Econometric Theory, **33**, 2017, 413–438.
doi:10.1017/S026646661500047X

SPECIFICATION TESTS FOR MULTIPLICATIVE ERROR MODELS

INDEEWARA PERERA AND MERVYN J. SILVAPULLE
Monash University

The family of multiplicative error models is important for studying non-negative variables such as realized volatility, trading volume, and duration between consecutive financial transactions. Methods are developed for testing the parametric specification of a multiplicative error model, which consists of separate parametric models for the conditional mean and the error distribution. The same method can also be used for testing the specification of the error distribution provided the conditional mean is correctly specified. A bootstrap method is proposed for computing the p-values of the tests and is shown to be consistent. The proposed tests have nontrivial asymptotic power against a class of $O(n^{-1/2})$-local alternatives. The tests performed well in a simulation study, and they are illustrated using a data example on realized volatility.

1. INTRODUCTION AND MOTIVATION

Statistical models for nonnegative random variables have been used in many areas, including finance, economics, health sciences, and engineering. In finance, the family of multiplicative error models for nonnegative variables plays a central role (Pacurar, 2008). They have been used for modelling the duration between financial transactions (Engle and Russell, 1998), trading volume of orders (Manganelli, 2005), high–low range of asset prices (Chou, 2005), and realized volatility (Engle and Gallo, 2006; Brownlees, Cipollini, and Gallo, 2012). Our main objective is to develop specification tests for multiplicative error models.

Let Z_i denote a nonnegative random variable, for example, realized volatility, at time i $(i = \ldots, -1, 0, 1 \ldots)$. A *multiplicative error model* [MEM] takes the form,

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

where $\Psi_i = \mathbb{E}(Z_i \mid \mathcal{H}_{i-1})$ is the mean of Z_i conditional on the past information $\mathcal{H}_{i-1}$, and $\{\ldots, \varepsilon_0, \varepsilon_1, \ldots\}$ are independent and identically distributed [iid]

We would like to thank three anonymous referees, the Co-editor Professor Giuseppe Cavaliere, and the editor Professor Peter Phillips for their comments and suggestions. We also thank Christian Brownlees for providing us with the dataset used in the empirical example, the participants of the seminar programs at the University of Warwick and the University of North Carolina at Chapel Hill, and the participants of the Second Econometric Society Meeting held in Xiamen, for their helpful suggestions. The authors were partially supported by Australian Research Council Discovery Grants. We thank Monash e-Science and Grid Engineering Lab for assistance with some computational aspects and use of the computing facilities. Please address correspondence to: Indeewara Perera/Mervyn Silvapulle, Department of Econometrics and Business Statistics, Monash Business School, Monash University, P.O. Box 197, Caulfield East, Australia 3145; e-mail: indeewara.perera@monash.edu; mervyn.silvapulle@monash.edu.

with mean 1 and common distribution function F^0. We develop new and easy-to-use methods for testing the specification of a given parametric form of the entire conditional distribution of Z_i, thus simultaneously testing the parametric forms of both, the conditional mean Ψ_i and the distribution function F^0 of the error term.

Several methods have been proposed in the literature for testing a given parametric specification of Ψ_i in MEM (e.g. Meitz and Teräsvirta, 2006; Chen and Hsieh, 2010). However, relatively less attention has been given to specification testing of the entire conditional distribution of Z_i. Such tests are important whenever the statistical inference involves the entire conditional distribution of Z_i rather than only its conditional mean. An example of this situation is the forecasting of the conditional density of Z_i by using an assumed parametric model. The methods developed in the paper are suitable for testing the goodness-of-fit of such parametric models. Other contexts where these tests would be useful include the following: (a) option pricing and risk management procedures that use intraday volatility estimated from price duration models (Giot, 2000; Fernandes and Grammig, 2005), (b) inference on the conditional quantiles of Z_i, also known as the *value at risk* in finance, and (c) modelling the link between duration and volatility (Ghysels, Gouriéroux, and Jasiak, 2004; Engle, 2000).

Fernandes and Grammig (2005) developed a specification test for F^0 under the assumption that Ψ_i has a known finite dimensional parametric form. An attractive feature of their test is that it is asymptotically distribution free. However, since the test is based on a nonparametric kernel density estimator, it has asymptotic power against local alternatives that converge slower than the regular parametric rate $n^{-1/2}$ but not against those that converge at the rate $n^{-1/2}$.

Corradi and Swanson (2006) developed a method for testing a hypothesis that the distribution of Z_i conditional on X_i follows a given parametric form, under the assumption that $\{(Z_i, X_i)\}_{i=-\infty}^{\infty}$ is an *observable* stochastic process. Their method is not applicable to the type of MEM in this paper because $\{X_i\}_{i=-\infty}^{\infty}$ may be *unobservable*. For example, for the model $\Psi_i = \phi_1 + \phi_2 Z_{i-1} + \phi_3 \Psi_{i-1}$, the variable X_i includes the unobservable conditional expectation Ψ_{i-1}. Other references related to the testing problem include evaluating density-forecasts of Z_i (Diebold, Gunther, and Tay, 1998; Bauwens, Grammig, Veredas, and Giot, 2004).

We propose a method based on the classical *Kolmogorov–Smirnov* and *Cramer–von Mises* type tests, except that our test statistics contain estimated parameters. Therefore, the limiting null distributions of the test statistics depend on unknown nuisance parameters, and hence, the critical values cannot be tabulated for general use. Therefore, we propose a bootstrap method to implement the tests. Establishing the validity of the bootstrap method forms the major technical content of the paper. It is shown that the tests have nontrivial asymptotic power against a sequence of local alternatives that converge at the regular parametric rate $O(n^{-1/2})$. In a comparative simulation study, the proposed tests performed better than their competitors do. Our tests can also be used for testing the specification of the error distribution alone, as in Fernandes and Grammig (2005) and Chen and Hsieh (2010).

The rest of the paper is structured as follows. Section 2 formulates the problem, defines the test statistics, and provides several asymptotic results. Section 3 develops the bootstrap implementation of the tests, and Section 4 provides results on the local power of the tests. Section 5 contains the results of the simulation study to evaluate the finite sample performance of the tests. A real-data example involving realized volatility is provided in Section 6. Section 7 concludes the paper. The Appendix provides the regularity conditions and some of the proofs. For the remaining proofs and some additional simulation results, readers may refer to the supplementary material associated with this article, available at Cambridge Journals Online (journals.cambridge.org); we refer to this simply as the *online Supplementary Material*.

2. THE TEST STATISTICS AND THEIR ASYMPTOTIC NULL DISTRIBUTIONS

Let the multiplicative error model [MEM] in (1) together with the definitions of $Z_i, \varepsilon_i, \mathcal{H}_i$, and F^0 be as in the previous section. Let $\mathcal{F} = \{F_\theta : \theta \in \Theta \subset \mathbb{R}^q\}$ be a given parametric family of distributions and let f_θ denote the probability density function [pdf] corresponding to F_θ, where q is a positive integer and each F_θ has mean 1 and finite variance. Further, let $\{\Psi_i(\phi) : \phi \in \Phi \subset \mathbb{R}^p\}$ be a given parametric family, where $\Psi_i(\cdot)$ has a known functional form that may depend on the past values $\{Z_t : t \leq i-1\}$, and p is a known positive integer. The notation $\Psi_i(\phi)$, as opposed to Ψ_i without the argument '$(\cdot)$', implicitly refers to a parametric form of the conditional mean, $\Psi_i = \mathbb{E}(Z_i \mid \mathcal{H}_{i-1})$. We assume that the multiplicative error term ε_i is independent of $\{(\Psi_t(\phi), Z_{t-1}) : t \leq i\}$, and that $\Psi_i(\phi)$ is of the form

$$\Psi_i(\phi) = \mathcal{J}\{Z_{i-1}, \ldots, Z_{i-p_1}, \Psi_{i-1}(\phi), \ldots, \Psi_{i-p_2}(\phi); \phi\}, \tag{2}$$

where $\mathcal{J}(\cdot;\phi)$ is a known and twice continuously differentiable function, and p_1 and p_2 are known positive integers. The general form in (2) includes a large class of conditional mean specifications of MEMs studied in the literature. For example, the linear multiplicative error model of Engle and Russell (1998), which we denote by 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 $\phi = (\alpha, \beta_1, \ldots, \beta_{p_1}, \gamma_1, \ldots, \gamma_{p_2})^\top$ with '$^\top$' denoting the transpose; it resembles the well-known GARCH(p_1, p_2).

Let the null and alternative hypotheses be defined by

$$H_0 : (\Psi_i, F^0) = \left(\Psi_i(\phi_0), F_{\theta_0}\right) \text{ for some } (\phi_0, \theta_0) \in \Phi \times \Theta, \text{ and } H_1 : \text{ Not } H_0. \tag{3}$$

To develop our tests of (3), we need to first estimate the model under H_0 and then construct a set of residuals. However, these computations encounter a hurdle

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

because it may not be possible to compute $\Psi_i(\phi)$ by using only the observed sample $\{Z_1,\ldots,Z_n\}$ and the value of ϕ. For example, for the MEM(1,1) model $\Psi_i(\phi)=\phi_1+\phi_2 Z_{i-1}+\phi_3\Psi_{i-1}(\phi)$, we have $\Psi_i(\phi)=\alpha(1-\gamma)^{-1}+\beta\sum_{j=1}^{\infty}\gamma^{j-1}Z_{i-j}$, which depends on the unobserved part of the process $\{\ldots,Z_{i-2},Z_{i-1}\}$ extending back to the infinite past. Therefore, we introduce some suitable starting values for the unobservable values $\{(Z_i,\Psi_i(\phi)), i\le 0\}$ and approximate $\Psi_i(\phi)$ by $\widetilde{\Psi}_i(\phi)$ defined as follows: $\widetilde{\Psi}_i(\phi)=1$ and $\widetilde{Z}_i=\bar{Z}:=[Z_1+\cdots+Z_n]/n$ for $i<1$, $\widetilde{Z}_i=Z_i$ for $i\ge 1$, and $\widetilde{\Psi}_i(\phi)=\mathcal{J}\{\widetilde{Z}_{i-1},\ldots,\widetilde{Z}_{i-p_1},\widetilde{\Psi}_{i-1}(\phi),\ldots,\widetilde{\Psi}_{i-p_2}(\phi);\phi\}$ for $i\ge 1$. The main theorems and propositions presented in Sections 3 and 4 show that the effect of the starting values on the distributions of the test statistics becomes negligible in large samples.

Let $\hat{\phi}$ denote the quasimaximum likelihood estimator [QMLE] of ϕ_0 based on the standard exponential distribution, which is defined by

$$\hat{\phi}=\arg\min_{\phi\in\Phi}\sum_{i=1}^{n}\ell_i(\phi),\quad \ell_i(\phi)=\log\widetilde{\Psi}_i(\phi)+Z_i/\widetilde{\Psi}_i(\phi). \tag{4}$$

Let the corresponding residuals $\{\widetilde{\varepsilon}_i, 1\le i\le n\}$ be defined by $\widetilde{\varepsilon}_i=Z_i/\widetilde{\Psi}_i(\hat{\phi})$, $i=1,\ldots,n$. We estimate θ_0 by

$$\hat{\theta}=\arg\max_{\theta\in\Theta}\sum_{i=1}^{n}g_\theta(\widetilde{\varepsilon}_i), \tag{5}$$

where $g_\theta(\cdot)$ is a suitably chosen function (see Assumption (E2) in Appendix A). For example, $g_\theta(\cdot)$ may be $\log f_\theta(\cdot)$, where f_θ is the pdf of F_θ ($\theta\in\Theta$). Let $I(\cdot)$ denote the indicator function and let $\widetilde{F}_n(x)=n^{-1}\sum_{i=1}^{n}I(\widetilde{\varepsilon}_i\le x)$, $x\ge 0$. Let the residual empirical process $\widetilde{W}_n$ estimated under the null hypothesis be defined by

$$\widetilde{W}_n(x)=\sqrt{n}\left\{\widetilde{F}_n(x)-F_{\hat{\theta}}(x)\right\},\qquad x\ge 0. \tag{6}$$

Next, we provide a heuristic argument to indicate that a Kolmogorov–Smirnov type test of H_0 against H_1 in (3) can be constructed based on $\widetilde{W}_n(\cdot)$.

If H_0 is true, then $F_{\hat{\theta}}$ and $\widetilde{F}_n$ are expected to be close to the unknown distribution F^0, and hence, $\widetilde{W}_n(\cdot)$ is likely to be close to zero. Now, suppose that H_1 is true. Then, either the conditional mean function $\Psi_i(\phi)$ or the parametric form F_θ for the error distribution is misspecified. If F_θ is misspecified and $\Psi_i(\phi)$ is correctly specified, then $\widetilde{F}_n$ is expected to be close to F^0 which is not close to $F_{\hat{\theta}}$. Hence, $\sup_x|\widetilde{W}_n(x)|$ is likely to be large. Alternatively, suppose that $\Psi_i(\phi)$ is misspecified. Then, the estimator $\{\Psi_i(\hat{\phi})\}$ is not expected to be close to the true conditional mean $\{\Psi_i\}$. Hence, $\widetilde{F}_n$ and $F_{\hat{\theta}}$ are likely to be close to different distribution functions, and $\sup_x|\widetilde{W}_n(x)|$ is expected to be large. These heuristic arguments provide a sufficient basis for constructing test statistics that resemble Kolomogorov–Smirnov type statistics based on $\widetilde{W}_n$.

