**}_n(\phi) - \widetilde{\Psi}_n(\phi)|]\} + \mathbb{E}\{\sup_{\phi\in\Lambda} |\widetilde{\Psi}_n(\phi) - \Psi_n(\phi)|\} = o(1).
\end{aligned}
$$

Consequently,

$$
\mathbb{E}\{\mathbb{E}^*[|\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)|]\} = o(1). \qquad \text{(A.4)}
$$

Next, consider the event $A_n = \{|\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)| \le 1\}$. On the event A_n,

$$
\mathbb{E}[\mathbb{E}^*\{\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)\}^2] \le \mathbb{E}[\mathbb{E}^*\{|\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)|\}].
$$

From (A.4) and applying Markov inequality, we obtain that

$$
P^*_n(|\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)| > 1) \le \mathbb{E}^*(|\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)| > 1) = o_p(1).
$$

Therefore, $P^*_n(A_n) \to_p 1$ as $n \to \infty$. Hence, it follows from (A.4) that

$$
\mathbb{E}[\mathbb{E}^*\{\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)\}^2] = o(1). \qquad \text{(A.5)}
$$

This together with (A.3) show that $d_2(G_{\breve{Z}^*,k}, G_{Z,k}) = o_p(1)$. Thus, the Part (a) of Theorem 1 follows.

Now, to prove Part (b), let $G_{\Psi,k}(x)$ be the distribution function of Ψ_{n+k} conditional on $\{Z_1, \ldots, Z_n\}$. By arguing as for (A.3), one obtains that

$$
[d_2(G_{\breve{\Psi}^*,k}, G_{\Psi,k})]^2 \le \mathbb{E}[\mathbb{E}^*\{\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)\}^2] + o_p(1).
$$

Further, as in (A.5), we have that $\mathbb{E}[\mathbb{E}^*\{\breve{\Psi}^*_{n+k}(\hat{\phi}^*) - \Psi_{n+k}(\phi_0)\}^2] \to 0$ as $n \to \infty$. Hence, Part (b) of Theorem 1 also follows. $\square$

We will next provide the proof of Theorem 2.

Let $\{F_n\}$ denote a sequence of cdf's in $D[0,\infty)$ with each F_n having unit mean and finite variance. The semi-parametric analogue of the data generating model (15) at a non-random (ϕ_n, F_n), where $(\phi_n, F_n) \to (\phi_0, F_0)$ as $n \to \infty$, is given by

$$\begin{aligned} Z_i^{(\phi_n,F_n)} &= \Psi_i^{(\phi_n,F_n)}(\phi_n)\varepsilon_i^{F_n}, \\ \Psi_i^{(\phi_n,F_n)}(\phi) &= g_\phi\big(Z_{i-1}^{(\phi_n,F_n)},\cdots,Z_{i-p_1}^{(\phi_n,F_n)},\Psi_{i-1}^{(\phi_n,F_n)}(\phi),\cdots,\Psi_{i-p_2}^{(\phi_n,F_n)}(\phi)\big), \quad i \in \mathbb{Z}, \end{aligned} \tag{A.6}$$

where $\varepsilon_i^{F_n} = F_n^{-1}(U_i) := \inf\{x \geq 0 : U_i \leq F_n(x)\}$, for $i \in \mathbb{Z}$. In what follows, for brevity, we write $(Z_{ni}^\dagger, \Psi_{ni}^\dagger, \varepsilon_{ni}^\dagger)$ for $(Z_i^{(\phi_n,F_n)}, \Psi_i^{(\phi_n,F_n)}, \varepsilon_i^{F_n})$, and continue to use P_n and $\mathbb{E}_n$ for the corresponding probability and expectation operators.

The next lemma extends Lemmas 3 and 4 to the setting in (A.6). We relegate the proof of this lemma to Appendix S.3 in the supplementary material.

