Econometric Theory, **28**, 2012, 1283–1312.
doi:10.1017/S0266466612000102

LACK-OF-FIT TESTING OF THE CONDITIONAL MEAN FUNCTION IN A CLASS OF MARKOV MULTIPLICATIVE ERROR MODELS

Hira L. Koul
Michigan State University

Indeewara Perera
Monash University

Mervyn J. Silvapulle
Monash University

The family of multiplicative error models, introduced by Engle (2002, *Journal of Applied Econometrics* 17, 425–446), has attracted considerable attention in recent literature for modeling positive random variables, such as the duration between trades at a stock exchange, volume transactions, and squared log returns. Such models are also applicable to other positive variables such as waiting time in a queue, daily/hourly rainfall, and demand for electricity. This paper develops a new method for testing the lack-of-fit of a given parametric multiplicative error model having a Markov structure. The test statistic is of Kolmogorov–Smirnov type based on a particular martingale transformation of a marked empirical process. The test is asymptotically distribution free, is consistent against a large class of fixed alternatives, and has nontrivial asymptotic power against a class of nonparametric local alternatives converging to the null hypothesis at the rate of $O(n^{-1/2})$. In a simulation study, the test performed better overall than the general purpose Ljung–Box Q-test, a Lagrange multiplier type test, and a generalized moment test. We illustrate the testing procedure by considering two data examples.

1. INTRODUCTION

Nonnegative random variables are frequently encountered in economics and finance and also in other fields of social and natural sciences. Examples include financial durations such as the duration between consecutive trades at a stock exchange, waiting time in a queue, daily/hourly rainfall, and demand for electricity.

We thank two anonymous referees and the co-editor Oliver Linton for their constructive comments and suggestions that led to substantial improvements in this paper. Hira Koul was partially supported by NSF DMS grant 0704130, and Mervyn Silvapulle and Indeewara Perera were partially supported by Australian Research Council Discovery grant DP0665710. We thank Monash e-Science and Grid Engineering Lab for assistance with some computational aspects and use of the computing facilities. Address correspondence to Indeewara Perera, Department of Econometrics and Business Statistics, Monash University, P.O. Box 197, Caulfield East, Australia 3145; e-mail: indeewara.perera@monash.edu.

As a general framework for modeling such nonnegative processes, Engle (2002) introduced a family of multiplicative error models.

To introduce this family of models, let y_i, $i \in \mathbb{Z} := \{0, \pm 1, \pm 2, \ldots\}$, be a discrete time process defined on $[0, \infty)$ and let $\mathcal{H}_{i-1}$ denote the information available for forecasting y_i. A multiplicative error model takes the form

$$y_i = \mathbb{E}[y_i \mid \mathcal{H}_{i-1}]\varepsilon_i, \qquad \varepsilon_i \mid \mathcal{H}_{i-1} \sim D^+(1, \sigma^2), \tag{1}$$

where $D^+(1, \sigma^2)$ is a probability distribution on the positive real line with unit mean and a finite variance σ^2. The error sequence ε_i, $i \in \mathbb{Z}$, is often assumed to be independent and identically distributed (i.i.d.).

This family of models has attracted considerable attention in recent literature. Multiplicative error specifications have been adopted for modeling durations (Engle and Russell, 1998), volume transactions (Manganelli, 2005), high–low price range (Chou, 2005), squared log returns (Engle and Gallo, 2006), equity market volatilities (Engle, Gallo, and Velucchi, 2012), and realized volatility (Lanne, 2006; Engle and Gallo, 2006), among many others.

The parametric form of $\mathbb{E}[y_i \mid \mathcal{H}_{i-1}]$ is the main component of interest in a multiplicative error model (MEM), and hence testing for the specification of this component has an important place in the process of model evaluation. However, a large proportion of MEMs have a complicated probabilistic structure (see Pacurar, 2008; Brownlees, Cipollini, and Gallo, 2011). Consequently, assessing the goodness-of-fit of such parametric specifications is a nontrivial task. A small number of general methods have been developed for testing the goodness-of-fit of a given parametric form for $\mathbb{E}[y_i \mid \mathcal{H}_{i-1}]$. We will shortly describe some of these tests. One aspect of these general methods is that they are designed for models with more dynamic parametric forms. For example, they would typically allow $\mathbb{E}[y_i \mid \mathcal{H}_{i-1}]$ to depend on the whole infinite past, $\{y_{t-1} : t \le i\}$. However, in empirical studies, one would want to consider such elaborate models only if a simpler one does not fit. This paper develops a formal test for this purpose. More specifically, a method is developed for testing the lack-of-fit of a given MEM having the Markov specification

$$\mathbb{E}[y_i \mid \mathcal{H}_{i-1}] = \tau(y_{i-1}), \tag{2}$$

where $\tau(\cdot)$ is a positive function defined on $\mathbb{R}^+ := [0, \infty)$. We shall call the model (1)–(2) a Markov MEM.

In what follows, we describe some of the existing methods of testing for MEMs. A common practice for evaluating a MEM is to carry out simple diagnostic tests to examine the dynamical and distributional properties of the estimated residuals; for example, see Jasiak (1998), Giot (2000), Ghysels, Gouriéroux, and Jasiak (2004), Bauwens and Veredas (2004), Luca and Gallo (2004), and Bauwens (2006). The Box–Pierce–Ljung type portmanteau test statistic applied to the estimated residuals or squared estimated residuals is among the most frequently used tests in the literature. This approach was employed for the class of multiplicative error models known as autoregressive conditional duration (ACD) by Engle and Russell (1998) and has frequently been employed in subsequent studies. Another method that was introduced by Engle and Russell (1998) is to test for no excess dispersion of the estimated residuals, paying particular attention to checking the first and second moments of the residuals when the error distribution is assumed to be either exponential or Weibull. These diagnostic tests may suggest that the MEM is misspecified as a result of a misspecification of the conditional mean and/or of the error distribution.

Several tests that exclusively test for misspecifications of the conditional mean function have also been proposed recently. Meitz and Teräsvirta (2006) developed Lagrange multiplier (LM) type tests, focusing on testing the specifications of the functional form of the conditional mean against various forms of parametric alternatives (e.g., tests against higher order models, tests of linearity, and tests of parameter constancy). Hautsch, (2006) also considered some LM tests. One aspect of such LM type tests is that they require the alternative hypothesis to specify a larger finite-dimensional parametric model, and so the null hypothesis is obtained by setting some components of the model parameter equal to zero. Thus, LM type tests are targeted to detect departures in the direction of the particular finite-dimensional parametric model in the alternative hypothesis.

Recently, Hong and Lee (2011) developed a class of tests using a generalized spectral derivative approach for testing the specification of the conditional mean, without assuming the knowledge of a parametric form under the alternative. These tests are based on kernel density estimators, and hence their finite-sample performances are subject to the choice of the kernal and the bandwidth. Chen and Hsieh (2010) proposed a set of generalized moment tests, unifying several existing parametric tests for the conditional mean. In addition, the general diagnostic test considered by Hong and Lee (2003) can also be used as a misspecification test for MEMs (see Meitz and Teräsvirta, 2006).

There are also methods available for testing the goodness-of-fit of a given parametric form for the distribution of the error term ε_i in (1) (see, e.g., Fernandes and Grammig, 2005; Chen and Hsieh, 2010). However, these methods are not applicable for testing the adequacy of a given parametric form for the conditional mean function, which is the topic of this paper.

This paper develops a new test for the conditional mean specification of a Markov MEM. Most of the existing tests that we mentioned earlier are developed for general models that are not necessarily Markov. Thus, by design, they do not make use of the simple dynamic structure of the Markov model (2). By contrast, the test proposed in this paper is designed specifically to exploit the special structure of the model (2). Therefore, there is some ground to conjecture that, for the case of testing the lack-of-fit of a given Markov MEM, the test proposed in this paper would be more suitable than the existing ones. In a simulation study, when testing for the conditional mean specification of a Markov model, the new test performed significantly better than a LM test, a generalized moment test, and a Box–Pierce–Ljung type portmanteau test. We expect that the new test complements these tests and spectral derivative–based tests for MEMs in a desirable fashion.

The test is introduced in Section 2. It is based on a marked empirical process of residuals, analogous to the ones in Stute, Thies, and Zhu (1998) and Koul and Stute (1999). The main result of Section 2 indicates that the asymptotic null distribution of the test statistic is that of the supremum of the standard Brownian motion on [0, 1]. Therefore, the test is asymptotically distribution free, and a set of asymptotic critical values is available for general use. Consistency against a fixed alternative and the asymptotic power against a sequence of local nonparametric alternatives are discussed in Section 3. Section 4 contains a simulation study. Two illustrative examples are discussed in Section 5. Section 6 concludes the paper. The proofs are relegated to an Appendix.

2. THE TEST STATISTIC AND ITS ASYMPTOTIC NULL DISTRIBUTION

This section provides an informal motivation for the test, defines the test statistic, and states its asymptotic null distribution. First, Section 2.1 provides a motivation for the test and a brief indication of the approach adopted in constructing the test statistic. Then, Section 2.2 introduces the regularity conditions, defines the test statistic, and states the main result on its asymptotic null distribution.

To introduce the null and alternative hypotheses of interest, let q be a known positive integer, $\Theta \subseteq \mathbb{R}^q$, and let $\mathcal{M} = \{\Psi(y,\theta) : y \geq 0, \theta \in \Theta\}$ be a parametric family of positive functions. We wish to test

$$H_0 : \tau(y) = \Psi(y,\theta), \quad \text{for some } \theta \in \Theta \quad \text{and} \quad \forall\, y \geq 0, \quad \text{vs.} \quad H_1 : \text{Not } H_0. \tag{3}$$

Let $\{Y_0, Y_1, \ldots, Y_n\}$ be observations of a positive, strictly stationary, and ergodic process $\{Y_i\}$ that obeys the model (1)–(2). Let G denote the stationary distribution function of Y_0. Let τ, Ψ, σ^2, and the testing problem be as in (1)–(2) and (3). Let θ denote the true parameter value under H_0 and ϑ denote an arbitrary point in Θ. Under H_0, G may depend on θ, but we do not exhibit this dependence.

2.1. Motivation for the Test Statistic

This section provides a motivation for the test and an overview of the general approach. The regularity conditions are not discussed here; they will be provided in Section 2.2. Let $T(y,\vartheta) = \int_0^y [\{\tau(x)/\Psi(x;\vartheta)\} - 1]dG(x)$ and

$$\mathcal{U}_n(y,\vartheta) = n^{-1/2} \sum_{i=1}^{n} \left\{ \frac{Y_i}{\Psi(Y_{i-1},\vartheta)} - 1 \right\} I(Y_{i-1} \leq y), \qquad y \geq 0, \vartheta \in \Theta. \tag{4}$$

First, consider the special case when the true parameter value θ in H_0 is given. Arguing analogously as in Stute et al. (1998), because θ is known, the integral transform $T(\cdot,\theta)$ is uniquely determined by $\tau(\cdot)$, assuming G is known. Therefore, inference about the functional form of $\tau(\cdot)$ could be based on an estimator of $T(\cdot,\theta)$. From (2) it follows that under H_0, $T(y,\theta) = \mathbb{E}I(Y_0 \le y)[\{Y_1/\Psi(Y_0,\theta)\} - 1] = 0$, for all $y \ge 0$. Further, an unbiased estimator of $T(y,\theta)$ is given by $n^{-1/2}$ $\mathcal{U}_n(y,\theta)$. It is shown later that, under H_0, $\mathcal{U}_n(y,\theta)$ converges weakly to $W \circ G$, where W is the standard Brownian motion. Therefore, a Kolmogorov–Smirnov type test could be based on $\sup_y |\mathcal{U}_n(y,\theta)|$, which converges weakly to $\sup_{0\le t\le 1} |W(t)|$, under H_0.