Let $D^{+} = \sup_{x\geq 0} \widetilde{W}_n(x)$ and $D^{-} = -\inf_{x\geq 0} \widetilde{W}_n(x)$, and define the test statistics

$$
\begin{aligned}
&T_1 \equiv KS = \sup_{x\geq 0} |\widetilde{W}_n(x)| = \max\{D^{+}, D^{-}\} && \text{[Kolmogorov–Smirnov]},\\
&T_2 \equiv Ku = D^{+} + D^{-} && \text{[Kuiper]},\\
&T_3 \equiv CvM = \int \widetilde{W}_n^2(x) dF_{\hat{\theta}}(x) && \text{[Cramér-von Mises]},\\
&T_4 \equiv A^2 = \int \widetilde{W}_n^2(x)[F_{\hat{\theta}}(x)\{1 - F_{\hat{\theta}}(x)\}]^{-1} dF_{\hat{\theta}}(x) && \text{[Anderson–Darling]},\\
&T_5 \equiv U^2 = \int \{\widetilde{W}_n(x) - \int [\widetilde{W}_n(x)] dF_{\hat{\theta}}(x)\}^2 dF_{\hat{\theta}}(x) && \text{[Watson]}.
\end{aligned}
$$

Let us introduce some additional notation. For a differentiable function $m(s,x)$ on $\Phi \times \mathbb{R}$ or on $\Theta \times \mathbb{R}$, the derivatives with respect to s and x are denoted by $\dot{m}(s,x)$ and $m'(s,x)$, respectively. For example, $f'_\theta(y) = \partial f_\theta(y)/\partial y$ and $\dot{\Psi}_i(\phi) = \left[(\partial/\partial\phi_1)\Psi_i(\phi), ..., (\partial/\partial\phi_p)\Psi_i(\phi)\right]^\top$. Let

$$\lambda_i(\phi) = \dot{\Psi}_i(\phi)/\Psi_i(\phi), \quad L_n(\phi) = \sum_{i=1}^{n} \ell_i(\phi),$$

$$h_\theta(t) = \{-\mathbb{E}[(\partial/\partial\theta)\dot{g}_\theta(\varepsilon_1)]\}^{-1} \dot{g}_\theta(t),$$

where $\ell_i(\cdot)$ and $g_\theta(\cdot)$ are as in (4) and (5), respectively. Let $D[0,1]$ denote the space of *càdlàg* functions on $[0,1]$ equipped with the uniform metric, and let $f_1 \circ f_2$ denote the composition of the functions f_1 and f_2 defined by $f_1 \circ f_2(x) = f_1\{f_2(x)\}$. Let $F_n(x) = n^{-1}\sum_{i=1}^{n} I(\varepsilon_i \leq x)$ denote the empirical distribution function of the unobserved errors $\{\varepsilon_1, \ldots, \varepsilon_n\}$, and let $W_n(x) = \sqrt{n}\{F_n(x) - F_{\theta_0}(x)\}$ be the corresponding empirical process under the null hypothesis.

Assume that Conditions (C1)−(C5) and Assumptions (E1) and (E2), stated in Appendix A, are satisfied and that the null hypothesis H_0 holds. An important result established under these assumptions in Appendix A.3, Lemma A.12, is that:

$$\sup_{x\geq 0} \left| \widetilde{F}_n(x) - F_n(x) - (\hat{\phi} - \phi_0)^\top \mathbb{E}\{\lambda_1(\phi_0)\} x f_{\theta_0}(x) \right| = o_p(n^{-1/2}). \tag{7}$$

Hence, in first-order asymptotic arguments, the estimated empirical process $\widetilde{W}_n(x)$ may be replaced by the more tractable process $W_n(x) + \sqrt{n}(\hat{\phi} - \phi_0)^\top \mathbb{E}\{\lambda_1(\phi_0)\} x f_{\theta_0}(x) + \sqrt{n}(\hat{\theta} - \theta_0)^\top \dot{F}_{\theta_0}(x)$. Our first main result is the following:

THEOREM 1. *Suppose that Conditions (C1)−(C5) and Assumptions (E1) and (E2) are satisfied and that the null hypothesis H_0 holds. Then, (a) the estimated empirical process $\widetilde{W}_n \circ F_{\theta_0}^{-1}(\cdot)$ converges weakly in $D[0,1]$ to a centered Gaussian process $G(\cdot)$, (b) there exists a continuous functional $\mathfrak{h}_j$ on $D[0,1]$ such that $T_j = \mathfrak{h}_j(\widetilde{W}_n \circ F_{\theta_0}^{-1}) + o_p(1)$, and (c) $T_j \xrightarrow{d} \mathfrak{h}_j(G)$ as $n \to \infty$ $(j = 1, \ldots, 5)$.*

Based on the preceding theorem, we propose the following asymptotic test: reject H_0 if $T_j > c_{j\alpha}$, where $c_{j\alpha}$ is the $(1-\alpha)$th quantile of $\mathfrak{h}_j(G)$, $j = 1, \ldots, 5$.

Since the distribution of $\mathfrak{h}_j(G)$ depends on the unknown nuisance parameter (θ_0, ϕ_0), the asymptotic critical values of T_j cannot be computed for general use $(j = 1, \ldots, 5)$. It does not seem possible to adapt a martingale transformation method as in Koul, Perera, and Silvapulle (2012) or a method based on an analytical approximation to compute critical values. Therefore, a bootstrap method is proposed in the next section.

3. BOOTSTRAP TESTS

In this section, we propose a bootstrap algorithm to implement the tests proposed in the previous section. Here, in order to highlight the fact that the bootstrap samples are generated from a process that starts at time $-m$, the superscript '$*(m)$' is used instead of the more familiar symbol '$*$'; in Appendix A, we use the superscript * for $^{*(\infty)}$. For $j = 1, \ldots, 5$, the steps of the bootstrap algorithm are as follows.

Step 1: Compute the estimates $\hat{\phi}$ and $\hat{\theta}$ based on the observed sample $\{Z_1, \ldots, Z_n\}$ and obtain the residuals $\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_n$, where $\widetilde{\varepsilon}_i = Z_i / \widetilde{\Psi}_i(\hat{\phi})$, $i = 1, \cdots, n$.
Step 2: Compute the test statistic T_j.
Step 3: Generate $m + n + 1$ independent observations, $\varepsilon^*_{-m}, \ldots, \varepsilon^*_n$ from $F_{\hat{\theta}}$.
Step 4: Assign the starting values $\Psi_i^{*(m)}(\phi) = 1$ and $Z_i^{*(m)} = (Z_1 + \cdots + Z_n)/n$ for $i < -m$, and compute $Z_{-m}^{*(m)}, \ldots, Z_n^{*(m)}$ recursively by using the model equation (other suitable starting values may also be used). Now, discard the first $m+1$ values $\{Z_{-m}^{*(m)}, \ldots, Z_0^{*(m)}\}$, and use $\{Z_1^{*(m)}, \ldots, Z_n^{*(m)}\}$ as the bootstrap sample.
Step 5: Repeat step 1 for the bootstrap sample $\{Z_1^{*(m)}, \ldots, Z_n^{*(m)}\}$, and compute the bootstrap counterparts of $\hat{\phi}$, $\{\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_n\}$, $\hat{\theta}$, $\widetilde{F}_n$ and $\widetilde{W}_n$, which we denote by $\hat{\phi}^{*(m)}$, $\{\widetilde{\varepsilon}_1^{*(m)}, \ldots, \widetilde{\varepsilon}_n^{*(m)}\}$, $\hat{\theta}^{*(m)}$, $\widetilde{F}_n^{*(m)}$, and $\widetilde{W}_n^{*(m)}$, respectively. Then, compute $T_j^{*(m)}$, the bootstrap analogue of T_j.
Step 6: Repeat steps 3–5 a sufficiently large number of times and compute $c_{j\alpha}^{*(m)}$, the $(1-\alpha)$th quantile of the sampled values of $T_j^{*(m)}$. Now, the *bootstrap test* based on T_j is the following: Reject H_0 at significance level α if $T_j > c_{j\alpha}^{*(m)}$.

Let P_n^* denote the bootstrap probability conditional on $\{Z_1, \ldots, Z_n\}$. In what follows, convergence results relating to bootstrapped processes such as $\widetilde{W}_n^{*(m)} \circ F_{\hat{\theta}}^{-1}$ are in probability, and they are valid irrespective of whether H_0 is true. Let '$\xrightarrow{d^*}$' denote the convergence in distribution of bootstrap statistics. For example, the statement '$T_j^{*(m)} \xrightarrow{d^*} \mathfrak{h}_j(\mathcal{V})$' means that $P_n^*(T_j^{*(m)} \le z) \longrightarrow P\{\mathfrak{h}_j(\mathcal{V}) \le z\}$, in probability, at every continuity point z of $P\{\mathfrak{h}_j(\mathcal{V}) \le z\}$. Similarly, we define the bootstrap orders $o_{p_n^*}(1)$ and $O_{p_n^*}(1)$ as follows: (a) $X_n^* = o_{p_n^*}(1)$ if $P_n^*\{|X_n^*| > \delta\} \xrightarrow{p} 0$ 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 $n \to \infty$. Now, we have

THEOREM 2. *Suppose that Conditions (C1)–(C6) and Assumptions (E1)–(E3) are satisfied. Then, conditional on* $\{Z_1, \ldots, Z_n\}$, *(a) the process*

$\widetilde{W}_n^{*(m)} \circ F_{\hat{\theta}}^{-1}(\cdot)$ *converges weakly [in probability] in* $D[0,1]$ *to a centered Gaussian process* $\mathcal{V}(\cdot)$ *with cov*$\{\mathcal{V}(s), \mathcal{V}(t)\} := \min\{s,t\} - st + \mathcal{G}(s,t,\theta_0,\phi_0)$, *where* $\mathcal{G}(s,t,\theta,\phi)$ *is given by (A.9) in Appendix A, and (b)* $T_j^{*(m)} \xrightarrow{d^*} \mathfrak{h}_j(\mathcal{V})$ *as* $n \to \infty$ $(j = 1,\ldots,5)$. *If, in addition,* H_0 *is also true then the Gaussian process* G *in Theorem 1 and* $\mathcal{V}$ *have the same law.*

In view of Theorem 2, under H_0, the distribution of $\mathfrak{h}_j(\mathcal{V})$ is the same as that of $\mathfrak{h}_j(G)$, the asymptotic null distribution of T_j. Therefore, the bootstrap test based on T_j is a valid level α asymptotic test $(j = 1,\ldots,5)$. Next, suppose that H_1 is true. Then, as indicated in the previous section, $T_j \xrightarrow{p} \infty$, except in some pathological situations $(j = 1,\ldots,5)$. However, the quantiles of $\mathfrak{h}_j(\mathcal{V})$ are finite because $\mathcal{V}$ is a centered Gaussian process and $\mathfrak{h}_j$ is continuous. Therefore, the critical values computed by the bootstrap method are finite. Consequently, the bootstrap test based on T_j has asymptotic power 1 against fixed alternatives $(j = 1,\ldots,5)$.

The proof of Theorem 2 requires several intermediate results of independent interest. One that is worthy of special mention is an asymptotic result that is crucial for establishing the weak convergence of $\widetilde{W}_n^{*(m)} \circ F_{\hat{\theta}}^{-1}(\cdot)$. To indicate this, let $\hat{\varepsilon}_i^{*(m)} = Z_i^{*(m)} / \Psi_i^{*(m)}(\hat{\phi}^{*(m)})$, $\hat{F}_n^{*(m)}(x) = n^{-1}\sum_{i=1}^n I(\hat{\varepsilon}_i^{*(m)} \le x)$, and $\widehat{W}_n^{*(m)}(x) = \sqrt{n}\{\hat{F}_n^{*(m)}(x) - F_{\hat{\theta}^{*(m)}}(x)\}$. The required result, Lemma A.13 in Appendix A, states that $\sup_{x\ge 0} n^{1/2}|\widetilde{F}_n^{*(m)}(x) - \hat{F}_n^{*(m)}(x)| = o_{p_n^*}(1)$. Consequently, $\widetilde{W}_n^{*(m)}(\cdot)$ can be replaced by the more tractable process $\widehat{W}_n^{*(m)}(\cdot)$ in first-order asymptotic arguments. The process $\{Z_i^{*(m)} : i \in \mathbb{N}\}$ is not stationary and has a complicated dependence pattern. Therefore, establishing the weak convergence of $\widehat{W}_n^{*(m)}(\cdot)$ is not a trivial task. The purpose of some of the technical lemmas in Appendix A is to develop the preliminary results needed for establishing the weak convergence of $\widehat{W}_n^{*(m)}(\cdot)$. For example, from Lemmas A.6, A.7, and A.8 we obtain that $\sqrt{n}\{\hat{\phi}^{*(m)} - \hat{\phi}\}$ and $\sqrt{n}\{\hat{\theta}^{*(m)} - \hat{\theta}\}$ have asymptotic expansions that resemble those for $\sqrt{n}\{\hat{\phi} - \phi_0\}$ and $\sqrt{n}\{\hat{\theta} - \theta_0\}$, respectively. We use these expansions in Lemmas A.9, A.10, and A.11 to develop the main technical results needed for deriving the weak convergence of $\widehat{W}_n^{*(m)}(\cdot)$.