Lemma 5. *Suppose that B.6 is satisfied, and $(\phi, \boldsymbol{z}, \boldsymbol{s}) \mapsto g_\phi(\boldsymbol{z}, \boldsymbol{s})$ is twice continuously differentiable. Further, assume that there exist a compact neighbourhood K_1 of ϕ_0 such that analogues of C.1 and C.3 hold for the DGP in (A.6) and the QMLE (10). Then,*
(a) $\max_{1\leq i\leq n} \|\lambda_{ni}^\dagger\|_{K_1} = o_{p_n}(n^{1/2})$, where $\lambda_{ni}^\dagger := \dot{\Psi}_{ni}^\dagger(\phi_n)/\Psi_{ni}^\dagger(\phi_n)$.
(b) Let $B \subset K_1$ be an open neighbourhood of ϕ_0 such that $\mathbb{E}[\|\ddot{\Psi}_0\|_B^{2+d}] < \infty$, $d > 0$. Let $0 < M < \infty$. Then, $n^{1/2} \sup |\Psi_{ni}^\dagger(t) - \Psi_{ni}^\dagger(s) - (t-s)^\top \dot{\Psi}_{ni}^\dagger(s)|/\Psi_{ni}^\dagger(\phi_n) = o_{p_n}(1)$, where the supremum is taken over $1 \leq i \leq n$, and over $\{(t,s) : t, s \in B, \sqrt{n}\|t-s\| \leq M\}$.

A key result required for the proof of Theorem 2 is to show that the empirical distribution function of the standardized residuals, say $\widetilde{F}_{n,qml}$, converges to the unknown error distribution F_0. Recall that $\tilde{\varepsilon}_t^{(std)} = \{n^{-1}\sum_{i=1}^n \tilde{\varepsilon}_i^{(qml)}\}^{-1}\tilde{\varepsilon}_t^{(qml)}$, $t = 1,\ldots,n$, and

$$\widetilde{F}_{n,qml}(x) := n^{-1}\sum_{i=1}^n I(\tilde{\varepsilon}_i^{(std)} \leq x), \quad x \geq 0.$$

The next lemmas show that $\widetilde{F}_{n,qml}$ converges in probability to F_0 as $n \to \infty$.

Lemma 6. *Suppose that the assumptions of Theorem 2 are satisfied. Then, we have that $d_2(\widetilde{F}_{n,qml}, F_0) \to_p 0$ as $n \to \infty$.*

Proof of Lemma 6. Let $G_n(x) = n^{-1}\sum_{i=1}^n I(\varepsilon_i \leq x)$, be the empirical distribution function of the unobserved errors $\{\varepsilon_1,\ldots,\varepsilon_n\}$. From the triangular inequality we have

$$d_2(\widetilde{F}_{n,qml}, F_0) \leq d_2(\widetilde{F}_{n,qml}, G_n) + d_2(G_n, F_0).$$

It can be shown that $d_2(G_n, F_0) \overset{a.s.}{\to} 0$ as $n \to \infty$ (see, for example, Lemma 8.4 of Bickel and Freedman, 1981).

Thus, it suffices to show that $d_2(\widetilde{F}_{n,qml}, G_n) \to_p 0$ as $n \to \infty$. To this end, let J be a random variable having Laplace distribution on $\{1, \ldots, n\}$, with $P(J = i) = 1/n$ for each $i = 1, \ldots, n$. Define two random variables $X^{(1)}$ and $Y^{(1)}$ by

$$X^{(1)} = \varepsilon_J \text{ and } Y^{(1)} = \tilde{\varepsilon}_J^{(std)}.$$

Then, $X^{(1)}$ and $Y^{(1)}$ have the marginal distributions G_n and $\widehat{F}_n$ respectively. Hence

$$\begin{aligned}\{d_2(\widetilde{F}_{n,qml}, G_n)\}^2 &= \inf\{E|X - Y|^2\} \le E\{X^{(1)} - Y^{(1)}\}^2 \\ &= n^{-1} \sum_{i=1}^{n} (\varepsilon_i - \tilde{\varepsilon}_i^{(std)})^2 = (n\tilde{\mu}_n^2)^{-1} \sum_{i=1}^{n} \{\tilde{\mu}_n \varepsilon_i - \tilde{\varepsilon}_i^{(qml)}\}^2, \qquad \text{(A.7)}\end{aligned}$$

where $\tilde{\mu}_n := n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_i^{(qml)}$.