Now, consider the testing problem (3), where H_0 specifies a parametric family $\Psi(\cdot,\theta)$ for $\tau(\cdot)$, for some unknown θ. Let $\widehat{\theta}$ be a $n^{1/2}$-consistent estimator of θ. Then an estimator of $T(y,\theta)$ is $n^{-1/2}\mathcal{U}_n(y,\widehat{\theta})$. The limiting null distribution of $\mathcal{U}_n(y,\widehat{\theta})$ depends on $\widehat{\theta}$ and the unknown parameter θ in a complicated fashion. Therefore, the method outlined in the previous paragraph for known θ is no longer applicable, and it does not lead to an asymptotically distribution free test. Constructing such a test based on $\mathcal{U}_n(y,\widehat{\theta})$ is the focus of Section 2.2.

The process $\mathcal{U}_n(y,\widehat{\theta})$ is an extension of the so-called cumulative sum process for the one sample setting to the current setup. The use of cumulative sum process for testing the lack-of-fit of a given regression function goes back to von Neumann (1941), who proposed a test of constant regression based on an analogue of this process. More recently, its analogues have been used by several authors to propose asymptotically distribution free (ADF) lack-of-fit tests in some other models.

Stute, Xu, and Zhu (2008) and Escanciano (2010) proposed residual empirical processes marked by appropriately chosen weighting functions, for additive regression and certain time series models, respectively. The testing procedure proposed in Escanciano (2010) could also be viewed as a generalization of the method developed in Wooldridge (1990) for obtaining ADF tests. This approach is not directly applicable for testing H_0 in (3) using $\mathcal{U}_n(y,\widehat{\theta})$, because $\mathcal{U}_n(y,\widehat{\theta})$ is a marked empirical process with the marks that are the centered residuals $(Y_i/\Psi(Y_{i-1},\widehat{\theta}) - 1)$ and not functions of a conditioning set containing past observations as in Escanciano (2010).

Escanciano and Mayoral (2010) and Ling and Tong (2011) have also developed ADF tests based on marked empirical processes similar to $\mathcal{U}_n(y,\widehat{\theta})$. However, the hypotheses considered in these papers are different from (3), and hence these tests are not applicable in the present context.

In this paper, we appeal to a particular martingale transformation method to construct an ADF test for H_0 in (3) based on $\mathcal{U}_n(y,\widehat{\theta})$. Such martingale transformation methods have been successfully applied to location, regression, and certain autoregressive models. More specifically, tests have been developed when the null hypothesis specifies a parametric family for the mean function of a regression and/or an autoregressive model, and the conditional variance function in a regression model; see, for example, Stute et al. (1998), Koul and Stute (1999), Khmaladze and Koul (2004), Dette and Hetzler (2009), and Koul and Song (2010). A common feature of all these studies is that they are all for additive models.

The Markov multiplicative time series models studied in this paper are structurally different (see Engle, 2002; Brownlees et al., 2011).

2.2. The Test and the Main Results

Let F denote the cumulative distribution function (cdf) of ε_1. In what follows, $\|a\|$ denotes euclidean norm of a vector $a \in \mathbb{R}^q$, and for a $q \times q$ real matrix D, $\|D\| := \sup\{\|a^T D\|;\ a \in \mathbb{R}^q,\ \|a\| = 1\}$. Now we shall introduce a set of regularity conditions.

(C1) The cdf G is continuous, $G(y) > 0$ for $y > 0$, and $\mathbb{E}Y_0^4 < \infty$. The sequence of random variables $\{\varepsilon_i\}$ is positive and i.i.d. with $\mathbb{E}(\varepsilon_1) = 1$, $0 < \sigma^2 < \infty$, and ε_i is stochastically independent of $\{Y_{j-1},\ j \leq i\}$.

(C2) The cdf F of ε_1 has a bounded Lebesgue density f.

(C3)

(a) $\Psi(y,\vartheta)$ is bounded away from zero, uniformly over $y \in \mathbb{R}^+$ and $\vartheta \in \Theta$.

(b) The true parameter value θ is in the interior of Θ, and $\int_0^\infty |\Psi(y,\theta)|^2\, dG(y) < \infty$. For all y, $\Psi(y,\vartheta)$ is continuously differentiable with respect to ϑ in the interior of Θ.

For $\vartheta \in \Theta$ and $y \geq 0$, let $\dot{\Psi}(y,\vartheta) = \Big[(\partial/\partial\vartheta_1)\Psi(y,\vartheta), \ldots, (\partial/\partial\vartheta_q)\Psi(y,\vartheta)\Big]^T$,

$$g(y,\vartheta) = \dot{\Psi}(y,\vartheta)/\Psi(y,\vartheta) \quad \text{and} \quad C(y,\vartheta) = \int_{z \geq y} g(z,\vartheta) g(z,\vartheta)^T\, dG(z).$$

(C4) For every $0 < K < \infty$,

$$\sup_{1 \leq i \leq n,\, \sqrt{n}\|\vartheta - \theta\| \leq K} \sqrt{n}|\Psi(Y_{i-1},\vartheta) - \Psi(Y_{i-1},\theta) - (\vartheta - \theta)'\dot{\Psi}(Y_{i-1},\theta)| = o_p(1).$$

(C5) There exist a $q \times q$ square matrix $\dot{g}(y,\theta)$ and a nonnegative function $h(y,\theta)$, both measurable in the y-coordinate, and satisfying the following conditions: $\forall \delta > 0, \exists \eta > 0$ such that $\|\vartheta - \theta\| \leq \eta$ implies

$$\|g(y,\vartheta) - g(y,\theta) - \dot{g}(y,\theta)(\vartheta - \theta)\| \leq \delta h(y,\theta)\|\vartheta - \theta\|, \quad \forall y \geq 0,$$

$$\mathbb{E}h^2(Y_0,\theta) < \infty, \qquad \mathbb{E}\|\dot{g}(Y_0,\theta)\|\|g(Y_0,\theta)\|^j < \infty, \qquad j = 0, 1.$$

(C6) $\int_0^\infty \|g(y,\theta)\|^2\, dG(y) < \infty$.

(C7) $C(y,\theta)$ is a positive definite matrix for all $y \in [0,\infty)$.

(C8) $\|g^T(\cdot,\theta)C^{-1}(\cdot,\theta)\|$ is bounded on bounded intervals.

(C9) $\int \|g^T(y,\theta)C^{-1}(y,\theta)\|\, dG(y) < \infty$.

(C10) There exists an estimator $\widehat{\theta}_n$ of θ satisfying $n^{1/2}\|\widehat{\theta}_n - \theta\| = O_p(1)$.

An example of $\widehat{\theta}_n$ satisfying condition (C10) is the quasi maximum likelihood (QML) estimator of θ given by

$$\widehat{\theta}_n = \arg\min_{\vartheta\in\Theta} Q_n(\vartheta), \quad \text{where } Q_n(\vartheta) = n^{-1}\sum_{i=1}^{n}\left\{\frac{Y_i}{\Psi(Y_{i-1},\vartheta)} + \ln\Psi(Y_{i-1},\vartheta)\right\}. \tag{5}$$

Conditions (C2), (C8), and (C9) are needed to ensure tightness of some sequences of stochastic processes appearing in the proofs. Conditions (C3)–(C6) are concerned with the smoothness of the parametric model being fitted to the conditional mean function. Similar conditions have been assumed for regression and autoregressive models in Stute et al. (1998) and Koul and Stute (1999). It can be shown that the conditions (C3)–(C9) are satisfied by linear Markov models of the form $\Psi(y,\vartheta) = \vartheta_1 + \vartheta_2 y,\ \ \vartheta_1,\vartheta_2 > 0$. The condition (C2) on the error distribution is satisfied by many distributions that are continuous and have positive supports, including exponential, Weibull, gamma, generalized gamma, Burr, and many others.

Now, let $\widehat{\mathcal{U}}_n(y) := \mathcal{U}_n(y,\widehat{\theta}_n)$, $\widehat{g}(y) := g(y,\widehat{\theta}_n)$, $G_n(y) := n^{-1}\sum_{i=1}^{n} I(Y_{i-1} \le y)$, and $\widehat{C}_y := \int_{x\ge y}\widehat{g}(x)\widehat{g}(x)^T\ dG_n(x)$. The proposed test is to be based on the following transform of the $\widehat{\mathcal{U}}_n$:

$$\widehat{\mathcal{W}}_n(y) := \widehat{\mathcal{U}}_n(y) - \int_0^y [\widehat{g}(x)]^T\widehat{C}_x^{-1}\int_{z\ge x}\widehat{g}(z)d\widehat{\mathcal{U}}_n(z)\,dG_n(x). \tag{6}$$

This transformation is similar to the Stute–Thies–Zhu transform of Stute et al. (1998), which in turn has its roots in the work of Khmaladze (1981).

The next theorem provides the required weak convergence result. It implies that the transform (6) of $\mathcal{U}(\cdot,\widehat{\theta}_n)$, under H_0, is asymptotically nuisance parameter free. Recall that the weak convergence in $D[0,\infty)$ is the weak convergence in $D[0,y]$, for every $0 \le y < \infty$ (see Stone, 1963). Here and in what follows, the symbol $\Longrightarrow$ denotes weak convergence and W is the standard Brownian motion.