4. ASYMPTOTIC LOCAL POWER

In this section, we study the asymptotic power of the tests against sequences of local alternatives. To illustrate the main idea, let us consider the Kolmogorov–Smirnov statistic $KS = \sup_{t\in[0,1]}|\widetilde{W}_n \circ F_{\theta_0}^{-1}(t)|$. Under a given sequence of local alternatives, we show that $\widetilde{W}_n \circ F_{\theta_0}^{-1}$ converges weakly in $D[0,1]$ to a process of the form $M + G$, where G is the zero mean Gaussian process in Theorem 1 and M is a nonstochastic term. Further, M is nonzero on a set of positive measure except in some pathological cases. Since $\widetilde{W}_n \circ F_{\theta_0}^{-1}$ converges weakly to G under H_0, M captures the shift away from zero that appears in the limiting law of $\widetilde{W}_n \circ F_{\theta_0}^{-1}$ under the sequence of local alternatives. Since G has a zero mean, we conclude that $\sup_{t\in[0,1]}|M(t)+G(t)|$ is stochastically larger than $\sup_{t\in[0,1]}|G(t)|$. Therefore, it follows that the KS test has nontrivial asymptotic local power.

We present the results for the following three types of local alternatives: (a) departure in the error distribution only, (b) departure in the conditional mean function only, and (c) departure in both. We consider them in turn.

Recall that, under H_0, (ϕ_0, θ_0) denotes the true value. Let $F_{(n)} = (1 - n^{-1/2}\delta)F_{\theta_0} + n^{-1/2}\delta\tilde{F}$, where $0 < \delta < 1$, $\tilde{F}$ is a distribution function satisfying the conditions imposed on the error distribution such that $[\int\{\ddot{g}_{\theta_0}(y)\}^{-1}\, dF_{\theta_0}(y)\int \dot{g}_{\theta_0}(\varepsilon)\, d\tilde{F}(\varepsilon)]^\top \dot{F}_{\theta_0} \neq [F_{\theta_0} - \tilde{F}]$ and $\tilde{F} \neq F^0$. Consider the sequence of local alternatives,

$$H_{an} : \Psi_i = \Psi_i(\phi_0) \text{ and } F^0 = F_{(n)}. \quad \textbf{(8)}$$

Thus, H_{an} converges to H_0 at the rate $O(n^{-1/2})$ and only the error distribution is misspecified under H_{an}.

PROPOSITION 1. *Suppose that Conditions (C1)–(C5) and Assumptions (E1) and (E2) are satisfied. Then, under H_{an}, we have the following for $j = 1,\ldots,5$:*

(i) $T_j \xrightarrow{d} \mathfrak{h}_j(W_a)$ *as $n \to \infty$, where $W_a(\cdot) = m_a(\cdot) + G(\cdot)$ and $m_a(\cdot)$ is a nonrandom function. An expression for $m_a(\cdot)$ is given in (A.15).*

(ii) *If (C6) and (E3) are also satisfied, then $T_j^{*(m)} \xrightarrow{d^*} \mathfrak{h}_j(G)$ as $n \to \infty$.*

Next, consider a sequence of local alternatives for departures from the conditional mean specification. To this end, let $r_i = r(Z_{i-1}, Z_{i-2}, \cdots)$ be a square integrable and twice continuously differentiable function. Assume that $\{r_i : i \in \mathbb{Z}\}$ forms a strictly stationary and ergodic process. Define a sequence of local hypothesis by

$$H_{bn} : \Psi_i = \Psi_i(\phi_0) + r_i/n^{1/2} \text{ and } F^0 = F_{\theta_0};$$

see Ling and Tong (2011) for similar local alternatives. For illustrative purposes, we obtain explicit expressions for the asymptotic local power for the special case when the error distribution is standard exponential. A general case is considered later in Proposition 3. Let $\mathcal{Q}_i(\phi_0, \varepsilon_i) = n^{-1}\sum_{j=1}^n h'_{\theta_0}(\varepsilon_j)\varepsilon_j\{\lambda_j(\phi_0)\}^\top \xi_i(\phi_0, \varepsilon_i)$, where $\xi_i(\phi_0, \varepsilon_i) = \tau(\phi_0)\lambda_i(\phi_0)(1 - \varepsilon_i)$. For $t \in [0,1]$, let

$$g_i(t) = a_i(t) - b_i(t) + c_i(t), \quad \textbf{(9)}$$

$$m_b(t) = \mathbb{E}[g_1(t)\{r_1/\Psi_1(\phi_0)\}(\varepsilon_1 - 1)], \quad \textbf{(10)}$$

where

$$a_i(t) = I\{\varepsilon_i \le F_{\theta_0}^{-1}(t)\} - t, \quad b_i(t) = \{h_{\theta_0}(\varepsilon_i) - \mathcal{Q}_i(\phi_0, \varepsilon_i)\}^\top \dot{F}_{\theta_0}\{F_{\theta_0}^{-1}(t)\},$$

$$c_i(t) = F_{\theta_0}^{-1}(t) f_{\theta_0}\{F_{\theta_0}^{-1}(t)\}\mathbb{E}[\lambda_1(\phi_0)]^\top \xi_i(\phi_0, \varepsilon_i),$$

and the expectation $\mathbb{E}$ is taken under the null hypothesis H_0.

PROPOSITION 2. *Suppose that Conditions (C1)–(C5) and Assumptions (E1) and (E2) are satisfied, and that F_{θ_0} is the standard exponential distribution. Then, under H_{bn}, we have the following for $j = 1, \ldots, 5$:*

(i) $T_j \xrightarrow{d} \mathfrak{h}_j(W_b)$ *as* $n \to \infty$, *where* $W_b(\cdot) = m_b(\cdot) + G(\cdot)$.

(ii) *If (C6) and (E3) are also satisfied, then* $T_j^{*(m)} \xrightarrow{d^*} \mathfrak{h}_j(G)$ *as* $n \to \infty$.

To obtain the local power against departures of both the conditional mean and the error distribution, let us consider the sequence of local alternatives

$$H_{cn} : Z_i = \left[\Psi_i(\phi_0) + r_i/n^{1/2}\right]\varepsilon_i, \quad \varepsilon_i \sim F_{(n)},$$

where $F_{(n)}$ is as in (8).

PROPOSITION 3. *Suppose that Conditions (C1)–(C5) and Assumptions (E1) and (E2) are satisfied, and* $m_c(t) := \lim_{n\to\infty} -n^{-1}\sum_{i<j\le n} \mathbb{E}[g_i(t)\{r_j/\Psi_j(\phi_0)\}]$ *exists for each* $t \in [0,1]$ *under* H_0. *Then, under* H_{cn}, *we have the following for* $j = 1, \ldots, 5$:
(i) $T_j \xrightarrow{d} \mathfrak{h}_j(W_c)$ *as* $n \to \infty$, *where* $W_c(\cdot) = m_c(\cdot) + m_a(\cdot) + G(\cdot)$, *with* m_a *as in Proposition 1. (ii) If (C6) and (E3) are also satisfied, then* $T_j^{*(m)} \xrightarrow{d^*} \mathfrak{h}_j(G)$ *as* $n \to \infty$.

It follows from Propositions 1, 2, and 3 that the proposed bootstrap tests have nontrivial asymptotic power against H_{an}, H_{bn}, and H_{cn}, respectively.

5. SIMULATION STUDY

We conducted a simulation study to compare the proposed tests with their competitors in terms of size, power, and relevance to density forecasting. Tests were evaluated when the parametric form $\Psi_i(\phi)$ under H_0 was MEM(1,1): $\Psi_i(\phi) = \phi_1 + \phi_2 Z_{i-1} + \phi_3 \Psi_{i-1}(\phi)$, $\phi_1 > 0, \phi_2, \phi_3 \ge 0, \phi_2 + \phi_3 < 1$. For the error distribution, we considered the following five families: Gamma, Weibull, Exponential, Generalized Gamma, and Burr. For the conditional mean Ψ_i of the true data generating process [DGP] under H_0 and under H_1, we considered the following five cases:

1. MEM(1,1): $\Psi_i = 0.20 + 0.10 Z_{i-1} + 0.70 \Psi_{i-1}$,
2. MEM(2,1): $\Psi_i = 0.10 + 0.20 Z_{i-1} + 0.10 Z_{i-2} + 0.60 \Psi_{i-1}$,
3. log-MEM(1,1): $\ln \Psi_i = -0.10 + 0.06 \ln Z_{i-1} + 0.90 \ln \Psi_{i-1}$,
4. Exp-MEM(1,1): $\ln \Psi_i = -0.10 + 0.15\varepsilon_{i-1} + 0.35 \mid \varepsilon_{i-1} - 1 \mid + 0.60 \ln \Psi_{i-1}$,
5. Threshold-MEM(1,1):

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

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

Fernandes and Grammig (2005) proposed a class of goodness-of-fit tests for multiplicative error models. In a numerical study, they compared several competing tests and observed that what was referred to as the D-test performed the best overall. Therefore, we included this test in our simulations for comparison, but we refer to it as the *FG* test to avoid confusion with the D^+ and D^- statistics introduced in Section 2. We also included the test in Janssen, Swanepoel, and Veraverbeke (2005) with their ε_i replaced by our $\widetilde{\varepsilon}_i$, $1 \leq i \leq n$. This is denoted by *JG*. Although the validity of the bootstrap method for the *JG* test has not yet been established for multiplicative error models, we included it in our study.

The sample sizes $n = 1000$, 2000, and 3000 were considered in the simulations. To start the recursive data generating process, the initial value of Ψ_i was set equal to its unconditional mean. To reduce the effect of initialization, we generated $(n+300)$ observations, discarded the first 300, and used the remaining n observations for the Monte Carlo sample. For the estimator $\hat{\theta}$ in (5), we used $g_\theta = \log f_\theta$, where f_θ is the pdf of F_θ. Type I error rates and estimates of power were obtained using 5000 Monte Carlo samples. To reduce the computational burden, we adopted the 'Warp-Speed' Monte Carlo method of Giacomini, Politis, and White (2013).

The results for $n = 1000$ and 5% significance level are given in Table 1. The main observations are as follows:
(a) All the tests performed well in terms of type I error rate at the nominal 5% significance level (see the top band in Table 1). The conclusion is the same for the tests at other levels of significance and sample sizes.
(b) Our tests exhibited higher power than *JG* and *FG* tests when the conditional mean was correctly specified and the error distribution was misspecified. In particular, the Anderson–Darling type A^2 test exhibited the best overall performance, followed by the Cramér–von Mises type test *CvM*.
(c) When the conditional mean was misspecified, our tests exhibited higher power than *JG* and *FG* tests did, with the A^2-test performing significantly better in most cases, and at least as well in the rest.
Thus, in terms of size and power, the tests introduced in this paper performed better than their competitors, with A^2 performing the best.

In parametric density forecasting, our tests can be used for testing the goodness-of-fit of the parametric model. To evaluate the potential contribution of the tests in this regard, a simulation study was performed in which we estimated the 'loss' in using an incorrect parametric family when the tests have adequate power to reject the incorrect family. Details of the study are provided in the online Supplementary Material.

6. AN EMPIRICAL ILLUSTRATION

We consider an example on *realized volatility*. The data for the example were obtained from Christian T. Brownlees. The dataset is based on 1989 observations of United Technologies [UTX] daily stock returns between January 2, 2001, and December 31, 2008. For details on the construction of realized volatility, see

TABLE 1. Performance of goodness-of-fit tests of 'H_0 : MEM(1,1) with error distribution F_θ' at 5% significance level

True DGP			Tests						
Cond. Mean	Error df	F_θ (in H_0)	JG	FG	CvM	U^2	A^2	KS	Ku
			Type I error rates (%)						
MEM(1,1)	Exponential	Exponential	5.1	4.5	4.7	4.9	5.1	4.7	4.6
		Weibull	4.9	4.4	4.5	4.5	4.6	4.2	4.1
	Gamma	Gamma	5.8	5.4	5.0	4.5	5.2	4.7	4.3
		Gen. Gamma	4.8	4.6	4.6	4.6	4.3	6.0	4.9
	Weibull	Weibull	4.3	5.7	5.5	5.6	5.4	5.3	4.7
		Gen. Gamma	4.2	5.5	4.6	4.7	4.6	5.2	5.1
	Gen. Gamma	Gen. Gamma	5.5	5.2	5.0	4.9	5.3	5.6	5.5
	Burr	Burr	5.8	4.6	4.8	5.0	4.7	4.7	5.2
			Estimated power (%)						
MEM(2,1)	Gamma	Weibull	29	25	59	55	73	41	49
	Weibull	Gamma	93	29	99	96	99	96	94
	Gen. Gamma	Weibull	75	53	96	94	99	86	90
log-MEM	Gamma	Weibull	36	28	74	66	83	54	60
	Weibull	Weibull	26	6	33	22	34	29	19
	Burr	Gen. Gamma	18	12	45	43	53	32	35
MEM(1,1)	Gamma	Weibull	28	22	59	54	75	42	48
	Weibull	Gamma	92	26	99	96	99	94	92
	Burr	Gen. Gamma	15	12	43	43	52	28	35
Threshold-MEM	Gamma	Gamma	22	11	42	30	47	30	24
	Exponential	Gamma	30	13	52	39	58	40	30
		Exponential	74	12	83	56	81	71	57
	Burr	Gen. Gamma	15	14	43	43	49	31	34
Exp-MEM	Gamma	Weibull	45	38	81	77	91	64	73
	Weibull	Gen. Gamma	59	17	63	54	61	60	51
		Burr	54	18	60	50	59	57	47

Note: (1) The error distribution F_θ in the third column completely determines the null hypothesis, 'H_0: MEM(1,1) with error distribution F_θ'. (2) The abbreviation *Gen. Gamma* refers to Generalized Gamma. (3) The test statistics are Cramér–von Mises [CvM], Watson's [U^2], Anderson-Darling [A^2], Kolmogorov–Smirnov [KS], Kuiper [Ku], Janssen *et al.* (2005) [JG], and the D-test of Fernandes and Grammig (2005) [FG]; for more details about the tests, see Section 2. (4) The sample size n is 1000.