Let

$$\mathcal{E}_{1i} := \frac{\widetilde{\Psi}_i(\hat{\phi}_{qml}) - \Psi_i(\hat{\phi}_{qml})}{\widetilde{\Psi}_i(\hat{\phi}_{qml})\Psi_i(\hat{\phi}_{qml})}, \quad \mathcal{E}_{2i} := \frac{\Psi_i(\hat{\phi}_{qml}) - \Psi_i(\phi_0)}{\Psi_i(\hat{\phi}_{qml})\Psi_i(\phi_0)}.$$

Then, we have that

$$\begin{aligned}\tilde{\varepsilon}_i^{(qml)} &= Z_i\left(\frac{1}{\widetilde{\Psi}_i(\hat{\phi}_{qml})} - \frac{1}{\Psi_i(\hat{\phi}_{qml})}\right) + Z_i\left(\frac{1}{\Psi_i(\hat{\phi}_{qml})} - \frac{1}{\Psi_i(\phi_0)}\right) + \frac{Z_i}{\Psi_i(\phi_0)} \\ &= -Z_i\mathcal{E}_{1i} - Z_i\mathcal{E}_{2i} + \varepsilon_i. \qquad \text{(A.8)}\end{aligned}$$

Let B be the compact neighbourhood of ϕ_0 in condition L.3. Note that, by condition B.6, we have $\Psi_i(\phi), \widetilde{\Psi}_i(\phi) > \alpha_L > 0$. Hence, on the event $\{\hat{\phi}_{qml} \in B\}$,

$$|\ \mathcal{E}_{1i}\ | \le\ \alpha_L^{-2}\ |\ \widetilde{\Psi}_i(\hat{\phi}_{qml}) - \Psi_i(\hat{\phi}_{qml})\ | \le\ \alpha_L^{-2} \sup_{\phi \in B} |\ \widetilde{\Psi}_i(\phi) - \Psi_i(\phi)\ | \xrightarrow{e.a.s.} 0, \text{ as } i \to \infty.$$

Therefore, an application of Lemma 2.1 of Straumann and Mikosch (2006) yields that $Z_i\mathcal{E}_{1i} \xrightarrow{e.a.s.} 0$ as $i \to \infty$, and hence, $n^{-1} \sum_{i=1}^{n} Z_i\mathcal{E}_{1i} \xrightarrow{a.s.} 0$ as $n \to \infty$.

Next, we study the limiting behaviour of $n^{-1} \sum_{i=1}^{n} Z_i\mathcal{E}_{2i}$. First, note that, from Lemmas 3 and 4 we have

L.4. *For every constant $M > 0$, $\sup \sqrt{n}\ |\ \Psi_i(t) - \Psi_i(s) - (t-s)^\top \dot{\Psi}_i(s)\ |\ /\Psi_i(\phi_0) = o_p(1)$, where the supremum is taken over $1 \le i \le n$ and $\{(t, s) : t, s \in \Phi, \sqrt{n}\|t - s\| \le M\}$. Furthermore, $\max_{1 \le i \le n} n^{-1/2} \|\lambda_i(\phi_0)\| = o_p(1)$.*

Let $\delta > 0$ be fixed but arbitrary constant. Since $n^{1/2}(\hat{\phi}_{qml} - \phi_0) = O_p(1)$, there exists a constant $M > 0$ such that $P(n^{1/2}\|\hat{\phi}_{qml} - \phi_0\| \le M) \ge 1 - \delta$ for all n sufficiently large. Hence, in view of L.4, on the event $\{n^{1/2}\|\hat{\phi}_{qml} - \phi_0\| \le M\}$, we have that

$$\begin{aligned}n^{1/2} \max_{1 \le i \le n} |\ \mathcal{E}_{2i}\ | &\le n^{1/2}\alpha_L^{-1} \max_{1 \le i \le n} \left|\frac{\Psi_i(\hat{\phi}_{qml}) - \Psi_i(\phi_0)}{\Psi_i(\phi_0)}\right| \\ &\le \alpha_L^{-1} \max_{1 \le i \le n} \|n^{1/2}(\hat{\phi}_{qml} - \phi_0)^\top \lambda_i(\phi_0)\| + o_p(1) \\ &\le \alpha_L^{-1} M \max_{1 \le i \le n} \|\lambda_i(\phi_0)\| + o_p(1).\end{aligned}$$