THEOREM 1. *Suppose that (1)–(2), (C1)–(C10), and H_0 hold. Further, suppose that, for some $\beta > 0$, $\gamma > 0$, we have that*

$$\begin{aligned}
&(a) \quad \mathbb{E}\|g(Y_0,\theta)\|^4 < \infty,\\
&(b) \quad \mathbb{E}\{\|g(Y_0,\theta)\|^4|Y_0|^{1+\beta}\} < \infty,\\
&(c) \quad \mathbb{E}\{(\|g(Y_1,\theta)\|^2\|g(Y_0,\theta)\|^2|\varepsilon_1 - 1|^2|Y_1|\}^{1+\gamma} < \infty.
\end{aligned} \tag{7}$$

Then, for any consistent estimator $\widehat{\sigma}$ of σ,

$$\widehat{\sigma}^{-1}\widehat{\mathcal{W}}_n(y) \Longrightarrow W\circ G(y), \quad \text{in } D[0,\infty) \text{ and the uniform metric.}$$

Let $0 < y_0 < \infty$. For the rest of this section, we assume that the conditions of Theorem 1 are satisfied unless the contrary is clear. Then it follows from the foregoing theorem that $\widehat{\sigma}^{-1}\widehat{\mathcal{W}}_n(y) \Longrightarrow W \circ G(y)$ on $D[0, y_0]$ with respect to the uniform metric. Therefore, $\widehat{\sigma}^{-1}\widehat{\mathcal{W}}_n(y)$ converges weakly to a centered Gaussian process. Further, as shown in the next section, $\widehat{\sigma}^{-1}\widehat{\mathcal{W}}_n(y)$ has a drift under H_1. This suggests that a test of H_0 vs. H_1 could be based on a suitably chosen functional of $\widehat{\sigma}^{-1}\widehat{\mathcal{W}}_n(y)$. To this end, define

$$\text{TKS} = \left\{\widehat{\sigma}\sqrt{G_n(y_0)}\right\}^{-1} \sup_{0\le y\le y_0} \left|\widehat{\mathcal{W}}_n(y)\right|. \tag{8}$$

Then, by arguments similar to those in Stute et al. (1998), it follows that $\text{TKS} \xrightarrow{d} \sup_{0\le t\le 1}|W(t)|$. Therefore, an asymptotic level-α test rejects H_0 if $\text{TKS} > c_\alpha$ where $P(\sup_{0\le t\le 1}|W(t)| > c_\alpha) = \alpha$. Although the foregoing result holds for any fixed y_0, in practice its choice would depend on the data. A practical choice of y_0 could be the 99th percentile of $\{Y_0, \ldots, Y_n\}$ (see Stute et al., 1998).

For computing $\widehat{\mathcal{W}}_n(y)$, the following equivalent expression may be used:

$$\widehat{\mathcal{W}}_n(y) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} r_i\left[I(Y_{i-1}\le y) - \frac{1}{n}\sum_{j=1}^{n} G_i I(Y_{j-1}\le Y_{i-1}\wedge y)\right], \tag{9}$$

where $r_i := \{Y_i/\Psi(Y_{i-1},\widehat{\theta}_n) - 1\}$ and $G_i := \widehat{g}^T(Y_{j-1})\widehat{C}_{Y_{j-1}}^{-1}\widehat{g}(Y_{i-1})$.

A candidate for $\widehat{\sigma}^2$ in the foregoing theorem is

$$\widehat{\sigma}^2 := n^{-1}\sum_{i=1}^{n}\left\{\frac{Y_i}{\Psi(Y_{i-1},\widehat{\theta}_n)} - 1\right\}^2. \tag{10}$$

By (C3)(b), (C6), and (C10), the $\widehat{\sigma}^2$ in (10) converges in probability to σ^2 under H_0.

3. ASYMPTOTIC POWER

In this section we show that the test introduced in the previous section is consistent against certain fixed alternatives and that it has nontrivial asymptotic power against a class of local nonparametric alternatives converging to the null hypothesis at the rate of $O(n^{-1/2})$.

3.1. Consistency

Let $\upsilon \notin \mathcal{M}$ be a known positive measurable function defined on $\mathbb{R}^+$. The alternative we are interested in is

$$H_a : \tau(y) = \upsilon(y), \quad \forall y \ge 0. \tag{11}$$

Consider the following set of conditions.

(C11)

(a) The estimator $\widehat{\theta}_n$ of θ, obtained under the assumption that H_0 holds, converges in probability to some point in Θ under H_a; we shall also denote this limit by θ.

(b) $\inf_{y \in \mathbb{R}^+} \upsilon(y) > 0$.

(c) $\mathbb{E}[\upsilon(Y_0)/\Psi(Y_0,\theta)] \neq 1$ and $\mathbb{E}\upsilon^2(Y_0) < \infty$ under H_a, where θ is as in part (a) of this condition, and conditions (C3)(b) and (C5)–(C7) are assumed to hold.

(d) There exist a $d > 0$ and a nonnegative function $t(y,\theta)$, measurable in the y-coordinate, such that, for $\dot{g}(y,\theta)$ as in (C5), $\mathbb{E}t^2(Y_0,\theta) < \infty$, $\mathbb{E}\|\dot{g}(Y_0,\theta)\| t(Y_0,\theta) < \infty$, and $\|\Psi(y,\vartheta) - \Psi(y,\theta)\| \leq t(y,\theta)\| \vartheta - \theta\|$ for $y \geq 0$ and $\|\vartheta - \theta\| \leq d$.

(e) There exists a $y > 0$ such that

$$\mathbb{E}\left(\left[\frac{\upsilon(Y_0)}{\Psi(Y_0,\theta)} - 1\right] I(Y_0 \leq y)\right) - B(y,\theta) \neq 0, \tag{12}$$

where $D(x,\theta) := \mathbb{E}\big([\upsilon(Y_0)/\Psi(Y_0,\theta) - 1] g(Y_0,\theta) I(Y_0 \geq x)\big)$, and

$$B(y,\theta) := \int_0^y g^T(x,\theta) C^{-1}(x,\theta) D(x,\theta) dG(x).$$

Now, the following theorem states the consistency of the proposed test.

THEOREM 2. *Assume that (1)–(2), H_a, (C1), (C3)(a), and (C11) hold and that the estimator $\widehat{\sigma}^2$ converges in probability to a constant $\sigma_a^2 > 0$. Then* $P(\text{TKS} > c_\alpha) \to 1$. *That is, the test that rejects H_0 whenever* TKS $> c_\alpha$ *is consistent for H_a.*

Under H_a, by (C1), (C3)(a), (C11), and the ergodic theorem, the $\widehat{\sigma}^2$ of (10) converges in probability to $\sigma_a^2 := \sigma^2 \mathbb{E}\{\upsilon(Y_0)/\Psi(Y_0,\theta)\}^2 + \mathbb{E}\{\upsilon(Y_0)/\Psi(Y_0,\theta) - 1\}^2 > 0$.

3.2. Local Power

Let $\gamma \notin \mathcal{M}$ be a positive measurable function on $\mathbb{R}^+$, let θ be as in H_0, and consider the following sequence of alternatives:

$$H_{n\gamma} : \tau(y) = \Psi(y,\theta) + n^{-1/2}\gamma(y), \qquad y \geq 0. \tag{13}$$

Assume that $\widehat{\theta}_n$ continues to be $\sqrt{n}$-consistent under $H_{n\gamma}$. Let

$$\rho(y) := \mathbb{E}\left[\frac{\gamma(Y_0)}{\Psi(Y_0,\theta)} g(Y_0) I(Y_0 \geq y)\right].$$

Then we have the following theorem.

THEOREM 3. *Assume that (1)–(2), $H_{n\gamma}$, (7), and conditions (C1)–(C10) hold and that the $\widehat{\sigma}$ in Theorem 1 continues to be a consistent estimator of σ. Also, assume that the function γ in (13) satisfies $\mathbb{E}[\gamma^2(Y_0)] < \infty$. Then, for all $y_0 > 0$,*

$$\lim_{n\to\infty} P(\text{TKS} > c_\alpha) = P\left(\sup_{0\le y\le y_0} |W \circ G(y) + \sigma^{-2} M(y)| \ge c_\alpha\right),$$

where $M(y) = \mathbb{E}\Big[\{\gamma(Y_0)/\Psi(Y_0,\theta)\} I(Y_0 \ge y)\Big] - \int_{x\le y} g^T(x) C_x^{-1} \rho(x)\, dG(x)$. *Consequently, the test based on* TKS *of (8) has nontrivial asymptotic power against $H_{n\gamma}$, for all γ for which $M \neq 0$.*

A routine argument shows that the estimator $\widehat{\theta}_n$ defined in (5) continues to satisfy (C10) under $H_{n\gamma}$. In fact, one can verify that under $H_{n\gamma}$ and the assumed conditions, $n^{1/2}(\widehat{\theta}_n - \theta) \to_D \mathcal{N}\big(C^{-1}(0,\theta)\mathbb{E}\{\gamma(Y_0) g(Y_0,\theta)/\Psi(Y_0,\theta)\}, \sigma^2 C^{-1}(0,\theta)\big)$.

Note also that the $\widehat{\sigma}^2$ in (10) continues to be a consistent estimator of σ^2 under $H_{n\gamma}$, because

$$\begin{aligned}
\widehat{\sigma}^2 = {} & n^{-1} \sum_{i=1}^n \left\{ \varepsilon_i \frac{\Psi(Y_{i-1},\theta)}{\Psi\left(Y_{i-1},\widehat{\theta}_n\right)} - 1 \right\}^2 + n^{-2} \sum_{i=1}^n \left\{ \varepsilon_i \frac{\gamma(Y_{i-1})}{\Psi\left(Y_{i-1},\widehat{\theta}_n\right)} \right\}^2 \\
& + 2n^{-1} \sum_{i=1}^n \left\{ n^{-1/2} \frac{\Psi(Y_{i-1},\theta)}{\Psi^2\left(Y_{i-1},\widehat{\theta}_n\right)} (\varepsilon_i - 1)\varepsilon_i \gamma(Y_{i-1}) \right\} \\
& + 2n^{-1} \sum_{i=1}^n \left\{ n^{-1/2} \frac{\left[\Psi(Y_{i-1},\theta) - \Psi(Y_{i-1},\widehat{\theta}_n)\right]}{\Psi^2\left(Y_{i-1},\widehat{\theta}_n\right)} \varepsilon_i \gamma(Y_{i-1}) \right\}.
\end{aligned}$$

It follows from (C3)(b), (C6), and (C10) that, under $H_{n\gamma}$, $\max_{1\le i\le n} |\Psi_i(Y_{i-1},\theta) - \Psi_i(Y_{i-1},\widehat{\theta}_n)| = o_p(1)$ (see (A.15) in the Appendix). Thus, the first term on the right of the last equality converges in probability to σ^2. Because $\mathbb{E}[\gamma^2(Y_0)] < \infty$ and ε_1 is independent of Y_0, by the ergodic theorem, the second term is $o_p(1)$. Because $n^{-1/2} \max_{1\le i\le n} |\Psi(Y_{i-1},\theta)| = o_p(1)$ by (C3)(b), and $\max_{1\le i\le n} |\Psi_i(Y_{i-1}, \theta) - \Psi_i(Y_{i-1},\widehat{\theta}_n)| = o_p(1)$, the last two terms are also $o_p(1)$.

4. A SIMULATION STUDY

A simulation study was carried out to investigate the finite-sample performance of the proposed test for checking the lack-of-fit of a given parametric form for the conditional mean $\mathbb{E}[y_t \mid \mathcal{H}_{t-1}]$ of a Markov MEM. For comparison, the Ljung–Box Q test (Ljung and Box, 1978), a LM test (Meitz and Teräsvirta, 2006), and a generalized moment test (Chen and Hsieh, 2010) were also considered. The Ljung–Box Q test was first used for evaluating a MEM by Engle and Russell (1998). Since then, it has been used in a large proportion of the subsequent studies involving MEMs. The Ljung–Box Q statistic is given by $\text{LBQ}(L) = T(T+2) \sum_{k=1}^{L} (T-k)^{-1} \rho_k^2$, where L is the number of autocorrelation lags included in the statistic and ρ_k^2 is the squared sample autocorrelation at lag k of the residuals corresponding to the estimated model.