Brownlees *et al.* (2012) and Brownlees and Gallo (2006). We are interested in evaluating the goodness-of-fit of several MEMs and to illustrate the relevance of the tests in density forecasting.

Goodness-of-fit of parametric models

First, we evaluate the goodness-of-fit of MEM(1,1) for the mean Ψ_i with unspecified error distribution. To this end, we considered the *Ljung-Box Q* test with lags 5, 10, 15, and 20 (Ljung and Box, 1978, Engle and Russell, 1998), a *Lagrange multiplier* [LM] test with MEM(2, 1) as the alternative (see Theorem 1 in Meitz and Teräsvirta, 2006), and the *generalized moment* test of Chen and Hsieh (2010) with $(\varepsilon_{i-1} - 1)$ as the 'misspecification indicator'. The smallest p-value for these six tests was 0.56. Therefore, the MEM(1,1) model appears to provide a good fit for the conditional mean, as further confirmed by visual inspection of the residual plots and the correlogram. In response to a reviewer's comment, we also evaluated the Asymmetric MEM model of Fernandes and Grammig (2006) that appears as model number 4 in their Table 1, but it did not improve the fit. Therefore, for the rest of this section, we restrict our attention to the simpler MEM(1,1) model for Ψ_i and evaluate it in combination with different error distributions.

We considered the six error distributions, Exponential, Weibull, Gamma, Generalized Gamma, Burr, and a mixture of Burr and Generalized Gamma. Further, we applied the seven tests studied in the simulations. For every test, the p-value corresponding to the first four error distributions (Exponential, Weibull, Gamma, Generalized Gamma) turned out to be nearly zero (see Table 2). Therefore, we rule out these four distributions and focus on Burr distribution and a mixture of Burr and Generalized Gamma distributions. In the previous section, we observed that the Anderson-Darling type statistic A^2 performed the best overall. Hence, we restrict our attention to A^2; the conclusions are practically the same for Cramer–von Mises and Kolmogorov–Smirnov tests as well. It appears from the results for A^2 in Table 2 that a mixture of Burr and Generalized Gamma provides the best fit, at least based on the p-values.

Apart from the p-values, it is also of interest to evaluate graphically the goodness-of-fit of each parametric error distribution. To this end, we constructed the QQ-plots of the residuals for every error distribution. Figure 1 shows the plots

TABLE 2. Goodness-of-fit of 'H_0 : MEM(1,1) with error distribution F_θ' for UTX realized volatility, and its performance in conditional-density forecasting

	Tests							
	JG	FG	CvM	U^2	A^2	KS	Ku	
F_θ (in H_0)	p-values							Log-score
Exponential	0.00	0.00	0.00	0.00	0.01	0.00	0.00	−2.03
Weibull	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−1.78
Gamma	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−1.66
Generalized Gamma	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−1.60
Burr	0.97	0.34	0.21	0.37	0.10	0.46	0.85	−1.56
Mixture of Burr Generalized Gamma	0.45	0.35	0.45	0.36	0.37	0.45	0.40	−0.57

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

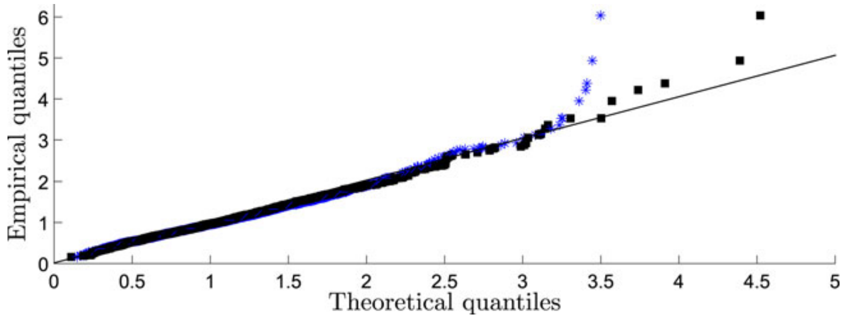

FIGURE 1. QQ plots of residuals when (a) the mean function is MEM(1,1) and the error distribution is Generalized Gamma [$*$], and (b) the mean function is MEM(1,1) and the error distribution is a mixture of Burr and Generalized Gamma [■].

for the Generalized Gamma and the mixture of Burr and Generalized Gamma error distributions. The latter is significantly closer to a straight line than the former is, confirming the conclusion based on the goodness-of-fit tests. Other plots not included here also indicated that Exponential, Weibull, Gamma, and Generalized Gamma error distributions do not fit as well as the Burr or mixture of Burr and Generalized Gamma distributions do. In summary, MEM(1,1) with a mixture of Burr and Generalized Gamma for the error distribution appears to provide the best fit.

Density/distribution forecasting:

We wish to evaluate how well the models in Table 2 perform in out-of-sample forecasting and the extent to which the forecast performance corresponds to measures of goodness-of-fit.

To evaluate the density forecasts, we applied a method of Diebold *et al.* (1998). To this end, first note that the one-step-ahead density forecast produced by an MEM, using information available up to time $i-1$, is $\widehat{f_i}(x|\mathcal{H}_{i-1}) := \{\widetilde{\Psi}_i(\hat{\phi})\}^{-1} f_{\hat{\theta}}(x/\widetilde{\Psi}_i(\hat{\phi}))$, and the *probability integral transform* of Z_i is $\widehat{F}_i(Z_i|\mathcal{H}_{i-1}) := \int_0^{Z_i} \widehat{f_i}(x|\mathcal{H}_{i-1})dx = F_{\hat{\theta}}(Z_i/\widetilde{\Psi}_i(\hat{\phi}))$, where Z_i is the variable to be forecast at time i. If the density forecast is correct, then the sequence of probability integral transforms $\{\widehat{F}_i(Z_i|\mathcal{H}_{i-1})\}_{i=1}^n$ is iid uniform on the unit interval (Diebold *et al.*, 1998). Therefore, density forecasts may be evaluated by using methods to detect departures of the empirical distribution of $\{\widehat{F}_i(Z_i|\mathcal{H}_{i-1})\}_{i=1}^n$ from the uniform distribution (see Bauwens *et al.*, 2004; Corsi, Mittnik, Pigorsch, and Pigorsch, 2008).

To apply the preceding method, we first estimated the six models corresponding to the six error distributions by using only the observations, which we denote by $\{Z_1, \ldots, Z_j\}$, for the subperiod, January 2, 2001 to June 29, 2007. Then, we re-estimated the model for the subsample $\{Z_1, \ldots, Z_{j+1}\}$, thus expanding the time period by 1. We repeated the process until the entire period was covered, and then examined the empirical cumulative distributions of the corresponding probability integral transforms. The main results are summarized in Appendix D of the online Supplementary Material. The plots (not included) for Exponential, Weibull and Gamma error distributions deviate significantly from that for the uniform distribution. The plot for Generalized Gamma also deviates from the uniform distribution. However, the plots for Burr and for mixture of Burr and Generalized Gamma are somewhat closer to being uniformly distributed, and therefore these two error distributions appear to have performed better for forecasting than the other four.

To complement the preceding visual inspections, we also considered the *logarithmic score* to evaluate the density forecasts (e.g. Corsi *et al.*, 2008). The logarithmic score is $\mathcal{S} = n^{-1}\sum_{t=1}^{n}\log\{\widehat{f_i}(Z_i|\mathcal{H}_{i-1})\}$. A large $\mathcal{S}$ value implies better predictive ability. Ranking of the models in terms of $\mathcal{S}$ turned out to be approximately in the order in which the goodness-of-fit tests also ranked them in terms of p-values (see Table 2). The consistency of the results between forecast performance and the goodness-of-fit tests illustrates the importance of the proposed tests for density forecasting.

7. DISCUSSION AND CONCLUSION

We have developed a family of Kolmogorov–Smirnov and Cramér–von Mises type tests for the specification of the multiplicative error models, which includes the well-known family of autoregressive conditional duration models. In a simulation study, the proposed tests performed better than their competitors did. A data example illustrated that the tests adequately complement the residual diagnostics, such as the QQ plots, for evaluating goodness-of-fit. Further, the empirical example and the simulations presented in the previous two sections illustrated that the use of the goodness-of-fit tests in density/distribution forecasting can be expected to reduce loss resulting from the use of a misspecified model.

Throughout, we assumed that the null model was estimated by the QMLE corresponding to the standard exponential error distribution; see also Fernandes and Grammig (2005). Another obvious possibility is to use a maximum likelihood estimator [MLE] based on the null model. However, an (unreported) Monte Carlo simulation based on a range of error distributions showed that none of the two approaches dominates the other. A third approach, based on MLE of a finite dimensional regular parametric model that nests both the null and the alternative, is likely to be asymptotically optimal; however, since in our setting the alternative hypothesis does not specify such a parametric model, in this paper we do not pursue this approach.

If for some $j \in \{1,\dots,5\}$, the test based on T_j rejects H_0, then the source of the misspecification of the model may be $\Psi_i(\phi)$ and/or F_θ. Unfortunately, the value of the test statistic T_j does not guide us to the particular source of the violation. Thus, a natural question that arises is, what would be a reasonable further step if H_0 is rejected by a T_j? We recommend a procedure that is similar to the one illustrated in the empirical example of Section 6. More specifically, we suggest

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

that tests designed only for the specification of $\Psi_i(\phi)$ be applied first. Once a suitable specification for $\Psi_i(\phi)$ has been chosen, our tests can be used to test each of the several possible error distributions combined with the $\Psi_i(\phi)$ chosen using other tests.

NOTE

1. A set $F \subset \mathbb{R}^+ \cup \{0\}$ is called a δ-net of $\left(\mathbb{R}^+ \cup \{0\}, \mu_b^B\right)$ if and only if for each $x \in \mathbb{R}^+ \cup \{0\}$, there exists a $y \in F$ such that $\mu_b^B(x, y) \leq \delta$. A minimal δ-net has the smallest cardinality amongst all possible δ-nets of $\left(\mathbb{R}^+ \cup \{0\}, \mu_b^B\right)$.

REFERENCES

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

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

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

Brownlees, C.T., F. Cipollini, & G.M. Gallo (2012) Multiplicative error models. In L. Bauwens, C. Hafner, & S. Laurent (eds.), *Handbook of Volatility Models and Their Applications*, Chapter 9, pp. 225–247. Wiley.

Chen, Y.-T. & C.-S. Hsieh (2010) Generalized moment tests for autoregressive conditional duration models. *Journal of Financial Econometrics* 8, 345–391.

Chou, R.Y. (2005) Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking* 37, 561–582.

Corradi, V. & N.R. Swanson (2006) Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics* 133, 779–806.

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

Diebold, F.X., T.A. Gunther, & A.S. Tay (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–883.

Engle, R.F. (2000) The Econometrics of ultra-high-frequency data. *Econometrica* 68, 1–22.

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

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

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

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

Freedman, D.A. (1975) On tail probabilities for martingales. *Annals of Probability* 3, 100–118.

Ghysels, E., C. Gouriéroux, & J. Jasiak (2004) Stochastic volatility duration models. *Journal of Econometrics* 119, 413–433.

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

Giot, P. (2000) Time transformations, intraday data, and volatility models. *Journal of Computational Finance* 4, 31–62.

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

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

Koul, H.L. & S. Ling (2006) Fitting an error distribution in some heteroscedastic time series models. *The Annals of Statistics* 34, 994–1012.

Koul, H.L. & M. Ossiander (1994) Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *The Annals of Statistics* 22, 540–562.

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

Ling, S. & H. Tong (2011) Score based goodness-of-fit tests for time series. *Statistica Sinica* 21, 1807–1829.

Ljung, G.M. & G.E.P. Box (1978) On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.

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

Meitz, M. & T. Teräsvirta (2006) Evaluating models of autoregressive conditional duration. *Journal of Business & Economic Statistics* 24, 104–124.

Ossiander, M. (1987) A central limit theorem under metric entropy with L_2 bracketing. *The Annals of Probability* 15, 897–919.