Since $\max_{1\le i\le n} n^{-1/2}\|\lambda_i(\phi_0)\| = o_p(1)$, then it follows that $\max_{1\le i\le n} \mid \mathcal{E}_{2i} \mid = o_p(1)$. Therefore, with the aid of the Ergodic Theorem, we obtain that

$$\left| n^{-1}\sum_{i=1}^{n} Z_i\mathcal{E}_{2i}\right| \le \max_{1\le i\le n} \mid \mathcal{E}_{2i} \mid n^{-1}\sum_{i=1}^{n} Z_i = o_p(1).$$

Since δ is arbitrary and $P(\hat{\phi}_{qml}\in B)\to 1$, in view of the above results and (A.8),

$$\tilde{\mu}_n = n^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}_i^{(qml)} = n^{-1}\sum_{i=1}^{n}\varepsilon_i + o_p(1),$$

and hence $\tilde{\mu}_n \to_p 1$ as $n\to\infty$; recall that $n^{-1}\sum_{i=1}^{n}\varepsilon_i \overset{a.s.}{\to} \mathbb{E}(\varepsilon_0) = 1$ by the strong law of large numbers. Therefore, in view of (A.7), for some constant $K>0$,

$$\{d_2(\widetilde{F}_{n,qml}, G_n)\}^2 \le Kn^{-1}\sum_{i=1}^{n}\{\varepsilon_i - \tilde{\varepsilon}_i^{(qml)}\}^2 + M_n \tag{A.9}$$

where M_n is a sequence of random variables that converges to zero in probability.

Since $Z_i\mathcal{E}_{1i} \overset{e.a.s.}{\longrightarrow} 0$, $\max_{1\le i\le n} \mid \mathcal{E}_{2i} \mid = o_p(1)$, and $\{Z_i; i\in\mathbb{Z}\}$ is strictly stationary and ergodic with $\mathbb{E}(Z_0^2)<\infty$, in view of (A.8), we obtain

$$n^{-1}\sum_{i=1}^{n}\{\varepsilon_i - \tilde{\varepsilon}_i^{(qml)}\}^2 = n^{-1}\sum_{i=1}^{n}\{Z_i\mathcal{E}_{1i} + Z_i\mathcal{E}_{2i}\}^2 = o_p(1).$$

This result together with (A.9) yield that $d_2(\widetilde{F}_{n,qml}, G_n) = o_p(1)$. □

Next, we provide the proof of Theorem 2.

Proof of Theorem 2. Let $G_{Z,k}(x)$ be the distribution function of Z_{n+k}, conditional on $\{Z_1,\ldots,Z_n\}$. Then, from the definition of the Mallows metric and applying the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
[d_2(G_{\check{Z}^{*qml},k}, G_{Z,k})]^2 &\le \inf_{\varepsilon^*\sim\widetilde{F}_{n,qml},\ \varepsilon\sim F} \mathbb{E}[\mathbb{E}^*\{\check{\Psi}_{n+k}^{*qml}(\hat{\phi}_{qml}^*)\varepsilon^* - \Psi_{n+k}(\phi_0)\varepsilon\}^2] \\
&\le \mathbb{E}[\mathbb{E}^*\{\check{\Psi}_{n+k}^{*qml}(\hat{\phi}_{qml}^*) - \Psi_{n+k}(\phi_0)\}^2] + \mathbb{E}\{\Psi_{n+k}^2(\phi_0)\} \inf_{\varepsilon^*\sim\widetilde{F}_{n,qml},\ \varepsilon\sim F_0} \mathbb{E}[\mathbb{E}^*(\varepsilon^*-\varepsilon)^2] \\
&\le \mathbb{E}[\mathbb{E}^*\{\check{\Psi}_{n+k}^{*qml}(\hat{\phi}_{qml}^*) - \Psi_{n+k}(\phi_0)\}^2] + \mathbb{E}\{\Psi_{n+k}^2(\phi_0)\}[d_2(\widetilde{F}_{n,qml}, F_0)]^2.
\end{aligned}$$