4.1. Design of the Study

To evaluate size performances of the tests, six data generating processes (DGPs) were considered for generating $Y_0, Y_1, \ldots, Y_n$. Five of them were based on the model

$$\text{ACD}(1,0): Y_i = \tau_i \varepsilon_i, \qquad \tau_i = 0.2 + 0.1 Y_{i-1}, \tag{14}$$

where the error distributions were, respectively, the Exp(1) (E) and the following Weibull (W), gamma (G), generalized gamma (GG), and Burr (B) distributions:

$$\begin{aligned}
\text{W:} &\quad f_W(x,a) = (a/b)(x/b)^{a-1} \exp\{-(x/b)^a\}, \qquad a = 0.6,\\
\text{G:} &\quad f_G(x) = \{4/\Gamma(2)\} x \exp(-2x),\\
\text{GG:} &\quad f_{GG}(x,a,c) = \{b^{ac}\Gamma(a)\}^{-1} c x^{ac-1} \exp\{-(x/b)^c\}, \qquad a = 3, \qquad c = 0.3,\\
\text{B:} &\quad f_B(x,a,d) = (a/b)(x/b)^{a-1}\{1+d(x/b)^a\}^{-(1+d^{-1})}, \quad a = 1.3, \quad d = 0.4.
\end{aligned}$$

For each of these distributions, the scale parameter b was chosen so that $\mathbb{E}(\varepsilon_1) = 1$. The sixth DGP for size was the following squared-ARCH(1) model:

$$\text{squared-ARCH}(1): X_i = \sqrt{\tau_i}\epsilon_i, \qquad \tau_i = 0.1 + 0.85 Y_{i-1}, \qquad Y_i = X_i^2, \tag{15}$$

where the errors $\{\epsilon_i\}$ are i.i.d. standard normal (N). Here, $Y_i = X_i^2$ follows a Markov MEM of the form (1)–(2).

To evaluate the powers of the tests, four DGPs were considered. The first three were based on the following Markov MEM:

$$\mathcal{M}(\alpha,\beta,\gamma): Y_i = \tau_i \varepsilon_i, \qquad \tau_i = \alpha + \beta Y_{i-1} + \gamma\sqrt{Y_{i-1}}. \tag{16}$$

The data were generated from $\mathcal{M}$(0.2, 0.1, 0.3), $\mathcal{M}$(0.2, 0.1, 0.5), and $\mathcal{M}$(0.2, 0.1, 0.7), keeping Exp(1) as the distribution of the error term ε_i. The fourth DGP to evaluate powers was the following squared-ARCH(2) model:

$$\text{squared-ARCH}(2): X_i = \sqrt{\tau_i}\epsilon_i, \quad \tau_i = 0.2 + 0.1 Y_{i-1} + 0.05 Y_{i-2}, \quad Y_i = X_i^2, \tag{17}$$

where $\{\epsilon_i\}$ are i.i.d. standard normal. The $Y_i = X_i^2$ in (17) follows a MEM. However, it is not Markov.

For each DGP, the null and alternative hypotheses were

$$H_0: \tau_i = \Psi(Y_{i-1}, \vartheta), \quad \text{where } \Psi(y,\vartheta) = \vartheta_1 + \vartheta_2 y, \quad \text{and} \quad H_1: \text{Not } H_0, \tag{18}$$

respectively, where $\vartheta = (\vartheta_1, \vartheta_2)' \in (\mathbb{R}^+)^2$, $Y_i = \tau_i \varepsilon_i$, and ε_i are i.i.d. Thus, ACD(1,0) in (14) and squared-ARCH(1) in (15) are models under H_0, whereas $\mathcal{M}(\alpha, \beta, \gamma)$ and squared-ARCH(2) in (16) and (17) are models under H_1.

To start the recursive DGPs to generate $Y_0, Y_1, \ldots, Y_n$, the initial value of τ_i was set equal to the unconditional mean of Y for every DGP. To ensure that the effect of initialization is negligible, we generated $(n+\ell+1)$ observations with $\ell = 300$, discarded the first ℓ observations, and used the remaining $n+1$ observations as $Y_0, Y_1, \ldots, Y_n$. All the simulation estimates are based on 1,000 repetitions.

It follows from the null hypothesis in (18) that the parametric family to be fitted is $\mathcal{M} = \{\Psi(\cdot, \vartheta) : \Psi(y, \vartheta) = \vartheta_1 + \vartheta_2 y,\ \vartheta_1, \vartheta_2 > 0,\ y \geq 0\}$. Let $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2)'$ denote the QML estimator (5) of θ and let $\widehat{\sigma}^2$ be given by (10). Then we have that $\dot{\Psi}(y, \vartheta) = (1,\ y)'$, $g(y, \vartheta) = \dot{\Psi}(y, \vartheta)/\Psi(y, \vartheta)$,

$$\widehat{g}(y) = g\left(y, \widehat{\theta}_n\right) = [1,\ y]'/\left\{\widehat{\theta}_1 + \widehat{\theta}_2 y\right\}, \qquad r_i = Y_i/\left\{\widehat{\theta}_1 + \widehat{\theta}_2 Y_{i-1}\right\} - 1, \quad \text{and}$$

$$\widehat{C}_y = n^{-1}\sum_{i=1}^{n} \widehat{g}(Y_{i-1})\widehat{g}(Y_{i-1})^T I(Y_{i-1} > y) = n^{-1}\sum_{i=1}^{n} V_{i-1}/\left\{\widehat{\theta}_1 + \widehat{\theta}_2 Y_{i-1}\right\}^2 \times I(Y_{i-1} > y),$$

where

$$V_{i-1} = \begin{bmatrix} 1 & Y_{i-1} \\ Y_{i-1} & Y_{i-1}^2 \end{bmatrix}.$$

Now, let $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ be the ordered values of the sample $Y_1, Y_2, \ldots, Y_n$. Then, with the foregoing choices and with y_0 being the 99% quantile, we have that

$$\mathrm{TKS} = \left\{\widehat{\sigma}\sqrt{0.99}\right\}^{-1} \sup_{0 \leq y \leq y_0} \left|\widehat{\mathcal{W}}_n(y)\right| = \left\{\widehat{\sigma}\sqrt{0.99}\right\}^{-1} \max_{1 \leq i \leq [n0.99]} \left|\widehat{\mathcal{W}}_n(Y_{(i-1)})\right|,$$

where $\widehat{\mathcal{W}}_n$ is as in (9).

The large-sample level-α critical value c_α of TKS is equal to the $100(1-\alpha)\%$ quantile of $\sup_{0 \leq t \leq 1} |W(t)|$. These values, prepared by Dr. R. Brownrigg, are available at http://homepages.mcs.vuw.ac.nz/~ray/Brownian. For $\alpha = 0.01, 0.05$, and 0.10, these critical values are 2.807034, 2.241403, and 1.959964, respectively.

For the LM test, we used the form in Theorem 1 of Meitz and Teräsvirta (2006), with ACD$(2,0)$ being the model under the maintained hypothesis. Because the null model is ACD$(1,0)$, the asymptotic null distribution of LM is χ_1^2 (Meitz and Teräsvirta, 2006). The generalized moment test is the M-test proposed in Chen and Hsieh (2010). Its test statistic M was computed as on page 354 of Chen and Hsieh (2010), using $(\varepsilon_{i-1} - 1)$ as the "misspecification indicator" g_i. Under H_0, the statistic M converges in distribution to χ_1^2 (see Chen and Hsieh, 2010, p. 353). As in Engle and Russell (1998), the critical values for the LBQ(L) were obtained from χ_L^2.

TABLE 1. Percentage of times the null hypothesis was rejected by 5% level tests for different DGPs

DGP	F	TKS	Q_5	Q_{15}	LM	M	TKS	Q_5	Q_{15}	LM	M
				$n = 200$					$n = 500$		
S1	E	2.5	1.4	4.1	2.4	21.8	4.9	2.9	2.9	5.1	8.2
	W	3.0	2.8	3.7	22.0	22.8	4.4	4.1	4.3	21.9	9.6
	G	3.6	2.0	3.2	0.4	22.1	4.5	2.3	3.7	0.3	10.9
	GG	2.9	2.8	3.3	7.7	22.1	3.7	2.6	3.7	8.6	9.9
	B	2.6	2.5	3.8	6.1	19.9	3.9	2.6	3.5	7.4	9.4
				$n = 500$					$n = 1{,}000$		
S2	N	5.7	2.8	4.2	14.0	4.4	5.2	3.9	4.8	11.4	5.4
P1	E	15.5	3.4	5.0	0.0	4.4	28.2	3.1	4.4	0.1	8.8
P2	E	26.7	3.3	4.8	0.1	7.6	55.1	5.4	6.2	0.1	8.2
P3	E	40.4	4.6	5.7	0.2	7.5	76.5	7.6	6.3	0.3	8.9
P4	N	16.3	9.4	7.6	35.2	10.5	17.4	14.7	11.6	50.6	10.8

Note: The DGPs under the null hypothesis are S1 and S2, where S1 is ACD(1,0) in (14) and S2 is squared-ARCH(1) in (15). The error distributions are exponential (E), Weibull (W), gamma (G), generalized gamma (GG), Burr (B), and normal (N). The DGPs P1–P4 are under the alternative hypothesis. Specifically, P1 is $\mathcal{M}(0.1, 0.2, 0.3)$, P2 is $\mathcal{M}(0.1, 0.2, 0.5)$, and P3 is $\mathcal{M}(0.1, 0.2, 0.7)$, where $\mathcal{M}(\alpha, \beta, \gamma)$ is the model in (16), and P4 is squared-ARCH(2) in (17). The test proposed in this paper is TKS. The other tests are the Ljung–Box Q test with lag length L (Q_L), the Lagrange multiplier test (LM), and the M test (M).

4.2. Results

We evaluated the performance of the TKS test with the LM, Ljung–Box Q (with lags 1, 5, 10, and 15), and M tests for the 1%, 5%, and 10% levels, considering different DGPs and sample sizes. A summary of the main results is presented in Table 1. More detailed results are available from the authors on request.

Each entry in Table 1 is the percentage of times H_0 was rejected out of 1,000 repetitions, at the 5% level. For each such entry, say p, a corresponding standard error is $\{p(1-p)/1{,}000\}^{1/2}$. It is evident from Table 1 that the TKS test proposed in this paper performed well in terms of size, although it was slightly undersized for $n = 200$. The M test exhibited serious size distortions for $n = 200$ but to a lesser extent for $n = 500$. Table 1 shows that, for $n = 200$, the type I error rate of the 5% level M test was at least 20%. However, this test performed well in terms of size when $n = 1{,}000, 2{,}000$, and 5,000. The results for LBQ and LM were not particularly good. The LM test exhibited significant sensitivity to the form of the error distribution. For example, for the ACD(1,0) model, the type I error rate of the 5% level LM test ranged from 0.3% for the gamma distribution to 22% for the Weibull distribution.

The estimated powers of the tests (in %) corresponding to the 5% level are also presented in Table 1. The main observations may be summarized as follows.

1. The test proposed in this paper, TKS, performed consistently well throughout. Its performance was either the best or close to the best. Especially, when

the DGP was Markov, the power of the TKS test was substantially better than those for the LBQ, LM, and M tests.