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

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

APPENDIX A: Assumptions, preliminary results and main proofs

In this Appendix, we provide proofs for the main results stated in Sections 2, 3, and 4, together with several preliminary lemmas. Proofs of some of the results are given in the online Supplementary Material. Let ‘plim’ denote the probability limit operator. For any nonnegative integer m, let $\{\Psi_i^{(m)}(\phi), Z_i^{(m)}\}$ denote the process that starts with $\Psi_i^{(m)}(\phi) = 1$ and $Z_i^{(m)} = \bar{Z}$ for $i < -m$ and follows the model defined by (1) and (2) for $i \geq -m$, with conditional mean $\Psi_i^{(m)} = \Psi_i^{(m)}(\phi_0)$ and error distribution F_{θ_0}, where $(\phi_0, \theta_0) = \text{plim}\,(\hat{\phi}, \hat{\theta})$. Let $\widetilde{\Psi}_i^{(m)}(\phi)$ denote the analogue of $\widetilde{\Psi}_i(\phi)$ for $\{Z_i^{(m)} : i \in \mathbb{N}\}$, where $\mathbb{N} := \{1, 2, 3, \cdots\}$. We say that a sequence of random variables $\{X_i : i = 1, 2, \cdots\}$ converges *exponentially almost surely [e.a.s]* to zero, denoted $X_i \overset{e.a.s.}{\to} 0$, if there exists a $\gamma > 1$ such that $\gamma^i X_i \overset{a.s.}{\to} 0$ as $i \to \infty$ (see Straumann and Mikosch, 2006).

First, we state two sets of regularity conditions.

Condition C.

(C1). *The process $\{Z_i : i \in \mathbb{Z}\}$ is strictly stationary and ergodic, and $\mathbb{E}(Z_i^{2+d}) < \infty$ for some $d > 0$. The function $\Psi_i(\phi)$ is twice continuously differentiable with respect to ϕ (a.s.). For each $\phi \in \Phi$ and $i \geq 1$, $\Psi_i(\phi)$ is bounded away from zero with probability one, and $\mathbb{E}[\Psi_i(\phi)^{2+d}]$ and $\mathbb{E}[\|\lambda_i(\phi)\|^{2+d}]$ are finite. The parameter space Φ is a compact subset of $\mathbb{R}^p$ and $\phi_0 = \text{plim}\,\hat{\phi}$ is an interior point of Φ.*

(C2). *For each $K < \infty$, $\sup \sqrt{n}\, |\, \Psi_i(t) - \Psi_i(s) - (t-s)^\top \dot{\Psi}_i(s)\, |\, / \,\Psi_i(\phi_0) = o_p(1)$, where the supremum is taken over $1 \leq i \leq n$ and over $\{(t,s) : t, s \in \Phi, \sqrt{n}\|t - s\| \leq K\}$.*

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

(C3). *(i) For each $\theta \in \Theta$, F_θ has a positive density f_θ almost everywhere (a.e.).*
(ii) F^0 has a positive density f^0 (a.e.).
(iii) $f_\theta(y)$, $F_\theta(y)$, $F_\theta^{-1}(y)$ and $f^0(y)$ are twice continuously differentiable in (θ, y).
(iv) There exists an open neighbourhood B of θ_0, such that $\sup_{y\geq 0,\theta\in B}(1+y)f_\theta(y)$, $\sup_{y\geq 0,\theta\in B}\|\dot{F}_\theta(y)\|^{1+d}$, $\sup_{y\geq 0,\theta\in B}\|(\partial^2/\partial\theta\partial\theta^T)F_\theta(y)\|^{1+d}$, $\sup_{\theta\in B}\int|1-y|^{1+d}f_\theta(y)dy$ are all finite for some $d>0$.
(v) There exists an $a>0$ such that $\sup_{y\geq 0,|u|<a}(1+y^2)f'\{(1+u)y\}<\infty$, and $\sup_{\theta\in\Theta}\int y^2 f(y)dy<\infty$, where f may be the true density of the error term ε_i or a member of the parametric family $\{f_\theta : \theta\in\Theta\}$. Further, $\|\dot{f}_\theta(y)\|\leq K_\theta(y)$ for some function $K_\theta(y)$ satisfying $\sup_{\theta\in B}\int K_\theta(y)dy<\infty$.

(C4). $\sup_{\phi\in\Phi}|\widetilde{\Psi}_i(\phi)-\Psi_i(\phi)|\overset{e.a.s.}{\to}0$*, and there exists an open neighbourhood B of ϕ_0 such that* $\sup_{\phi\in B}\|\dot{\widetilde{\Psi}}_i(\phi)-\dot{\Psi}_i(\phi)\|\overset{e.a.s.}{\to}0$*, as $i\to\infty$.*

(C5). $\max_{1\leq i\leq n}n^{-1/2}\|\lambda_i(\phi_0)\|=o_p(1)$.

(C6). *(i)* $\sup_{\phi\in\Phi}|\Psi_i^{(m)}(\phi)-\Psi_i^{(\infty)}(\phi)|$, $\sup_{\phi\in\Phi}|\widetilde{\Psi}_i^{(m)}(\phi)-\Psi_i^{(m)}(\phi)|\overset{e.a.s.}{\to}0$ *as $i\to\infty$.*
(ii) There exists an open neighbourhood B of ϕ_0 such that, as $i\to\infty$,
(a) $\sup_{\phi\in B}\|\dot{\Psi}_i^{(m)}(\phi)-\dot{\Psi}_i^{(\infty)}(\phi)\|\overset{e.a.s.}{\to}0$, *and (b)* $\sup_{\phi\in B}\|\dot{\widetilde{\Psi}}_i^{(m)}(\phi)-\dot{\Psi}_i^{(m)}(\phi)\|\overset{e.a.s.}{\to}0$.

Assumption E.

(E1). *If Ψ_i is of the form $\Psi_i(\phi)$ then ϕ_0 is the true value satisfying $\Psi_i=\Psi_i(\phi_0)$. Otherwise, $\Psi_i\neq\Psi_i(\phi_0)$. There exists an open neighbourhood B of ϕ_0 such that, (a) $\lim\mathbb{E}\{-n^{-1}L_n(\phi)\}$ has a maximum at $\phi=\phi_0$ on B, (b) the eigenvalues of the hessian of $-n^{-1}L_n(\phi)$ for $\phi\in B$ are all less than $-\xi$ for some $\xi>0$, with probability tending to one, and (c) $\|\sqrt{n}(\hat{\phi}-\phi_0)-[n^{-1/2}\tau(\phi_0)\sum_{i=1}^n\lambda_i(\phi_0)(1-\varepsilon_i)]\|=o_p(1)$, where $\tau(\phi)$ is a $p\times p$ matrix that is uniformly continuous as a function of ϕ on B.*

(E2). *The parameter space Θ is compact, and the estimator $\hat{\theta}$ converges in probability to an interior point θ_0 of Θ. If H_0 is true, then θ_0 is the true value satisfying $F^0=F_{\theta_0}$. Otherwise, $F^0\neq F_{\theta_0}$. The function $g_\theta(t)$ is twice continuously differentiable with respect to both θ and t and it satisfies (a) $\mathbb{E}[\dot{g}_{\theta_0}(\varepsilon_1)]=0$ where $\dot{g}_\theta(t):=(\partial/\partial\theta)g_\theta(t)$, and (b) there exist $d>0$, $0<K_0<\infty$, and an open neighbourhood B of θ_0, such that $\sup_{\theta\in B}\int\|\{\mathbb{E}[(\partial/\partial\theta)\dot{g}_\theta(\varepsilon_1)]\}^{-1}\dot{g}_\theta(t)\|^{2+d}f_\theta(t)dt<K_0$.*

(E3). *The Conditions (C1)–(C6) and Assumptions (E1)–(E2) continue to hold when (F^0,ϕ_0) is replaced by (F_{θ_n},ϕ_n), where $(\phi_n,\theta_n)\to(\phi_0,\theta_0)$ as $n\to\infty$.*

The asymptotic expansion of $\sqrt{n}(\hat{\phi}-\phi_0)$ in Assumption (E1) is typically satisfied by the QML estimator with $\tau(\phi)=\{\mathbb{E}[-\lambda_i(\phi)\lambda_i(\phi)^\top]\}^{-1}$ (see Bauwens and Giot, 2001; Hautsch, 2011). We would expect (C2) to be satisfied by a large class of multiplicative error models because it essentially says that the remainder term in a one-term Taylor expansion of $\Psi_i(\phi)$ is small compared to $\Psi_i(\phi_0)$. These two conditions imply that $n^{1/2}\{\Psi_i(\hat{\phi})-\Psi_i(\phi_0)\}/\Psi_i(\phi_0)$ is bounded in probability with asymptotic mean zero. For most parametric families, Assumption (E2) would typically be satisfied if $g_\theta=\log f_\theta$. Condition (C3) introduces some smoothness and tail conditions on F_θ and F^0. Conditions (C4)–(C6) will be used in the Appendix to obtain certain technical results, for example, to show that a metric entropy is small and residual empirical processes converge

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

uniformly (see the proof of Lemma A.5). The preceding regularity conditions are satisfied by the well-known MEM(p_1, p_2) model.

A.1. A general result for weighted empirical processes based on triangular arrays of nonnegative random variables

This section obtains some general results on empirical processes for residuals from a nonlinear time series when the data generating process depends on the sample size and hence takes the form of a triangular array. These results are required to derive the asymptotic distributions of the test statistics, $T_1, \ldots, T_5$, and to establish the consistency of the bootstrap. To formulate the setting, let us introduce Assumption T.

Assumption T [A general setting for triangular arrays]:

(a) Θ is a compact subset of $\mathbb{R}^q$, and ϑ_0 is an interior point of Θ.

(b) $\{H_\vartheta : \vartheta \in \Theta \subset \mathbb{R}^q\}$ is a family of distribution functions on $[0,\infty)$.

(c) $H_\vartheta(x)$ is continuous in ϑ, twice continuously differentiable in x, and it has positive density $h_\vartheta(x)$ (a.e) for every $\vartheta \in \Theta$.

(d) $\sup_{\vartheta \in \Theta} \sup_{x \geq 0} h_\vartheta(x) < \infty$, $\sup_{\vartheta \in \Theta} \int_{x \geq 0} |x| h_\vartheta(x)\, dx < \infty$.

(e) $(\eta_{ni}, \gamma_{ni}, \rho_{ni}), 1 \leq i \leq n$, is an array of random variables defined on a probability space $(\Omega, \mathcal{A}, P_n)$, and η_{ni} is independent of $(\gamma_{ni}, \rho_{ni}), 1 \leq i \leq n$.

(f) $\{\eta_{ni}, 1 \leq i \leq n\}$ is iid with distribution function H_{ϑ_n}, $\vartheta_n \to \vartheta_0$ as $n \to \infty$.

(g) There exists a triangular array of sub sigma-fields $\{\mathcal{F}_{ni}\}$ such that $\mathcal{F}_{ni} \subset \mathcal{F}_{n(i+1)}$ for $1 \leq i < n$, $\{\eta_{n,i-1}, \gamma_{ni}, \rho_{ni}, 2 \leq i \leq j\}$ are $\mathcal{F}_{nj}$-measurable for $2 \leq j \leq n$, (γ_{n1}, ρ_{n1}) is $\mathcal{F}_{n1}$-measurable, and η_{nj} is independent of $\mathcal{F}_{nj}$.

(h) $n^{-1}\sum_{i=1}^n \gamma_{ni}^2 = O_{p_n}(1)$, $\max_{1\leq i \leq n} n^{-1/2}|\gamma_{ni}| = o_{p_n}(1)$, $\max_{1 \leq i \leq n} |\rho_{ni}| = o_{p_n}(1)$.

Let us define

$$U_n(x) = n^{-1/2}\sum_{i=1}^n \gamma_{ni}[I(\eta_{ni} \leq x) - H_{\vartheta_n}(x)],$$
$$\widetilde{U}_n(x) = n^{-1/2}\sum_{i=1}^n \gamma_{ni}[I(\eta_{ni} \leq x + x\rho_{ni}) - H_{\vartheta_n}(x + x\rho_{ni})], \quad x \geq 0.$$

The main result of this subsection is that $\sup_{x\geq 0}|\widetilde{U}_n(x) - U_n(x)| = o_{p_n}(1)$. Since the proof of this result is long, we segment it into a few lemmas.

LEMMA A.1. *Let $M_n = \sum_{i=1}^n D_{ni}$ be a sum of martingale differences defined on the triangular array of sub sigma-fields $\{\mathcal{A}_{ni}\}$, where $\mathcal{A}_{ni} \subset \mathcal{A}_{n(i+1)}, 1 \leq i \leq n$. Assume that $|D_{ni}| \leq a$ (a.s.) for $1 \leq i \leq n$. Then, for any $\eta, \alpha > 0$, we have that*
$P_n\{[M_n > \eta] \cap [\sum_{i=1}^n \mathbb{E}(D_{ni}^2 \mid \mathcal{A}_{n(i-1)}) \leq \alpha]\} \leq \exp\{-\eta^2/2(a\eta + \alpha)\}$.

The proof of the lemma follows from Proposition 2.1 in Freedman (1975). This is a general result on the sum of martingale differences defined in the form of a triangular array; it is not specifically for MEM.

LEMMA A.2. *Suppose that Assumption T is satisfied. Let $a > 0, b > 0$ and $x \geq 0$, such that* $\sup_{\vartheta_n \in B} \sup_{|z|\leq b} |H_{\vartheta_n}\{x(1+z)\} - H_{\vartheta_n}(x)| < a$. *Let B be an open neighbourhood of*

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

the limit ϑ_0*, and let* $\Pi_n = \{\max_i |\gamma_{ni}| \le an^{1/2}\} \cap \{\max_i |\rho_{ni}| \le b\} \cap \{n^{-1}\sum_{i=1}^n \gamma_{ni}^2 \le c\}$*. Then, for* $\eta, c > 0$ *and* $\vartheta_n \in B$*, we have that*

$$P_n\big([|\widetilde{U}_n(x) - U_n(x)| > \eta] \cap \Pi_n\big) \le \exp\{-\eta^2/2a(\eta + c)\}.$$

Proof. Detailed proof is given in the online Supplementary Material. The main part is to show that the quadratic variation of $\widetilde{U}_n(x) - U_n(x)$ is bounded from above by $(a/n)\sum_{i=1}^n \gamma_{ni}^2$. Then, the proof follows by applying Lemma A.1. ■

The proofs of Lemmas A.3 and A.4 stated below are given in the online Supplementary Material. Lemmas A.3 states the asymptotic equivalence of $\widetilde{U}_n$ and U_n at a given point x. This is similar to Corollary 2.1 of Koul and Ossiander (1994) for the analogous location problem.