Lemma 6 yields $d_2(\widetilde{F}_{n,qml}, F_0) \to_p 0$, and hence $\mathbb{E}\{\Psi_{n+k}^2(\phi_0)\}[d_2(\widetilde{F}_{n,qml}, F_0)]^2 = o_p(1)$. Since an analogue of (13) holds for the current setup, by using the analogues of C.1, C.2 and C.3 we also obtain that $n^{1/2}(\hat{\phi}_{qml}^* - \phi_0) = O_{p_n^*}(1)$. The rest of the proof follows from Lemmas 3, 4 and 5, and a repetition of the arguments used in the proof of Theorem 1. □

A.3 Proofs of the main results of Section 2.4

We relegate the proofs of Theorems 3 and 4 and Corollaries 1 and 2 stated in Section 2.4 to Appendix S.3 in the supplementary material.

Acknowledgement

The authors would like to thank two anonymous referees, the associate editor, and the Co-editor Professor Jianqing Fan for their comments and suggestions that led to significant improvements. The authors also thank Professor Gael Martin for providing them with the dataset used in the empirical example. Mervyn Silvapulle gratefully acknowledges support from the Australian Research Council Discovery Grant scheme (Grant number DP150100210).

References

Allen, D., Chan, F., McAleer, M., and Peiris, S. (2008). Finite sample properties of the QMLE for the Log-ACD model: application to Australian stocks. *Journal of Econometrics*, **147**(1), 163–185.

Andrews, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica*, **65**(5), 1097–1128.

Bauwens, L. and Giot, P. (2000). The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique*, **60**, 117–149.

Bauwens, L. and Giot, P. (2001). *Econometric Modeling of Stock Market Intraday Activity*. Kluwer Academic Publishers, Boston, Dordrecht, London. Advanced Studies in Theoretical and Applied Econometrics, vol. 38.

Bauwens, L., Grammig, J., Veredas, D., and Giot, P. (2004). A comparison of financial duration models via density forecasts. *International Journal of Forecasting*, **20**(4), 589–609.

Berkes, I. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics*, **32**(2), 633–655.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**(6), 1196–1217.

Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.

Brownlees, C. T. and Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: data handling concerns. *Computational Statistics & Data Analysis*, **51**(4), 2232–2245.

Brownlees, C. T., Cipollini, F., and Gallo, G. M. (2012). Multiplicative error models. In L. Bauwens, C. Hafner, and S. Laurent, editors, *Handbook of Volatility Models and Their Applications*, chapter 9, pages 225–247. John Wiley and Sons, Inc, Hoboken, New Jersey.

Chan, K.-S., Li, D., Ling, S., and Tong, H. (2014). On conditionally heteroscedastic AR models with thresholds. *Statistica Sinica*, **24**(2), 625–652.

Chen, B., Gel, Y. R., Balakrishna, N., and Abraham, B. (2011). Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes. *Journal of Forecasting*, **30**(1), 51–71.

Chen, S. X. (2000). Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, **52**(3), 471–480.

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking*, **37**(3), 561–82.

Christensen, T., Hurn, A., and Lindsay, K. (2012). Forecasting spikes in electricity prices. *International Journal of Forecasting*, **28**(2), 400–411.

Christoffersen, P. and Goncalves, S. (2005). Estimation risk in financial risk management. *Journal of Risk*, **7**(3), 1–28.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, **27**(1-3), 46–78.

D'Agostino, R. B. and Stephens, M. A., editors (1986). *Goodness-of-fit techniques*, volume 68 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York.

Drost, F. C. and Werker, B. J. M. (2004). Semiparametric duration models. *Journal of Business & Economic Statistics*, **22**(1), 40–50.

Dufour, A. and Engle, R. F. (2000). The ACD model: Predictability of the time between consecutive trades. ICMA centre discussion papers in finance, Henley Business School, Reading University.

Engle, R. and Russell, J. R. (1997). Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of Empirical Finance*, **4**(2-3), 187–212.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, **17**(5), 425–446.

Engle, R. F. and 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. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, **66**(5), 1127–1162.

Fernandes, M. and Grammig, J. (2005). Nonparametric specification tests for conditional duration models. *Journal of Econometrics*, **127**(1), 35–68.

Fernandes, M. and Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics*, **130**(1), 1–23.

Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, **10**(4), 605–637.

Francq, C. and Zakoïan, J.-M. (2010). *GARCH models: structure, statistical inference and financial applications*. Wiley. John Wiley & Sons Ltd.