2. The LM test implemented in this study was defined for testing against a squared-ARCH(2) alternative. Consequently, as expected, the LM test performed the best when the DGP was squared-ARCH(2) in (17). However, when the DGP was not squared-ARCH(2), the LM test performed substantially worse than the TKS test.
3. The LBQ test appeared somewhat sensitive to the number of lags included in the statistic. In comparison, the performance of the M test was more consistent, throughout. However, its power was considerably lower than that of the TKS test in all the simulations.

5. EMPIRICAL EXAMPLES

In this section, we illustrate our testing procedure by considering two data examples. For comparison, the LBQ, LM, and M tests considered in Section 4 are also employed.

Example 1

We first apply the testing procedures to monthly squared log returns of Intel stocks. The data for this example were downloaded from the home page of Professor Ruey S. Tsay. The sample consists of monthly log stock returns of Intel Corporation from January 1973 to December 2008. Tsay (2010) studied this data set and concluded that an ARCH(1) model appears to fit well. This indicates that a multiplicative Markov model of the form (1)–(2) may be adequate for squared log returns. We employ the proposed testing procedure for checking the adequacy of the following Markov MEM:

$$Y_i = \Psi(Y_{i-1}, \vartheta)\varepsilon_i, \qquad \Psi(y, \vartheta) = \vartheta_1 + \vartheta_2 y, \qquad \vartheta = (\vartheta_1, \vartheta_2)^T \in (\mathbb{R}^+)^2. \quad \textbf{(19)}$$

First, the model (19) is estimated by the QML estimator (5). The estimated model is $\Psi(y, \widehat{\theta}) = 0.0115 + 0.3661y$, where the standard errors of the parameters are 0.0015 and 0.1226, respectively. Figure 1 provides a plot of $\{\widehat{\sigma}\sqrt{0.99}\}^{-1}|\widehat{\mathcal{W}}_n(y)|$ against y. From this graph, we obtain that TKS $= 1.9192$, which is the supremum of the graph. Hence, the TKS test does not reject the null model in (19) at the 10% level. For the other three tests, the observed test statistics are LBQ(1) $= 0.8001$, LBQ(5) $= 6.7610$, LBQ(10) $= 13.1506$, LBQ(15) $= 30.5898$, LM $= 0.8153$, and $M = 3.5415$, with the corresponding p-values 0.3708, 0.2390, 0.2154, 0.010, 0.3666, and 0.06, respectively.

Thus, apart from LBQ(15), all other tests fail to reject the null model (19) at the 5% level. In the simulations carried out in the previous section, it was evident that the LBQ test was somewhat sensitive to the number of lags included in the statistic. Therefore, it is prudent not to rely on the conclusion of LBQ(15) alone. The LM test, too, in the simulations, was considerably sensitive to the distribution of the error term ε_i, which is unknown to us in this case. However, the TKS test,

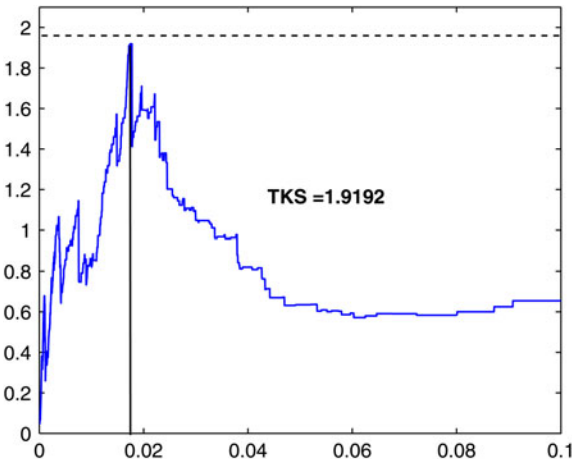

FIGURE 1. Plot of $\{\widehat{\sigma}\sqrt{0.99}\}^{-1}|\widehat{\mathcal{W}}_n(y)|$ against y for squared log returns of Intel stocks. The horizontal dashed line corresponds to the 10% level critical value of the TKS test. The observed test statistic TKS = 1.9192 is the supremum of the curve.

which displayed the best overall performance in the simulations, fails to reject the null model (19). The conclusions from the LBQ(L) (at $L = 1, 5, 10$) and M tests are also similar. Therefore, the indications are that the model (19) provides a good fit for the data.

Example 2

For the second example we consider the viscosity data (Series D) from Box and Jenkins (1976). This sample consists of 310 observations of hourly viscosity readings of a particular chemical process. Based on preliminary diagnostic tests, Datta and McCormick (1995), who studied this data set, concluded that a positive AR(1) model appears to fit well. This suggests that an additive Markov structure is adequate for the hourly viscosity readings of the process. To check whether a multiplicative Markov structure would also be adequate, we test the lack-of-fit of the Markov MEM in (19) for the data set.

We first estimate the model using the QML estimator (5) and obtain that $\Psi(y, \widehat{\theta}) = 4.2814 + 0.5299y$, where the standard errors of the parameters are 0.3111 and 0.0342, respectively. A plot of $\{\widehat{\sigma}\sqrt{0.99}\}^{-1}|\widehat{\mathcal{W}}_n(y)|$ against y for this data set is given in Figure 2. As appears from this graph, we have that TKS = 10.9627. Therefore, our test would reject the null model (19) at any reasonable level of significance. This is also consistent with the conclusions of the M and LBQ tests. However, the LM test fails to reject the null model (19) even when α is as high as 0.7. Specifically, we obtain that LBQ(1) = 80.8609, LBQ(5) = 229.3536, LBQ(10) = 287.1413, LBQ(15) = 306.0199, LM = 0.133, and $M = 338.6616$, with the p-value of LM being 0.7154 and the p-values of M and LBQ(L), at all lags, being almost zero.

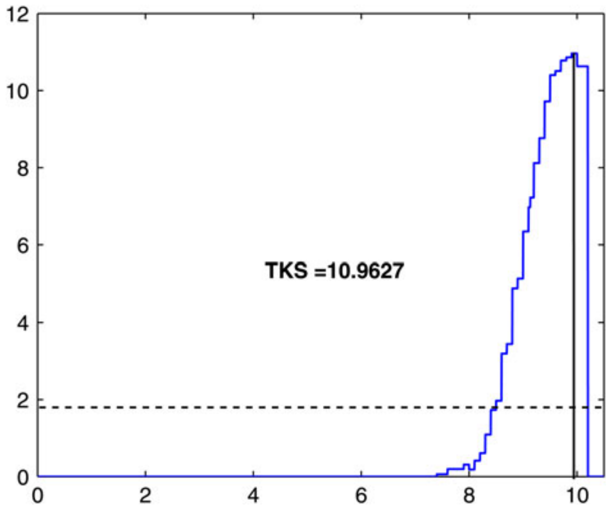

FIGURE 2. Plot of $\{\widehat{\sigma}\sqrt{0.99}\}^{-1}|\widehat{\mathcal{W}}_n(y)|$ against y for the viscosity data (Series D in Box and Jenkins, 1976). The horizontal dashed line corresponds to the 10% level critical value of the TKS test.

Therefore, as indicated by the TKS, M, and LBQ tests, it does not appear that the multiplicative model (19) is a good fit for the data. The LM test performed poorly in the simulations carried out in Section 4 within the Markov setting. Therefore, the conclusion of the LM test may not be very reliable in this case.

6. CONCLUSION

The contribution of this paper has methodological and theoretical components. We developed a new test for the lack-of-fit of a given multiplicative error model having a Markov structure. The family of such Markov MEMs is a simple subfamily of the general class of multiplicative models introduced by Engle (2002). In empirical studies, one would want to consider a general non-Markov model only if a simpler Markov model does not fit. The development of the test proposed in this paper makes use of the Markov structure in the proposed model. Therefore, this test is fundamentally different from the more general ones such as the LM type tests (Meitz and Teräsvirta, 2006; Hautsch, 2006), spectral derivative–based tests (Hong and Lee, 2011), and generalized moment tests (Chen and Hsieh, 2010) for MEMs. Thus, it is reasonable to conjecture that these may also have properties that are complementary. In fact, in a simulation study, the new test performed better overall than the Ljung–Box Q-test, a LM test, and a generalized moment test. Therefore, the indications are that the test proposed in this paper would be useful in empirical studies involving MEMs. We illustrated the testing procedure by considering two data examples.

This paper also makes a theoretical contribution. The approach of constructing a process such as $\widehat{\mathcal{W}}_n(\cdot)$ through a particular martingale transformation of an empirical process marked by the residuals, and then using it to construct an asymptotically distribution free test, is fairly recent. At this stage, this method has been developed for location, regression, and certain autoregressive models. This paper is the first to develop the method for multiplicative time series models.

The technical details of this method, included in the Appendix to this paper, would provide valuable insight and facilitate extension to other models.

REFERENCES

Bauwens, L. (2006) Econometric analysis of intra-daily trading activity on the Tokyo stock exchange. *Monetary and Economic Studies* 24, 1–23.

Bauwens, L. & D. Veredas (2004) The stochastic conditional duration model: A latent factor model for the analysis of financial durations. *Journal of Econometrics* 119, 381–412.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Box, G.E.P. & G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control*, rev. ed. Holden-Day.

Brownlees, C.T., F. Cipollini, & G.M. Gallo (2011) Multiplicative Error Models. Econometrics working papers archive, Università degli Studi di Firenze, Dipartimento di Statistica "G. Parenti."

Chang, M.N. (1990) Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Annals of Statistics* 18, 391–404.

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.

Chow, Y.S. & H. Teicher (1978) *Probability Theory*. Springer-Verlag.

Datta, S. & W.P. McCormick (1995) Bootstrap inference for a first-order autoregression with positive innovations. *Journal of the American Statistical Association* 90, 1289–1300.

Dette, H. & B. Hetzler (2009) Khmaladze transformation of integrated variance processes with applications to goodness-of-fit testing. *Mathematical Methods of Statistics* 18, 97–116.

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

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., G.M. Gallo, & M. Velucchi (2012) Volatility spillovers in East Asian financial markets: A MEM-based approach. *Review of Economics and Statistics* 94, 222–233.

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

Escanciano, J.C. (2010) Asymptotic distribution-free diagnostic tests for heteroskedastic time series models. *Econometric Theory* 26, 744–773.

Escanciano, J.C. & S. Mayoral (2010) Data-driven smooth tests for the martingale difference hypothesis. *Computational Statistics & Data Analysis* 54, 1983–1998.

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

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

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

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic Press.

Hautsch, N. (2006) Testing the Conditional Mean Function of Autoregressive Conditional Duration Models. FRU Working papers 2006/06, University of Copenhagen, Department of Economics, Finance Research Unit.

Hong, Y. & T.-H. Lee (2003) Diagnostic checking for the adequacy of nonlinear time series models. *Econometric Theory* 19, 1065–1121.

Hong, Y. & Y.-J. Lee (2011) Detecting misspecifications in autoregressive conditional duration models and non-negative time-series processes. *Journal of Time Series Analysis* 32, 1–32.

Jasiak, J. (1998) Persistence in intertrade durations. *Finance* 19, 166–195.