LEMMA A.3. *Suppose that Assumption T is satisfied. Then, for each fixed* $x \ge 0$*,* $|\widetilde{U}_n(x) - U_n(x)| = o_{p_n}(1)$.

The next lemma is the analogue of condition (A5) of Theorem 1.1 in Koul and Ossiander (1994) for the current setup. This is the crucial result needed for the chaining argument used in the proof of Lemma A.5 given below.

LEMMA A.4. *Let* H_ϑ, ϑ_0*, and* Θ *be as in Assumption T. Let B be a given open neighbourhood of* ϑ_0*, and* $\mu_b^B(x,y) = \sup_{\vartheta\in B}\sup_{|z|\le b} \mid H_\vartheta\{x(1+z)\} - H_\vartheta\{y(1+z)\} \mid^{1/2}$. *Then,* μ_b^B *forms a totally bounded pseudo-metric on* $\mathbb{R}^+ \cup \{0\}$ *for every* $b > 0$*. Let the entropy integral be defined as* $I(b) := \int_0^1 [\log \mathcal{N}(\delta, b)]^{1/2} d\delta < \infty$ *for all* $0 \le b < 1$*, where* $\mathcal{N}(\delta, b)$ *is the cardinal number of a minimal δ-net of* $(\mathbb{R}^+ \cup \{0\}, \mu_b^B)$.[1] *Then,* $I(b) < \infty$.

The next lemma is the main result of this subsection. This is an extension of Theorem 4.1 of Koul and Ling (2006) and Theorem 1.1 of Koul and Ossiander (1994).

LEMMA A.5. *Suppose that Assumption T holds. Then,* $\sup_{x\ge 0} |\widetilde{U}_n(x) - U_n(x)| = o_{p_n}(1)$.

Proof. Let $\mathcal{D}_{ni}^{(b)}(x) = [I\{\eta_{ni} \le x(1+\rho_{ni})\} - H_{\vartheta_n}\{x(1+\rho_{ni})\}] I\{\max_i |\rho_{ni}| \le b\}$ for $1 \le i \le n$. By using Lemma A.4 and a method of stratification and chaining based on $\mathcal{D}_{ni}^{(b)}$, for example as in Section 3 of Ossiander (1987), it may be shown that the processes $\widetilde{U}_n$ and U_n are asymptotically uniformly equicontinuous in the μ_b^B metric, $0 \le b < 1$. Therefore, for every $0 \le b < 1$ and $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} \bar{P}_n\left(\sup_{\mu_b^B(x,y)\le\delta,\ x,y\ge 0} \mid \widetilde{U}_n(x) - \widetilde{U}_n(y) \mid > \varepsilon\right) < \varepsilon, \tag{A.1}$$

$$\limsup_{n\to\infty} \bar{P}_n\left(\sup_{\mu_b^B(x,y)\le\delta,\ x,y\ge 0} \mid U_n(x) - U_n(y) \mid > \varepsilon\right) < \varepsilon, \tag{A.2}$$

where $\bar{P}_n$ denotes the outer probability. In the sequel, for brevity of notation, we omit the bar in $\bar{P}_n$.

Next, let $\Pi_n := \{\max_i |\gamma_{ni}| \le an^{1/2}\} \cap \{\max_i |\rho_{ni}| \le b\} \cap \{n^{-1}\sum \gamma_{ni}^2 \le c\}$. Fix an $\varepsilon > 0$ and $0 \le b_0 < 1$. Then, $I(b) < \infty$ and (A.1) and (A.2) hold for every $0 \le b \le b_0$.

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

Therefore, it may be argued as in the proof of Theorem 1.1 in (Koul and Ossiander, 1994, pp. 556–557) that there exist constants $a, \delta > 0$, $c \geq 1$ and $b \leq b_0$ with $\exp\{-\varepsilon^2/2a(\varepsilon + c)\} \leq \varepsilon^2/\mathcal{N}(\delta, b)$ and $\limsup_n P_n(n^{-1}\sum_i \gamma_{ni}^2 > c) < \varepsilon$, such that for all $n > 16c\{1 + \ln \mathcal{N}(u, b)\}/\delta^2$,

$$P_n\left\{\left[\sup_{x\geq 0}|\widetilde{U}_n(x) - U_n(x)| > 3\varepsilon\right] \cap \Pi_n\right\}$$
$$\leq \mathcal{N}(\delta, b)\sup_{x\geq 0} P_n\left\{\left[|\widetilde{U}_n(x) - U_n(x)| > \varepsilon\right] \cap \Pi_n\right\} + 2\varepsilon.$$

Because $n^{-1}\sum_{i=1}^n \gamma_{ni}^2 = O_{p_n}(1)$, $\max_{1\leq i\leq n} n^{-1/2}|\gamma_{ni}| = o_{p_n}(1)$ and $\max_{1\leq i\leq n}|\rho_{ni}| = o_{p_n}(1)$, one obtains that $\limsup_n P_n(\Pi_n^c) = \limsup_n P_n(n^{-1}\sum_i \gamma_{ni}^2 > c) < \varepsilon$. Further, by Lemmas A.1 and A.2,

$$\sup_{x\geq 0} P_n\left\{\left[|\widetilde{U}_n(x) - U_n(x)| > \varepsilon\right] \cap \Pi_n\right\} \leq \exp\left\{-\varepsilon^2/2a(\varepsilon + c)\right\} \leq \varepsilon^2/\mathcal{N}(\delta, b).$$

Since ε is arbitrary and $\limsup_n P_n(\Pi_n^c) < \varepsilon$, $\sup_{x\geq 0}|\widetilde{U}_n(x) - U_n(x)| = o_{p_n}(1)$. ■

A.2. Some asymptotic representations and convergence results for bootstrap estimators and test statistics

Throughout this subsection, we assume that Conditions (C1) – (C6) and Assumptions (E1) – (E3) are satisfied. The uniform metric is assumed for weak convergence in $D[0, 1]$. When the DGP is the MEM in (1) with $\Psi_i = \Psi_i(\phi_n)$ and $F^0 = F_{\theta_n}$, we say that the *DGP corresponds to* (ϕ_n, θ_n) and denote the probability measure by P_n.

LEMMA A.6. *Let (ϕ_n, θ_n) be a given sequence converging to (ϕ_0, θ_0) as $n \to \infty$, and let $\upsilon_{ni}(\phi) = \{\Psi_i(\phi) - \Psi_i(\phi_n)\}/\Psi_i(\phi_n)$. Suppose that the DGP corresponds to (ϕ_n, θ_n). Let $B_n(x) = x f_{\theta_n}(x) n^{1/2}(\phi - \phi_n)^\top \mathbb{E}\{\lambda_1(\phi_n)\}$. Then, for any $M < \infty$,*

$$\sup_{x,\phi,M}|n^{-1/2}\sum_{i=1}^n\left[F_{\theta_n}\{x + x\upsilon_{ni}(\phi)\} - F_{\theta_n}(x)\right] - B_n(x)| = o_{p_n}(1),$$

where $\sup_{x,\phi,M}$ is the supremum over $\{(x,\phi) : x \geq 0, \phi \in \Phi \sqrt{n}\|\phi - \phi_n\| \leq M\}$.

Proof. A detailed proof is given in the online Supplementary Material. The main strategy is to first expand $F_{\theta_n}\{x + x\upsilon_{ni}(\phi)\}$ as

$$F_{\theta_n}\{x + x\upsilon_{ni}(\phi)\} = F_{\theta_n}(x) + x\upsilon_{ni}(\phi) f_{\theta_n}(x) + o_{p_n}(1). \quad \textbf{(A.3)}$$

Then, express $\upsilon_{ni}(\phi)$ as a linear term in $(\phi_n - \phi)$ such that

$$\upsilon_{ni}(\phi) = \{\Psi_i(\phi_n)\}^{-1}\{\Psi_i(\phi) - \Psi_i(\phi_n)\} \approx (\phi - \phi_n)^\top \dot{\Psi}_i(\phi_n)/\Psi_i(\phi_n). \quad \textbf{(A.4)}$$

The remainder terms in (A.4) are all $o_{p_n}(1)$ uniformly over $\{\sqrt{n}\|\phi - \phi_n\| \leq M\}$. Hence, the desired result can be obtained by combining (A.3) and (A.4). ■

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

Recall that, in Section 3, we used the superscript '$*(m)$' for the bootstrap process that starts at time $-m$. Now, for the limiting case $m=\infty$, we use the superscript '$*$' instead of '$*(\infty)$'; for example, (Z_i^*, Ψ_i^*) denotes $(Z_i^{*(\infty)}, \Psi_i^{*(\infty)})$ and $(\hat{\phi}^*, \hat{\theta}^*)$ denotes $(\hat{\phi}^{*(\infty)}, \hat{\theta}^{*(\infty)})$. We establish that the bootstrap is valid for * and show that the error resulting from using $^{*(m)}$ in place of *, becomes negligible for large n. Some of the preliminary results required for this are obtained in the next few lemmas.

LEMMA A.7. *Suppose that Conditions (C1), (C2), (C4), (C5), and Assumptions (E1) and (E2) are satisfied. Then, under H_0, for each $j=1,\ldots,q$,*

$$\hat{\theta}_j - \theta_{0j} = n^{-1}\textstyle\sum_{i=1}^n \left\{h_{\theta_0 j}(\varepsilon_i) - \left(\hat{\phi}-\phi_0\right)^\top \varepsilon_i \lambda_i(\phi_0) h'_{\theta_0 j}(\varepsilon_i)\right\} + o_p\left(n^{-1/2}\right),$$

where $h_\theta(t) = \{-\mathbb{E}[(\partial/\partial\theta)\dot{g}_\theta(\varepsilon_1)]\}^{-1}\dot{g}_\theta(t) = [h_{\theta 1}(t),\ldots,h_{\theta q}(t)]^\top$. Further, if Assumption (E3) is also satisfied then, under H_0 and under H_1,

$$\hat{\theta}_j^* - \hat{\theta}_j = n^{-1}\sum_{i=1}^n \left\{h^*_{\hat{\theta}_j}(\varepsilon_i^*) - \left(\hat{\phi}^*-\hat{\phi}\right)^\top \varepsilon_i^* \lambda_i^*\left(\hat{\phi}\right) h^{*\prime}_{\hat{\theta}_j}(\varepsilon_i^*)\right\} + o_{p_n^*}\left(n^{-1/2}\right), \qquad \textbf{(A.5)}$$

in probability, where $h_\theta^*(t) = -\{\mathbb{E}_*[(\partial/\partial\theta)\dot{g}_\theta(\varepsilon_1^*)]\}^{-1}\dot{g}_\theta(t) = [h^*_{\theta 1}(t),\ldots,h^*_{\theta q}(t)]^\top$ and $\mathbb{E}_*$ is the bootstrap expectation for the DGP corresponds to $(\hat{\phi},\hat{\theta})$.

Proof. See the online Supplementary Material. ■

The Lemmas A.8–A.11 below are obtained without imposing the null hypothesis. Lemma A.8 yields that the differences $(\hat{\phi}^{*(m)} - \hat{\phi}^*)$ and $(\hat{\theta}^{*(m)} - \hat{\theta}^*)$ converge to zero faster than $n^{-1/2}$. Consequently, we are able to replace $(\hat{\phi}^*, \hat{\theta}^*)$ by $(\hat{\phi}^{*(m)}, \hat{\theta}^{*(m)})$ in asymptotic arguments. The proofs of these results make use of the power of the stronger form of convergence $\overset{e.a.s.}{\to} 0$ introduced in Section 2. These results form an important part in establishing that the nonstationary bootstrap process that starts at time $-m$ and the more tractable stationary hypothetical bootstrap process that starts at time $-\infty$ are close. The proofs of Lemmas A.8 and A.9 are given in the online Supplementary Material.

LEMMA A.8. *There exists an $\eta > 0.5$ such that the following hold in probability: (a) $n^\eta \sup_{\phi\in\Phi} |(\partial/\partial\phi)\{n^{-1}L_n^{*(m)}(\phi) - n^{-1}L_n^*(\phi)\}| \overset{a.s.}{\to} 0$, (b) $\hat{\phi}^{*(m)} - \hat{\phi}^* = O_{p_n^*}(n^{-\eta})$, (c) $\hat{\theta}^{*(m)} - \hat{\theta}^* = o_{p_n^*}(n^{-\eta})$.*

The next lemma yields an asymptotic uniform expansion for the bootstrap process $n^{1/2}\{\hat{F}_n^*(x) - F_n^*(x)\}$. We use the resulting expansion in the proof of Lemma A.10 for showing that the process $\widehat{W}_n^*$ converges weakly.