Gao, J., Kim, N. H., and Saart, P. W. (2015). A misspecification test for multiplicative error models of non-negative time series processes. *Journal of Econometrics*, **189**(2), 346–359. Frontiers in Time Series and Financial Econometrics.

Giacomini, R., Politis, D. N., and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, **29**(3), 567–589.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**(477), 359–378.

Grammig, J. and Maurer, K.-O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal*, **3**(1), 16–38.

Grammig, J. and Wellner, M. (2002). Modeling the interdependence of volatility and inter-transaction duration processes. *Journal of Econometrics*, **106**(2), 369–400.

Guo, B. and Li, S. (2018). Diagnostic checking of markov multiplicative error models. *Economics Letters*, **170**, 139 – 142.

Hautsch, N. (2011). *Econometrics of Financial High-Frequency Data*. Springer.

Hautsch, N., Malec, P., and Schienle, M. (2014). Capturing the zero: A new class of zero-augmented distributions and multiplicative error processes. *Journal of Financial Econometrics*, **12**(1), 89–121.

Hidalgo, J. and Zaffaroni, P. (2007). A goodness-of-fit test for ARCH(∞) models. *Journal of Econometrics*, **141**(2), 835–875.

Horváth, L., Kokoszka, P., and Teyssière, G. (2004). Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. *Journal of Statistical Computation and Simulation*, **74**(7), 469–485.

Janssen, P., Swanepoel, J., and Veraverbeke, N. (2005). Bootstrapping modified goodness-of-fit statistics with estimated parameters. *Statistics & Probability Letters*, **71**(2), 111–121.

Koul, H. L., Perera, I., and Silvapulle, M. J. (2012). Lack-of-fit testing of the conditional mean function in a class of Markov multiplicative error models. *Econometric Theory*, **28**(6), 1283–1312.

Lee, S.-W. and Hansen, B. E. (1994). Asymptotic theory for the GARCH$(1, 1)$ quasi-maximum likelihood estimator. *Econometric Theory*, **10**(1), 29–52.

Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics*, **140**(2), 849–873.

Luca, G. D., Gallo, G. M., and Carità, D. (2017). Evaluating Combined Forecasts for Realized Volatility Using Asymmetric Loss Functions. *Econometric Research in Finance*, **2**(2), 99–111.

Lunde, A. (1999). A generalized gamma autoregressive conditional duration model. *Discussion paper, Aalborg University.*

Mancini, L. and Trojani, F. (2011). Robust value at risk prediction. *Journal of Financial Econometrics*, **9**(2), 281–313.

Maneesoonthorn, W., Martin, G. M., Forbes, C. S., and Grose, S. D. (2012). Probabilistic forecasts of volatility and its risk premia. *Journal of Econometrics*, **171**(2), 217–236. Bayesian Models, Methods and Applications.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, **8**(4), 377–399.

Mazzeu, J. H. G., González-Rivera, G., Ruiz, E., and Veiga, H. (2017). A Bootstrap Approach for Generalized Autocontour Testing. Implications for VIX Forecast Densities. Working Papers 201709, University of California at Riverside, Department of Economics.

Meintanis, S. G., Milošević, B., and Obradović, M. (2017). Goodness-of-fit tests in conditional duration models. *Statistical Papers*.

Pacurar, M. (2008). Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys*, **22**(4), 711–751.

Pascual, L., Romo, J., and Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in {GARCH} models. *Computational Statistics & Data Analysis*, **50**(9), 2293–2312.

Perera, I. and Koul, H. L. (2017). Fitting a two phase threshold multiplicative error model. *Journal of Econometrics*, **197**(2), 348–367.

Perera, I. and Silvapulle, M. J. (2017). Specification tests for multiplicative error models. *Econometric Theory*, **33**(2), 413–438.

Perera, I., Hidalgo, J., and Silvapulle, M. J. (2016). A goodness-of-fit test for a class of autoregressive conditional duration models. *Econometric Reviews*, **35**(6), 1111–1141.

Pfanzagl, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika*, **14**(1), 249–272.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics*, **34**(5), 2449–2495.

Zhang, M. Y., Russell, J. R., and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, **104**(1), 179–207.