Khmaladze, È.V. (1981) A martingale approach in the theory of goodness-of-fit tests. *Akademiya Nauk SSSR. Teoriya Veroyatnosteĭ i ee Primeneniya* 26, 246–265.

Khmaladze, E.V. & H.L. Koul (2004) Martingale transforms goodness-of-fit tests in regression models. *Annals of Statistics* 32, 995–1034.

Koul, H.L. & W. Song (2010) Conditional variance model checking. *Journal of Statistical Planning and Inference* 140, 1056–1072.

Koul, H.L. & W. Stute (1999) Nonparametric model checks for time series. *Annals of Statistics* 27, 204–236.

Lanne, M. (2006) A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics* 4, 594–616.

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.

Luca, G.D. & G. Gallo (2004) Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics & Econometrics* 8, 1–12.

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.

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

Stone, C. (1963) Weak convergence of stochastic processes defined on semi-infinite time intervals. *Proceedings of the American Mathematical Society* 14, 694–696.

Stute, W., S. Thies, & L.X. Zhu (1998) Model checks for regression: An innovation process approach. *Annals of Statistics* 26, 1916–1934.

Stute, W., W.L. Xu, & L.X. Zhu (2008) Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika* 95, 451–467.

Tsay, R.S. (2010) *Analysis of Financial Time Series*, 3rd ed. Wiley Series in Probability and Statistics. Wiley.

von Neumann, J. (1941) Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics* 12, 367–395.

Wooldridge, J.M. (1990) A unified approach to robust, regression-based specification tests. *Econometric Theory* 6, 17–43.

APPENDIX: Proofs

In this section we present the proofs of Theorems 1 and 2. We first obtain several needed preliminaries. The following lemma provides a general weak convergence result about the marked empirical process

$$\alpha_n(y) = n^{-1/2} \sum_{i=1}^{n} \ell(Y_{i-1})(\varepsilon_i - 1) I(Y_{i-1} \le y),$$

where ℓ is a nonnegative measurable function on $\mathbb{R}^+$. This result will be used in the proofs of the other results in this section.

LEMMA 1. *Assume that model (1)–(2), (C1), and (C2) hold and that* $\inf_{y\in\mathbb{R}^+}\tau(y) > 0$. *Suppose, in addition, that for some* $\beta > 0$, $\gamma > 0$,

$$(a)\ \mathbb{E}\ell^4(Y_0) < \infty, \qquad (b)\ \mathbb{E}\Big\{\ell^4(Y_0)|Y_0|^{1+\beta}\Big\} < \infty,$$

$$(c)\ \mathbb{E}\Big\{\ell^2(Y_0)\ell^2(Y_1)|\varepsilon_1 - 1|^2|Y_1|\Big\}^{1+\gamma} < \infty. \tag{A.1}$$

Then $\alpha_n \Longrightarrow W \circ \rho$, *in the space* $D[0,\infty]$ *with respect to uniform metric, where* $\rho(y) := \sigma^2\mathbb{E}\ell^2(Y_0)I(Y_0 \le y)$.

This lemma is similar to Lemma 3.1 of Koul and Stute (1999), but it does not directly follow from that lemma. The main reason is that the present model is multiplicative, whereas the one considered in Koul and Stute (1999) is additive.

Proof of Lemma 1. The convergence of finite-dimensional distributions of $\alpha_n(\cdot)$ follows by an application of the central limit theorem (CLT) for martingales (Hall and Heyde, 1980, Cor. 3.1). To show the tightness of $\alpha_n(\cdot)$ we now argue as in Koul and Stute (1999). First fix $0 \le t_1 < t_2 < t_3 \le \infty$. Then

$$[\alpha_n(t_3) - \alpha_n(t_2)]^2[\alpha_n(t_2) - \alpha_n(t_1)]^2 = n^{-2}\sum_{i,j,k,l} U_iU_jV_kV_l,$$

where $U_i = \ell(Y_{i-1})(\varepsilon_i - 1)I(t_2 < Y_{i-1} \le t_3)$ $\;V_i = \ell(Y_{i-1})(\varepsilon_i - 1)I(t_1 < Y_{i-1} \le t_2)$. Because ε_i is independent of $\{Y_{i-1}, Y_{i-2}, \ldots, Y_0\}$ and $\mathbb{E}(\varepsilon_i) = 1$,

$$\mathbb{E}\left\{n^{-2}\sum_{i,j,k,l} U_iU_jV_kV_l\right\} = n^{-2}\sum_{i,j<k}\mathbb{E}\Big\{V_iV_jU_k^2\Big\} + n^{-2}\sum_{i,j<k}\mathbb{E}\Big\{U_iU_jV_k^2\Big\}. \tag{A.2}$$

Note that by (A.1)(a) the preceding expectations exist.

We shall now find bounds for the two sums on the right-hand side. We only consider the first sum. A bound for the second sum can be obtained similarly. First, let k be an arbitrary integer in the range $2 \le k \le n$. Then, by the inequality $(a+b)^2 \le 2a^2 + 2b^2$, $a, b \in \mathbb{R}$, and the stationarity of $\{Y_i\}$,

$$\sum_{i,j<k}\mathbb{E}\Big\{V_iV_jU_k^2\Big\} \le 2\sigma^2\Bigg[\mathbb{E}\Bigg\{\Bigg(\sum_{i=1}^{k-2} V_i\Bigg)^2\ell^2(Y_{k-1})I(t_2 < Y_{k-1} \le t_3)\Bigg\} + \mathbb{E}\Big\{V_1^2\ell^2(Y_1)I(t_2 < Y_1 \le t_3)\Big\}\Bigg]. \tag{A.3}$$

By conditioning on Y_{k-2} and using Fubini's theorem and the Cauchy–Schwarz inequality, the first expectation inside brackets is the same as

$$\mathbb{E}\Bigg\{\Bigg(\sum_{i=1}^{k-2} V_i\Bigg)^2\int_{t_2}^{t_3}\frac{\ell^2(y)}{\tau(Y_{k-2})}f\left(\frac{y}{\tau(Y_{k-2})}\right)dy\Bigg\}$$

$$\le \int_{t_2}^{t_3}\Bigg\{\mathbb{E}\Bigg(\sum_{i=1}^{k-2} V_i\Bigg)^4\Bigg\}^{1/2}\Bigg\{\ell^4(y)\mathbb{E}\Bigg[\frac{1}{\tau^2(Y_0)}f^2\left(\frac{y}{\tau(Y_0)}\right)\Bigg]\Bigg\}^{1/2}dy.$$

Because the V_i's form a centered martingale difference array, by Burkholder's inequality (Chow and Teicher, 1978, p. 384) and the fact that $\left(\sum_{i=1}^{k-2} V_i^2\right)^2 \le (k-2)\left(\sum_{i=1}^{k-2} V_i^4\right)$,

$$\mathbb{E}\left(\sum_{i=1}^{k-2} V_i\right)^4 \le K\mathbb{E}\left(\sum_{i=1}^{k-2} V_i^2\right)^2 \le K(k-2)^2\mathbb{E}V_1^4.$$

Here and in the rest of the proof, K is a generic constant that does not depend on n, k, or the chosen t_1, t_2, and t_3 but may vary from expression to expression. Now, let

$$F_1(t) := \mathbb{E}(\varepsilon_1 - 1)^4\mathbb{E}\Big(\ell^4(Y_0)I(Y_0 \le t)\Big), \qquad 0 \le t \le \infty,$$

$$F_2(t) := \int_0^t \left\{\ell^4(y)\mathbb{E}\left[\frac{1}{\tau^2(Y_0)} f^2\left(\frac{y}{\tau(Y_0)}\right)\right]\right\}^{1/2} dy, \qquad 0 \le t \le \infty.$$

Then we obtain $\mathbb{E}V_1^4 = \mathbb{E}(\varepsilon_1 - 1)^4\mathbb{E}\big(\ell^4(Y_0)I(t_1 < Y_0 \le t_2)\big) = [F_1(t_2) - F_1(t_1)]$ and $\int_{t_2}^{t_3}\{\ell^4(y)\mathbb{E}[f^2(y/\tau(Y_0))/\tau^2(Y_0)]\}^{1/2}\, dy = [F_2(t_3) - F_2(t_2)]$. Hence, the first expectation inside brackets in (A.3) is bounded from above by

$$K(k-2)[F_1(t_2) - F_1(t_1)]^{1/2}[F_2(t_3) - F_2(t_2)]. \tag{A.4}$$

Because $\mathbb{E}Y_1^4 < \infty$, we have that $\mathbb{E}(\varepsilon_1 - 1)^4 < \infty$. Then, by assumption (A.1)(a), F_1 is a continuous nondecreasing bounded function on $\mathbb{R}^+$. Clearly, F_2 is also nondecreasing and continuous. We shall now show that $F_2(\infty)$ is finite.

To this end, let r be a strictly positive continuous Lebesgue density on $\mathbb{R}^+$ such that $r(y) \sim y^{-1-\beta}$ as $y \to \infty$, where β is as in (A.1)(b). Then, by the Cauchy–Schwarz inequality and Fubini's theorem and using the fact that f/τ is uniformly bounded,

$$\begin{aligned} F_2(\infty) &\le \left[\int_0^\infty \ell^4(y)\mathbb{E}\left[\frac{1}{\tau^2(Y_0)} f^2\left(\frac{y}{\tau(Y_0)}\right)\right] r^{-1}(y)\, dy\right]^{1/2} \\ &\le K\left[\mathbb{E}\left\{\int_0^\infty \ell^4(y)\frac{1}{\tau(Y_0)} f\left(\frac{y}{\tau(Y_0)}\right) r^{-1}(y)\, dy\right\}\right]^{1/2} < \infty, \end{aligned}$$

where the finiteness of the last expectation follows from (A.1)(b).