LEMMA A.9. *Let $U_n^*(x) = n^{1/2}\{\hat{F}_n^*(x) - F_n^*(x)\}$ where $F_n^*(x) = n^{-1}\sum_{i=1}^n I(\varepsilon_i^* \le x)$. Then $\sup_{x\ge 0}|U_n^*(x) - n^{1/2}(\hat{\phi}^*-\hat{\phi})^\top x\mathbb{E}_*\{\lambda_1^*(\hat{\phi})\} f_{\hat{\theta}}(x)| = o_{p_n^*}(1)$, in probability.*

The weak convergence of $\widehat{W}_n^*$ is established in the next lemma.

LEMMA A.10. *Conditional on $\{Z_1,\ldots,Z_n\}$, the hypothetical bootstrap empirical process $\widehat{W}_n^* \circ F_{\hat{\theta}}^{-1}(\cdot) = n^{1/2}\big[\hat{F}_n^*\{F_{\hat{\theta}}^{-1}(\cdot)\} - F_{\hat{\theta}^*}\{F_{\hat{\theta}}^{-1}(\cdot)\}\big]$ converges weakly to $\mathcal{V}(\cdot)$ [in probability], where $\mathcal{V}(\cdot)$ is as in Theorem 2.*

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

Proof. For showing that a quantity is $o_{p_n^*}(1)$ in probability, one may assume without loss of generality that $(\hat\phi,\hat\theta)\to(\phi_0,\theta_0)$ along almost all sample paths; this is explained in the online Supplementary Material. In what follows, we restrict attention to such a fixed sample path. First, partition the process $\widehat{W}_n^*(\cdot)=n^{1/2}\{\hat F_n^*(\cdot)-F_{\hat\theta^*}(\cdot)\}$ as

$$\widehat{W}_n^*(\cdot)=n^{1/2}\left\{F_n^*(\cdot)-F_{\hat\theta}(\cdot)\right\}-n^{1/2}\left\{F_{\hat\theta^*}(\cdot)-F_{\hat\theta}(\cdot)\right\}+n^{1/2}\left\{\hat F_n^*(\cdot)-F_n^*(\cdot)\right\}. \tag{A.6}$$

Because $F_n^*(x)=n^{-1}\sum_{i=1}^n I\left(\varepsilon_i^*\le x\right)$, the first term on the right hand side of (A.6) may be written as $n^{1/2}\{F_n^*(x)-F_{\hat\theta}(x)\}=n^{-1/2}\sum_{i=1}^n\left\{I(\varepsilon_i^*\le x)-F_{\hat\theta}(x)\right\}$. Next, we obtain asymptotic uniform expansions for the last two terms in (A.6).

Let $\xi_i^*(\hat\phi,\varepsilon_i^*)=\tau(\hat\phi)\lambda_i^*(\hat\phi)(1-\varepsilon_i^*)$. Then, $\hat\phi^*-\hat\phi=n^{-1}\sum_{i=1}^n\xi_i^*(\hat\phi,\varepsilon_i^*)+o_{p_n^*}\left(n^{-1/2}\right)$. By expanding $F_{\hat\theta^*}(x)$ about $\hat\theta$ and substituting an asymptotic representation for $n^{1/2}\left(\hat\theta^*-\hat\theta\right)$, one obtains that uniformly in $x\ge 0$,

$$n^{1/2}\left\{F_{\hat\theta^*}(x)-F_{\hat\theta}(x)\right\}=n^{-1/2}\sum_{i=1}^n\left\{h_{\hat\theta}^*\left(\varepsilon_i^*\right)-\mathcal{Q}_{ni}^*\left(\hat\phi,\varepsilon_i^*\right)\right\}^\top\dot F_{\hat\theta}(x)+o_{p_n^*}(1), \tag{A.7}$$

where $\mathcal{Q}_{ni}^*(\hat\phi,\varepsilon_i^*)=n^{-1}\sum_{j=1}^n h_{\hat\theta}^{*'}(\varepsilon_j^*)\varepsilon_j^*\{\lambda_j^*(\hat\phi)\}^\top\xi_i^*(\hat\phi,\varepsilon_i^*)$ and h_θ^* is as in Lemma A.7.

Substituting $n^{1/2}(\hat\phi^*-\hat\phi)=n^{-1/2}\sum_{i=1}^n\xi_i^*(\hat\phi,\varepsilon_i^*)+o_{p_n^*}(1)$ in the uniform expansion of Lemma A.9 yields that, uniformly in $x\ge 0$,

$$n^{1/2}\left\{\hat F_n^*(x)-F_n^*(x)\right\}=xf_{\hat\theta}(x)\mathbb{E}_*\left[\lambda_1^*(\hat\phi)\right]^\top n^{-1/2}\sum_{i=1}^n\xi_i^*\left(\hat\phi,\varepsilon_i^*\right)+o_{p_n^*}(1). \tag{A.8}$$

Let $G_n^*(t):=n^{-1/2}\sum_{i=1}^n g_{ni}^*(t)$, where

$$\begin{aligned}g_{ni}^*(t)=&\,I\left(\varepsilon_i^*\le F_{\hat\theta}^{-1}(t)\right)-t-\left\{h_{\hat\theta}^*\left(\varepsilon_i^*\right)-\mathcal{Q}_{ni}^*\left(\hat\phi,\varepsilon_i^*\right)\right\}^\top\dot F_{\hat\theta}\left(F_{\hat\theta}^{-1}(t)\right)\\&+F_{\hat\theta}^{-1}(t)f_{\hat\theta}\left(F_{\hat\theta}^{-1}(t)\right)\mathbb{E}_*\left[\lambda_1^*(\hat\phi)\right]^\top\xi_i^*\left(\hat\phi,\varepsilon_i^*\right).\end{aligned}$$

In view of (A.7) and (A.8), $\widehat{W}_n^*\circ F_{\hat\theta}^{-1}(t)=G_n^*(t)+o_{p_n^*}(1)$, uniformly in $t\in[0,1]$. Therefore, the weak convergence of $\widehat{W}_n^*\circ F_{\hat\theta}^{-1}(\cdot)$ follows from that of $G_n^*(\cdot)$.

Let

$$\begin{aligned}\mathcal{G}(s,t,\theta,\phi):=&\sum_{j=1}^q\sum_{l=1}^q\left[\frac{\partial F_\theta}{\partial\theta_j}\left(F_\theta^{-1}(t)\right)\frac{\partial F_\theta}{\partial\theta_l}\left(F_\theta^{-1}(s)\right)\mathbb{E}\{h_{\theta_j}(\varepsilon_1)h_{\theta_l}(\varepsilon_1)\}\right]\\&+\mathbb{E}\left\{\left[\lambda_1(\phi)\right]^\top\right\}\tau(\phi)\mathbb{E}\left\{\left[\lambda_1(\phi)\right]\left[\lambda_1(\phi)\right]^\top\right\}\tau(\phi)^\top\\&\times\mathbb{E}\left[\lambda_1(\phi)\right]\sigma_\theta^2C_\theta\left(F_\theta^{-1}(t)\right)C_\theta\left(F_\theta^{-1}(s)\right)\\&-\sum_{j=1}^q\left[\frac{\partial F_\theta}{\partial\theta_j}\left(F_\theta^{-1}(t)\right)\int_0^s h_{\theta_j}\left(F_\theta^{-1}(u)\right)du\right]\\&-\sum_{j=1}^q\left[\frac{\partial F_\theta}{\partial\theta_j}\left(F_\theta^{-1}(s)\right)\int_0^t h_{\theta_j}\left(F_\theta^{-1}(u)\right)du\right]\end{aligned}$$

$$
\begin{aligned}
&+K_\theta\left(t-\int_0^t F_\theta^{-1}(u)du\right)C_\theta\left(F_\theta^{-1}(s)\right)\\
&+K_\theta\left(s-\int_0^s F_\theta^{-1}(u)du\right)C_\theta\left(F_\theta^{-1}(t)\right)\\
&+K_\theta\sum_{j=1}^{q}\left[C_\theta\left(F_\theta^{-1}(s)\right)\frac{\partial F_\theta}{\partial\theta_j}\left(F_\theta^{-1}(t)\right)+C_\theta\left(F_\theta^{-1}(t)\right)\frac{\partial F_\theta}{\partial\theta_j}\left(F_\theta^{-1}(s)\right)\right]\\
&\times\mathbb{E}\left\{\varepsilon_1 h_{\theta_j}(\varepsilon_1)\right\},
\end{aligned}
\tag{A.9}
$$

with $\sigma_\theta^2 := \int(x-1)^2 f_\theta(x)dx$, $K_\theta := \mathbb{E}\{[\lambda_1(\phi)]^\top\}\tau(\phi)\mathbb{E}[\lambda_1(\phi)]$, $h_{\theta_j}(t) := \{-\mathbb{E}[(\partial/\partial\theta_j)\dot{g}_\theta(\varepsilon_1)]\}^{-1}\dot{g}_\theta(t)$, $j=1,\ldots,q$, and $C_\theta(y) := \sum_{j=1}^q(\partial/\partial\theta_j)F_\theta(y)\int x h'_{\theta_j}(x)f_\theta(x)dx + yf_\theta(y)$.

It may be verified that $\mathrm{cov}^*\{G_n^*(s),G_n^*(t)\} := \min\{s,t\}-st+\mathcal{G}(s,t,\theta_0,\phi_0)+o_p(1)$. Hence, $\mathrm{cov}^*\{G_n^*(s),G_n^*(t)\}=\mathrm{cov}\{\mathcal{V}(s),\mathcal{V}(t)\}+o_p(1)$. It follows from a martingale CLT that the finite dimensional distributions of $G_n^*(\cdot)$ converge in probability to those of $\mathcal{V}(\cdot)$ as $n\to\infty$. By Markov's inequality and using Conditions (C3), (C5), (E2), and (E3), one also obtains that $n^{-1/2}\sum g_{ni}(t)$ is asymptotically stochastically equicontinuous. Therefore, $G_n^*(\cdot)$ converges weakly to $\mathcal{V}(\cdot)$ [in probability]. ■

Lemma A.10 and the continuous mapping theorem imply that the bootstrap tests based on $\widehat{W}_n^*$ are asymptotically valid. However, we are interested in the truncated process $\widehat{W}_n^{*(m)}$. We wish to show that the error resulting from substituting ${}^{*(m)}$ for * becomes negligible for large n, more specifically, $\sup_{y\geq 0}|\widehat{W}_n^*(y)-\widehat{W}_n^{*(m)}(y)| = o_{p_n^*}(1)$. To this end, it suffices to show that the differences $(\hat{F}_n^{*(m)}-\hat{F}_n^*)$ and $(F_{\hat{\theta}^{*(m)}}-F_{\hat{\theta}^*})$ converge to zero uniformly, at a rate faster than $n^{-1/2}$. These two sufficient conditions are obtained in the next lemma; proof is in the online Supplementary Material.

LEMMA A.11. *The following hold in probability:*
(a) $\sup_{y\geq 0} n^{1/2}\left|\hat{F}_n^{*(m)}(y)-\hat{F}_n^*(y)\right| = o_{p_n^*}(1)$, *(b)* $\sup_{y\geq 0} n^{1/2}\left|F_{\hat{\theta}^{*(m)}}(y)-F_{\hat{\theta}^*}(y)\right| = o_{p_n^*}(1)$.

A.3. Proofs of the main results stated in Sections 2 and 3

LEMMA A.12. *Suppose that Conditions (C1)–(C5) are satisfied. Then, under the null hypothesis,* $\sup_{x\geq 0}\left|\widetilde{F}_n(x)-F_n(x)-(\hat{\phi}-\phi_0)^\top\mathbb{E}\{\lambda_1(\phi_0)\}xf_{\theta_0}(x)\right| = o_p\left(n^{-1/2}\right)$.

Proof. A simplified version of the proof of Lemma A.9 with $(\hat{\phi},\hat{\theta})=(\phi_0,\theta_0)$ yields that

$$
\sup_{x\geq 0}\left|\left\{\hat{F}_n(x)-F_n(x)\right\}-\left(\hat{\phi}-\phi_0\right)^\top\mathbb{E}\{\lambda_1(\phi_0)\}xf_{\theta_0}(x)\right| = o_p\left(n^{-1/2}\right). \tag{A.10}
$$

By arguing as in the proof of Lemma A.11 with $\left(\hat{\phi},\hat{\theta}\right)=(\phi_0,\theta_0)$, one also obtains that $\sup_{x\geq 0} n^{1/2}\left|\widetilde{F}_n(x)-\hat{F}_n(x)\right| = o_p(1)$. Hence, the statement in (A.10) continues to hold when $\hat{F}_n$ is replaced by $\widetilde{F}_n$. ■

Proof of Theorem 1. Partition $\widetilde{W}_n(x)=n^{1/2}\{\widetilde{F}_n(x)-F_{\hat{\theta}}(x)\}$ as

$$
\widetilde{W}_n(\cdot)=n^{1/2}\left\{F_n(\cdot)-F_{\theta_0}(\cdot)\right\}-n^{1/2}\left\{F_{\hat{\theta}}(\cdot)-F_{\theta_0}(\cdot)\right\}+n^{1/2}\left\{\widetilde{F}_n(\cdot)-F_n(\cdot)\right\}. \tag{A.11}
$$

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

Since $F_n(x) = n^{-1}\sum_{i=1}^{n} I(\varepsilon_i \le x)$, the first term on the right hand side of (A.11) is $n^{1/2}\{F_n(x) - F_{\theta_0}(x)\} = n^{-1/2}\sum_{i=1}^{n}\{I(\varepsilon_i \le x) - F_{\theta_0}(x)\}$. By expanding $F_{\hat\theta}(x)$ about θ_0 and substituting an asymptotic representation for $n^{1/2}(\hat\theta - \theta_0)$, we have, uniformly in $x \ge 0$,

$$n^{1/2}\left\{F_{\hat\theta}(x) - F_{\theta_0}(x)\right\} = n^{-1/2}\sum_{i=1}^{n}\left\{h_{\theta_0}(\varepsilon_i) - \mathcal{Q}_i(\phi_0, \varepsilon_i)\right\}^{\top}\dot F_{\theta_0}(x) + o_p(1),$$

where $\mathcal{Q}_i(\phi_0, \varepsilon_i) = n^{-1}\sum_{j=1}^{n} h'_{\theta_0}(\varepsilon_j)\varepsilon_j\{\lambda_j(\phi_0)\}^{\top}\{\tau(\phi_0)\lambda_i(\phi_0)(1-\varepsilon_i)\}$.