By conditioning on Y_0, using Fubini's theorem, Hölder's inequality, and the γ as in (A.1)(c), we obtain that the second expectation inside brackets in (A.3) is the same as

$$\begin{aligned} &\int_{t_2}^{t_3} \mathbb{E}\left\{I(t_1 < Y_0 \le t_2)\ell^2(Y_0)\ell^2(y)\left(\frac{y}{\tau(Y_0)} - 1\right)^2 \frac{1}{\tau(Y_0)} f\left(\frac{y}{\tau(Y_0)}\right)\right\} dy \\ &\quad \le \Big\{\mathbb{E}I(t_1 < Y_0 \le t_2)\Big\}^{\gamma/(1+\gamma)} \\ &\qquad \times \int_{t_2}^{t_3}\left[\mathbb{E}\left\{\ell^2(Y_0)\ell^2(y)\left(\frac{y}{\tau(Y_0)} - 1\right)^2 \frac{1}{\tau(Y_0)} f\left(\frac{y}{\tau(Y_0)}\right)\right\}^{1+\gamma}\right]^{1/(1+\gamma)} dy. \end{aligned}$$

Thus,

$$\mathbb{E}\{V_1^2\ell^2(Y_1)I(t_2 < Y_1 \le t_3)\} \le [G(t_2) - G(t_1)]^{\gamma/(1+\gamma)}[F_3(t_3) - F_3(t_2)], \tag{A.5}$$

where, for $t \in [0, \infty]$,

$$F_3(t) := \int_0^t \left[\mathbb{E}\left\{\ell^2(Y_0)\ell^2(y)\left(\frac{y}{\tau(Y_0)} - 1\right)^2 \frac{f\big(y/\tau(Y_0)\big)}{\tau(Y_0)}\right\}^{1/(1+\gamma)}\right]^{1+\gamma} dy.$$

Clearly, F_3 is a nondecreasing and continuous function on $\mathbb{R}^+$. For the boundedness, we shall show that $F_3(\infty)$ is finite. Toward this end, let s be a strictly positive continuous Lebesgue density on $\mathbb{R}^+$ such that $s(y) \sim y^{-1-1/\gamma}$ as $y \to \infty$, where γ is as in (A.1)(c). Arguing as in the case of F_2, we obtain that $F_3(\infty)$ is less than or equal to

$$\left[\int_0^\infty \mathbb{E}\left\{\ell^2(Y_0)\ell^2(y)\left(\frac{y}{\tau(Y_0)}-1\right)^2 \frac{1}{\tau(Y_0)} f\left(\frac{y}{\tau(Y_0)}\right)\right\}^{1+\gamma} s^{-\gamma}(y)\,dy\right]^{1/(1+\gamma)}$$
$$\leq K\left[\mathbb{E}\left\{\ell^2(Y_0)\ell^2(Y_1)(\varepsilon_1-1)^2 s^{-\gamma/(1+\gamma)}(Y_1)\right\}^{1+\gamma}\right]^{1/(1+\gamma)} < \infty.$$

This yields that F_3 is also a continuous nondecreasing and bounded function on $\mathbb{R}^+$. Now, by (A.3) – (A.5) and summing from $k=2$ to $k=n$ we obtain

$$n^{-2} \sum_{i,j<k} \mathbb{E}\{V_i V_j U_k^2\} \leq K\Big\{[F_1(t_2)-F_1(t_1)]^{1/2}[F_2(t_3)-F_2(t_2)]$$
$$+\, n^{-1}[G(t_2)-G(t_1)]^{\gamma/(1+\gamma)}[F_3(t_3)-F_3(t_2)]\Big\}.$$

By similar arguments, the second sum on the right-hand side of (A.2) also has a similar bound. Consequently, tightness of $\{\alpha_n\}$ follows from Theorem 15.6 in Billingsley (1968). This completes the proof of Lemma 1. ■

For the proof of Theorem 1 we need some additional results. The next lemma gives the needed weak convergence result for $\mathcal{U}_n(y,\theta)$.

LEMMA 2. *Suppose (1)–(2), (C1), (C2), (C3)(a), and H_0 hold. Then $\sigma^{-1}\mathcal{U}_n(y,\theta) \Longrightarrow W \circ G(y)$, in $D[0,\infty]$ and uniform metric.*

Proof. Under H_0 and (C3)(a), $\tau(y) = \Psi(y,\theta)$ is bounded away from zero uniformly in y. Because, by (C1), $\mathbb{E}Y_0^4 < \infty$ then condition (A.1) is satisfied for $\ell(y) \equiv 1$. Thus, an application of Lemma 1 completes the proof. ■

For brevity, write $\mathcal{U}_n(y) = \mathcal{U}_n(y,\theta)$, $g(y) = g(y,\theta)$, and $C_y = C(y,\theta)$ and define

$$\mathcal{W}_n(y) := \mathcal{U}_n(y) - \int_0^y g^T(x) C_x^{-1}\left[\int_x^\infty g(z)\,d\mathcal{U}_n(z)\right] dG(x), \qquad \mu_i(y) := I(Y_{i-1} \geq y).$$

The following lemma gives the weak convergence of $\mathcal{W}_n$.

LEMMA 3. *Under (1)–(2), (C1)–(C9), and H_0, $\sigma^{-1}\mathcal{W}_n(y) \Longrightarrow W \circ G(y)$, in $D[0,\infty]$ and uniform metric.*

Proof. Arguing as in Stute et al. (1998) and using a conditioning argument, one can verify that $\mathrm{Cov}\{\sigma^{-1}\mathcal{W}_n(r), \sigma^{-1}\mathcal{W}_n(s)\} = G(r \wedge s)$.

To establish the convergence of finite-dimensional distributions, let $\mathcal{F}_i$ be the σ-algebra generated by $\{\varepsilon_i, \varepsilon_{i-1}, \ldots, Y_{i-1}, Y_{i-2}, \ldots\}$, $i \in \mathbb{Z}$ and

$$h_i(y) = \sigma^{-1}(\varepsilon_i - 1)\Big\{I(Y_{i-1} \leq y) - \int_0^{y \wedge Y_{i-1}} g^T(x) C_x^{-1} g(Y_{i-1})\,dG(x)\Big\}, \qquad i = 1, \ldots, n.$$

Note that $\mathbb{E}\big(h_i(y)|\mathcal{F}_{i-1}\big) = 0$, for all i and $\sigma^{-1}\mathcal{W}_n(y) = n^{-1/2}\sum_{i=1}^n h_i(y)$, for all y. Because $\mathrm{Cov}\big(\sigma^{-1}\mathcal{W}_n(x), \sigma^{-1}\mathcal{W}_n(y)\big) = \mathrm{Cov}\big(W \circ G(x), W \circ G(y)\big)$, by the CLT for martingales (e.g., cf. Hall and Heyde, 1980, Cor. 3.1), all finite-dimensional distributions of $\sigma^{-1}\mathcal{W}_n$ converge to those of $W \circ G$.

Lemma 2 implies the tightness of the process $\mathcal{U}_n(\cdot)$ in uniform metric. It remains to prove the tightness of the second term in $\mathcal{W}_n$. Denote it by $\mathcal{W}_{2n}$. Then,

$$\mathcal{W}_{2n}(y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\varepsilon_i - 1) \int_0^y g^T(x) C_x^{-1} g(Y_{i-1}) \mu_i(x) dG(x).$$

Proceeding as on page 231 of Koul and Stute (1999), let $A(y) := \int_0^y \|g^T(x)C_x^{-1}\| \, dG(x)$, $y \in [0,\infty]$. By condition (C9), $0 < A(\infty) < \infty$. Because G is continuous, the function $\mathcal{A}(y) := A(y)/A(\infty)$ is a strictly increasing continuous distribution function on $[0,\infty]$. Moreover, using the fact that $\|C_x\| \le \int \|g\|^2 dG$, for all $0 \le x \le \infty$, and by the Fubini theorem, for $y_1 < y < y_2$,

$$\begin{aligned} \mathbb{E}\big[\mathcal{W}_{2n}(y_1) - \mathcal{W}_{2n}(y_2)\big]^2 &= \sigma^2 \int_{y_1}^{y_2} \int_{y_1}^{y_2} g^T(x_1) C_{x_1}^{-1} C_{x_1 \vee x_2} C_{x_2}^{-1} g(x_2)\, dG(x_1)\, dG(x_2) \\ &\le \sigma^2 \int \|g(y)\|^2 \, dG(y) [\mathcal{A}(y_2) - \mathcal{A}(y_1)]^2 A^2(\infty). \end{aligned}$$

This bound, together with Theorem 12.3 of Billingsley (1968), implies that $\mathcal{W}_{2n}$ is tight. This completes the proof of Lemma 3. ■

For the proof of Theorem 1 we shall make use of the following lemma which follows from Lemma 3.1 of Chang (1990).

LEMMA 4. *Let V be a relatively compact subset of $D[0, y_0]$. Then with probability 1, for all $y_0 < \infty$, $\int_0^y v(x)[dG_n(x) - dG(x)] \longrightarrow 0$ as $n \to \infty$, uniformly in $0 \le y \le y_0$ and $v \in V$.*

Proof of Theorem 1. Fix a $y_0 > 0$. Recall that $\mathcal{U}_n(y) = \mathcal{U}_n(y,\theta)$ and $\widehat{\mathcal{U}}_n(y) = \mathcal{U}_n(y, \widehat{\theta}_n)$. Let

$$\widetilde{\mathcal{W}}_n(y) := \mathcal{U}_n(y) - \int_0^y \widehat{g}^T(x) \widehat{C}_x^{-1} \Big[\int_x^\infty \widehat{g}(z)\, d\mathcal{U}_n(z) \Big] \, dG_n(x).$$

We shall first show that $\sup_{0 \le y \le y_0} \big| \widehat{\mathcal{W}}_n(y) - \widetilde{\mathcal{W}}_n(y) \big| = o_p(1)$. Write

$$\widehat{\mathcal{W}}_n(y) - \widetilde{\mathcal{W}}_n(y) = \widehat{\mathcal{U}}_n(y) - \mathcal{U}_n(y) - \int_0^y \widehat{g}^T(x) \widehat{C}_x^{-1} J_n(x)\, dG_n(x), \tag{A.6}$$

where $J_n(y) := \int_y^\infty \widehat{g}(z)\, d\widehat{\mathcal{U}}_n(z) - \int_y^\infty \widehat{g}(z)\, d\mathcal{U}_n(z)$.

First, consider $\widehat{\mathcal{U}}_n(y) - \mathcal{U}_n(y)$. Let $\Delta_n := n^{1/2}(\widehat{\theta}_n - \theta)$. By the mean value theorem, there is a sequence of random vectors $\{\theta_n^*\}$ in Θ with $\|\theta_n^* - \theta\| \le \|\widehat{\theta}_n - \theta\|$ and such that

$$\widehat{\mathcal{U}}_n(y) - \mathcal{U}_n(y) = -\Delta_n^T n^{-1} \sum_{i=1}^{n} \varepsilon_i g(Y_{i-1}) I(Y_{i-1} \le y) + \Delta_n^T R_n(y), \tag{A.7}$$

where $R_n(y) := -n^{-1} \sum_{i=1}^{n} \Big(\frac{\Psi(Y_{i-1},\theta)}{\Psi(Y_{i-1},\theta_n^*)} g(Y_{i-1}, \theta_n^*) - g(Y_{i-1}) \Big) \varepsilon_i I(Y_{i-1} \le y).$

Because, by (C3)(a), Ψ is bounded from below, $\kappa := 1/\inf_{y,\vartheta} \Psi(y,\vartheta) < \infty$. By the triangle inequality, $\sup_{y\geq 0}\|R_n(y)\|$ is bounded from above by

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\left\|\left(\frac{\Psi(Y_{i-1},\theta)}{\Psi(Y_{i-1},\theta_n^*)}-1\right)g(Y_{i-1},\theta_n^*)\varepsilon_i+\big(g(Y_{i-1},\theta_n^*)-g(Y_{i-1})\big)\varepsilon_i\right\| \\
&\quad \leq \kappa \max_{1\leq i\leq n}|\Psi(Y_{i-1},\theta)-\Psi(Y_{i-1},\theta_n^*)|\Big(n^{-1}\sum_{i=1}^{n}\|g(Y_{i-1},\theta_n^*)\varepsilon_i\|\Big) \\
&\qquad + n^{-1}\sum_{i=1}^{n}\big\|\big(g(Y_{i-1},\theta_n^*)-g(Y_{i-1})\big)\varepsilon_i\big\|. \qquad \textbf{(A.8)}
\end{aligned}
$$