Next, consider the last term $n^{1/2}\{\widetilde F_n(\cdot) - F_n(\cdot)\}$ in (A.11). By applying Lemma A.12 and using an asymptotic representation for $n^{1/2}(\hat\phi - \phi_0)$, similar to the one used for $n^{1/2}(\hat\phi^* - \hat\phi)$ in the proof of Lemma A.10, one obtains that uniformly in $x \ge 0$,

$$n^{1/2}\{\widetilde F_n(x) - F_n(x)\} = x f_{\theta_0}(x)\mathbb{E}[\lambda_1(\phi_0)]^{\top} n^{-1/2}\sum_{i=1}^{n}\{\tau(\phi_0)\lambda_i(\phi_0)(1-\varepsilon_i)\} + o_p(1).$$

By using these asymptotic expansions and mimicking the arguments in the proof of Lemma A.10 for the special case $(\hat\phi, \hat\theta) = (\phi_0, \theta_0)$, we obtain that the process $\widetilde W_n(\cdot)$ converges weakly to the centered Gaussian process $G(\cdot)$ with covariance kernel $\operatorname{cov}\{G(s), G(t)\} := \min\{s,t\} - st + \mathcal{G}(s,t,\theta_0,\phi_0)$, where $\mathcal{G}(s,t,\theta,\phi)$ is as in (A.9). This completes the proof of part (a).

Let us illustrate part (b) for the *CvM* statistic $T_3 = \int \widetilde W_n^2(x)dF_{\hat\theta}(x)$. Write

$$T_3 = n\int\left\{\widetilde F_n(x) - F_{\hat\theta}(x)\right\}^2 dF_{\theta_0}(x) + n\int\left\{\widetilde F_n(x) - F_{\hat\theta}(x)\right\}^2\left(f_{\hat\theta}(x) - f_{\theta_0}(x)\right)dx.$$

The term $n\left|\int\left\{\widetilde F_n(x) - F_{\hat\theta}(x)\right\}^2(f_{\hat\theta}(x) - f_{\theta_0}(x))dx\right|$ is bounded from above by

$$\left(\sup_x\left|\sqrt{n}\left\{\widetilde F_n(x) - F_{\hat\theta}(x)\right\}\right|\right)^2\int\left|f_{\hat\theta}(x) - f_{\theta_0}(x)\right|dx.$$

One obtains by part (a) and the continuous mapping theorem that $(\sup_x|\sqrt{n}\{\widetilde F_n(x) - F_{\hat\theta}(x)\}|)^2 = O_p(1)$. Let B be an open neighbourhood of θ_0 as in Condition (C3). Then, by Condition (C3) and applying the mean value theorem, for every $\theta \in B$,

$$\int\left|f_\theta(x) - f_{\theta_0}(x)\right|dx \le \sum_{j=1}^{q}\left|\theta_j - \theta_{0j}\right|\int K_{\bar\theta}(x)dx \tag{A.12}$$

for some $\bar\theta \in B$ satisfying $\|\bar\theta - \theta_0\| \le \|\theta - \theta_0\|$, with the function K_θ as in (C3). Since $n^{1/2}(\hat\theta - \theta_0) = O_p(1)$ and $\sup_{\theta\in B}\int K_\theta(y)dy < \infty$ by (C3), one obtains by (A.12) that $\int|f_{\hat\theta}(x) - f_{\theta_0}(x)|dx = o_p(1)$. Hence, $n\int\{\widetilde F_n(x) - F_{\hat\theta}(x)\}^2(f_{\hat\theta}(x) - f_{\theta_0}(x))dx = o_p(1)$. Therefore, we have $T_3 = \mathfrak{h}_3(\widetilde W_n \circ F_{\theta_0}^{-1}) + o_p(1)$ with $\mathfrak{h}_3(a) := \int_0^1\{a(t)\}^2dt$ for $a \in D[0,1]$. Using similar arguments, one also obtains that part (b) holds for $j = 1, 2, 4$, and 5. Part (c) follows from parts (a), (b) and the continuous mapping theorem. ∎

The relevance of the next lemma was discussed in the last paragraph of Section 3.

LEMMA A.13. *Suppose that Conditions (C1)–(C6) and Assumptions (E1)–(E3) are satisfied. Then,* $\sup_{y\ge 0} n^{1/2}\left|\widetilde F_n^{*(m)}(y) - \hat F_n^{*(m)}(y)\right| = o_{p_n^*}(1)$*, in probability.*

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

Proof. In the following arguments, for brevity, a statement such as '$X_n^* = o_{p_n^*}(1)$, in probability', is abbreviated to '$X_n^* = o_{p_n^*}(1)$'. Let

$$\widetilde{\upsilon}_{ni}^{*(m)} = n^{1/2}\left[\widetilde{\Psi}_i^{*(m)}\left(\hat{\phi}^{*(m)}\right) - \Psi_i^{*(m)}(\hat{\phi})\right] / \Psi_i^{*(m)}(\hat{\phi}).$$

By (E3) and Conditions (C1) and (C6), for some open neighbourhood B of ϕ_0, $\sup_{\phi \in B} \left\| \dot{\widetilde{\Psi}}_i^{*(m)}(\phi) - \dot{\Psi}_i^{*(m)}(\phi) \right\| \overset{e.a.s.}{\to} 0$ and $\sup_{\phi \in \Phi} \left| \widetilde{\Psi}_i^{*(m)}(\phi) - \Psi_i^{*(m)}(\phi) \right| \overset{e.a.s.}{\to} 0$, as $i \to \infty$. Hence, Lemma 2.1 of Straumann and Mikosch (2006) imply that

$$\sum_{i=1}^{n} \left| \widetilde{\Psi}_i^{*(m)}\left(\hat{\phi}^{*(m)}\right) - \widetilde{\Psi}_i^{*(m)}(\hat{\phi}) \right| / \Psi_i^{*(m)}(\hat{\phi}) = O_{p_n^*}(1).$$

Therefore, by a one-term Taylor expansion, $\widetilde{\upsilon}_{ni}^{*(m)} = n^{1/2}(\hat{\phi}^{*(m)} - \hat{\phi})^\top \lambda_i^{*(m)}(\hat{\phi}) + \widetilde{r}_{ni}^{*(m)}$, for some random array $\{\widetilde{r}_{ni}^{*(m)}\}$ satisfying $n^{-1}\sum_{i=1}^n \widetilde{r}_{ni}^{*(m)} = O_{p_n^*}(n^{-1/2})$.

Let $\upsilon_{ni}^{*(m)} = n^{1/2}[\Psi_i^{*(m)}(\hat{\phi}^{*(m)}) - \Psi_i^{*(m)}(\hat{\phi})]/\Psi_i^{*(m)}(\hat{\phi})$. From the proof of Lemma A.11 we have $\max_{1 \le i \le n} |n^{-1/2}\upsilon_{ni}^{*(m)}| = o_{p_n^*}(1)$. By Lemma A.8, $n^{1/2}(\hat{\phi}^{*(m)} - \hat{\phi}) = O_{p_n^*}(1)$. Hence,

$$\max_{1 \le i \le n} \left| n^{-1/2}\widetilde{\upsilon}_{ni}^{*(m)} \right| \le n^{-1/2} \max_{1 \le i \le n} \left| \widetilde{\upsilon}_{ni}^{*(m)} - \upsilon_{ni}^{*(m)} \right| + \max_{1 \le i \le n} \left| n^{-1/2}\upsilon_{ni}^{*(m)} \right| = o_{p_n^*}(1).$$

Therefore, by Lemma A.5 with $\gamma_{ni} = 1$ and $\rho_{ni} = n^{-1/2}\widetilde{\upsilon}_{ni}^{*(m)}$, we have the following, uniformly in $y \ge 0$:

$$\begin{aligned} n^{1/2}\widetilde{F}_n^{*(m)}(y) &= n^{-1/2}\sum_{i=1}^n I\left(\widetilde{\varepsilon}_i^{*(m)} \le y\right) = n^{-1/2}\sum_{i=1}^n I\left\{\varepsilon_i^* \le y + yn^{-1/2}\widetilde{\upsilon}_{ni}^{*(m)}\right\} \\ &= n^{1/2}F_n^{*(m)}(y) + n^{-1/2}\sum_{i=1}^n \left\{F_{\hat{\theta}}\left(y + yn^{-1/2}\widetilde{\upsilon}_{ni}^{*(m)}\right) - F_{\hat{\theta}}(y)\right\} + o_{p_n^*}(1) \\ &= n^{1/2}F_n^{*(m)}(y) + yf_{\hat{\theta}}(y)n^{1/2}\left(\hat{\phi}^{*(m)} - \hat{\phi}\right)^\top n^{-1}\sum_{i=1}^n \widetilde{\lambda}_i^{*(m)}(\hat{\phi}) \\ &\quad + o_{p_n^*}(1). \end{aligned} \qquad \textbf{(A.13)}$$

From the proof of Lemma A.11 we have that

$$\sup_{y \ge 0} \left| \hat{F}_n^{*(m)}(y) - F_n^{*(m)}(y) - yf_{\hat{\theta}}(y)\left(\hat{\phi}^{*(m)} - \hat{\phi}\right)^\top n^{-1}\sum_{i=1}^n \lambda_i^{*(m)}\left(\hat{\phi}\right) \right| = o_{p_n^*}\left(n^{-1/2}\right). \qquad \textbf{(A.14)}$$

Since $\sup_{\phi \in B} \left\|\widetilde{\lambda}_i^{*(m)}(\phi) - \lambda_i^{*(m)}(\phi)\right\| \overset{e.a.s.}{\to} 0$ and $\sup_{\theta \in B, y \ge 0}(1+y)f_\theta(y) < \infty$ for some open neighbourhood B of θ_0, it follows from (A.13), (A.14) and Lemma 2.1 of Straumann and Mikosch (2006) that $\sup_{y \ge 0} n^{1/2}|\widetilde{F}_n^{*(m)}(y) - \hat{F}_n^{*(m)}(y)| = o_{p_n^*}(1)$. ■

https://doi.org/10.1017/S026646661500047X Published online by Cambridge University Press

Proof of Theorem 2. Part (a) follows from Lemmas A.10, A.11, and A.13. Since $T_j = \mathfrak{h}_j(\widetilde{W}_n^{*(m)} \circ F_{\hat{\theta}}^{-1}) + o_{p_n^*}(1)$, in probability ($j = 1, \cdots, 5$), part (b) follows from part (a) and the continuous mapping theorem. ■

The next lemma is used for deriving the results on local power stated in Section 4.

LEMMA A.14. *Suppose that Conditions (C1)−(C5) and Assumptions (E1) and (E2) are satisfied. Then, under* H_{an}, $\widetilde{W}_n \circ F_{\theta_0}^{-1}(\cdot)$ *converges weakly to the Gaussian process* $W_a(\cdot)$ *given by* $W_a(\cdot) = m_a(\cdot) + G(\cdot)$, *with* $G(\cdot)$ *is as in Theorem 1, and*

$$m_a(t) := \delta\left\{\tilde{F}\left[F_{\theta_0}^{-1}(t)\right] - t - \mathcal{R}^\top \dot{F}_{\theta_0}\left[F_{\theta_0}^{-1}(t)\right]\right\}, \quad t \in [0,1], \tag{A.15}$$

where $\mathcal{R} := \left[\int \{\ddot{g}_{\theta_0}(y)\}^{-1}\, dF_{\theta_0}(y)\right]\left[\int \dot{g}_{\theta_0}(\varepsilon)\, d\tilde{F}(\varepsilon)\right]$. *Further, if (C6) and (E3) are also satisfied, then conditional on* $\{Z_1, \cdots, Z_n\}$, *under* H_{an}, $\widetilde{W}_n^{*(m)} \circ F_{\hat{\theta}}^{-1}(\cdot)$ *converges weakly to* $G(\cdot)$ *[in probability].*

Proof. See the online Supplementary Material. ■

Proof of Proposition 1. Because $T_j = \mathfrak{h}_j\left(\widetilde{W}_n \circ F_{\theta_0}^{-1}\right) + o_p(1)$ and $T_j^{*(m)} = \mathfrak{h}_j\left(\widetilde{W}_n^{*(m)} \circ F_{\hat{\theta}}^{-1}\right) + o_{p^*}(1)$ [in probability], the proof follows from Lemma A.14 and the continuous mapping theorem. ■

Proofs of Propositions 2 and 3 are given in the online Supplementary Material.