By condition (C4),

$$
\max_{1\leq i\leq n}|\Psi(Y_{i-1},\theta)-\Psi(Y_{i-1},\theta_n^*)| \leq \|\Delta_n\|n^{-1/2}\max_{1\leq i\leq n}\|\dot{\Psi}(Y_{i-1},\theta)\|+o_p(n^{-1/2}). \qquad \textbf{(A.9)}
$$

Because (C3)(b) gives $\int|\Psi(y,\theta)|^2\,dG(y)<\infty$, along with (C6), we obtain

$$
\int\|\dot{\Psi}(y,\theta)\|^2\,dG(y) \leq \int\|g(y)\|^2\,dG(y)\int|\Psi(y,\theta)|^2\,dG(y)<\infty.
$$

This in turn implies that

$$
n^{-1/2}\max_{1\leq i\leq n}\|\dot{\Psi}(Y_{i-1},\theta)\|=o_p(1). \qquad \textbf{(A.10)}
$$

Thus, in view of (A.9) and (C10), $\max_{1\leq i\leq n}|\Psi(Y_{i-1},\theta)-\Psi(Y_{i-1},\theta_n^*)|=o_p(1)$. By the triangle inequality

$$
n^{-1}\sum_{i=1}^{n}\|g(Y_{i-1},\theta_n^*)\varepsilon_i\| \leq n^{-1}\sum_{i=1}^{n}\|g(Y_{i-1})\|\varepsilon_i+n^{-1}\sum_{i=1}^{n}\|g(Y_{i-1},\theta_n^*)-g(Y_{i-1})\|\varepsilon_i.
$$

Because $\mathbb{E}(\varepsilon_1)=1$ and ε_1 is independent of Y_0, by the ergodic theorem and (C6), the first term on the right-hand side converges almost surely to $\mathbb{E}\|g(Y_0)\|<\infty$. By (C5), the second term, on the set $\{\|\widehat{\theta}_n^*-\theta\|\leq\eta\}$, with η and h as in (C5), is less than or equal to $\{n^{-1}\sum_{i=1}^{n}\|\dot{g}(Y_{i-1})\|\varepsilon_i+n^{-1}\sum_{i=1}^{n}\delta h(Y_{i-1},\theta)\varepsilon_i\}\|\widehat{\theta}^*-\theta\|$. Then (C5) and (C10) together with the ergodic theorem imply

$$
n^{-1}\sum_{i=1}^{n}\|g(Y_{i-1},\theta_n^*)\varepsilon_i\|=O_p(1). \qquad \textbf{(A.11)}
$$

From these derivations, we obtain that the first term in the upper bound (A.8) is $o_p(1)$. A similar argument together with condition (C5) shows that the second term in this bound tends to zero in probability.

Thus, $\sup_{y\geq 0}\|R_n(y)\|=o_p(1)$, and uniformly over $0\leq y\leq\infty$,

$$
\begin{aligned}
\widehat{\mathcal{U}}_n(y)-\mathcal{U}_n(y) &= -\Delta_n^T n^{-1}\sum_{i=1}^{n}\varepsilon_i g(Y_{i-1})I(Y_{i-1}\leq y)+o_p(1) \\
&= -\Delta_n^T n^{-1}\sum_{i=1}^{n}g(Y_{i-1})I(Y_{i-1}\leq y)+o_p(1). \qquad \textbf{(A.12)}
\end{aligned}
$$

The last claim is proved as follows. Because ε_i is independent of $\{Y_{i-1}, Y_{i-2}, \ldots, Y_0\}$, $\mathbb{E}(\varepsilon_1) = 1$ and, by (C6), $\mathbb{E}\|g(Y_0)\| < \infty$, the ergodic theorem implies the pointwise convergence in (A.12). The uniformity is obtained by adapting a Glivenko–Cantelli type argument for the strictly stationary case as explained under (4.1) in Koul and Stute (1999).

Next, consider J_n in (A.6). For the sake of brevity, write $\widehat{g}_{i-1} = \widehat{g}(Y_{i-1})$ and $g_{i-1} = g(Y_{i-1})$. Because $\varepsilon_i = Y_i / \Psi(Y_{i-1}, \theta)$,

$$J_n(y) = -n^{-1/2} \sum_{i=1}^{n} \widehat{g}_{i-1} \left(\frac{\Psi(Y_{i-1}, \widehat{\theta}) - \Psi(Y_{i-1}, \theta)}{\Psi(Y_{i-1}, \widehat{\theta})} \right) \varepsilon_i \mu_i(y)$$
$$= J_{1n}(y)\Delta_n + J_{2n}(y)\Delta_n + J_{3n}(y)\Delta_n + J_{4n}(y) + J_{5n}(y)\Delta_n + J_{6n}(y)\Delta_n,$$

where

$$J_{1n}(y) = -\frac{1}{n} \sum_{i=1}^{n} \widehat{g}_{i-1} \widehat{g}_{i-1}^T \mu_i(y), \qquad J_{2n}(y) = \frac{1}{n} \sum_{i=1}^{n} g_{i-1} g_{i-1}^T (1 - \varepsilon_i) \mu_i(y),$$

$$J_{3n}(y) = \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{g}_{i-1} \widehat{g}_{i-1}^T - g_{i-1} g_{i-1}^T \right) (1 - \varepsilon_i) \mu_i(y),$$

$$J_{4n}(y) = \frac{-1}{\sqrt{n}} \sum_{i=1}^{n} \left[\Psi(Y_{i-1}, \widehat{\theta}) - \Psi(Y_{i-1}, \theta) - (\widehat{\theta} - \theta)^T \dot{\Psi}(Y_{i-1}, \theta) \right] \frac{\widehat{g}_{i-1} \varepsilon_i}{\Psi(Y_{i-1}, \widehat{\theta})} \mu_i(y),$$

$$J_{5n}(y) = \frac{1}{n} \sum_{i=1}^{n} \widehat{g}_{i-1} \left(\widehat{g}_{i-1} - g_{i-1} \right)^T \varepsilon_i \mu_i(y),$$

$$J_{6n}(y) = \frac{1}{n} \sum_{i=1}^{n} \widehat{g}_{i-1} g_{i-1}^T \left(\frac{\Psi(Y_{i-1}, \widehat{\theta}) - \Psi(Y_{i-1}, \theta)}{\Psi(Y_{i-1}, \widehat{\theta})} \right) \varepsilon_i \mu_i(y).$$

By definition $J_{1n}(y) = -\widehat{C}_y$. We now show that

$$\sup_{y \geq 0} \|J_{jn}(y)\| = o_p(1), \qquad j = 2, \ldots, 6. \tag{A.13}$$

Arguing as for (A.12) and (A.11), one obtains, respectively, $\sup_{y \geq 0} \|J_{2n}(y)\| = o_p(1)$ and $n^{-1} \sum_{i=1}^{n} \|\widehat{g}_{i-1}\| \varepsilon_i = O_p(1)$. Then, as Ψ is bounded below by $1/\kappa$, condition (C4) implies that $\sup_{y \geq 0} \|J_{4n}(y)\| \leq \sqrt{n} \max_{1 \leq i \leq n} |\Psi(Y_{i-1}, \widehat{\theta}) - \Psi(Y_{i-1}, \theta) - (\widehat{\theta} - \theta)^T \dot{\Psi}(Y_{i-1}, \theta)| \kappa\, n^{-1} \sum_{i=1}^{n} \|\widehat{g}_{i-1}\| \varepsilon_i = o_p(1)$.

Next, consider $J_{3n}(y)$. Let $\dot{g}_{i-1} = \dot{g}(Y_{i-1}, \theta)$, $h_{i-1} = h(Y_{i-1}, \theta)$ where h is as in assumption (C5), $\gamma_n := \widehat{\theta}_n - \theta$, and $\eta_i = 1 - \varepsilon_i$. Then (C5) and the triangle inequality imply that, on the set $\{\|\gamma_n\| \leq \eta\}$, where η is as in (C5),

$$\sup_{y \geq 0} \|J_{3n}(y)\| \leq \frac{1}{n} \sum_{i=1}^{n} \left[\|\widehat{g}_{i-1} - g_{i-1}\|^2 + 2\|g_{i-1}\| \left(\|\widehat{g}_{i-1} - g_{i-1}\| \right) \right] |\eta_i|$$
$$\leq \frac{1}{n} \sum_{i=1}^{n} \left[\left(\delta h_{i-1} + \|\dot{g}_{i-1}\| \right)^2 \|\gamma_n\|^2 + 2\|g_{i-1}\| \left(\delta h_{i-1} + \|\dot{g}_{i-1}\| \right) \|\gamma_n\| \right] |\eta_i|.$$

Then by (C5), the ergodic theorem, and (C10), $\sup_{y \geq 0} \|J_{3n}(y)\| = o_p(1)$. A similar argument proves (A.13) for $j = 5$. For the case of $j = 6$, note that $\sup_{y \geq 0} \|J_{6n}(y)\|$ is bounded above by

$$\kappa \left(\max_{1 \leq i \leq n} |\Psi(Y_{i-1}, \widehat{\theta}) - \Psi(Y_{i-1}, \theta)| \right) \left\| \frac{1}{n} \sum_{i=1}^{n} \widehat{g}_{i-1} g_{i-1}^T \varepsilon_i \right\|, \tag{A.14}$$

where $1/\kappa$ is the lower bound on Ψ. By (A.10), (C4), and (C10),

$$\max_{1\le i\le n} |\Psi(Y_{i-1},\widehat{\theta}) - \Psi(Y_{i-1},\theta)| = o_p(1). \tag{A.15}$$

By (C5), (C10), and the ergodic theorem, on the set $\{\|\gamma_n\| \le \eta\}$, where η is as in (C5),

$$\begin{aligned}\Big\|\frac{1}{n}\sum_{i=1}^n \widehat{g}_{i-1} g_{i-1}^T \varepsilon_i\Big\| &\le \frac{1}{n}\sum_{i=1}^n (\|\dot{g}_{i-1}\|\|\gamma_n\| + \delta h_{i-1}\|\gamma_n\| + \|g_{i-1}\|)\|g_{i-1}\|\varepsilon_i \\ &= \|\gamma_n\|\Big(\mathbb{E}\|\dot{g}_0\|\|g_0\| + \delta\mathbb{E}(h_0\|g_0\|) + o_p(1)\Big) + \mathbb{E}\|g_0\|^2 + o_p(1) \\ &= \mathbb{E}\|g_0\|^2 + o_p(1) \;=\; O_p(1).\end{aligned}$$

Hence, the upper bound (A.14) is $o_p(1)$. We have thus proved that

$$\sup_{y\ge 0} \|J_n(y) + \widehat{C}_y \Delta_n\| = o_p(1). \tag{A.16}$$

Next, observe $\sup_{y\ge0}\|\widehat{C}_y - C_y\| \le \sup_{y\ge 0}\big\|n^{-1}\sum_{i=1}^n \big(\widehat{g}_{i-1}\widehat{g}_{i-1}^T - g_{i-1}g_{i-1}^T\big)\mu_i(y)\big\| + \sup_{y\ge0}\big\|n^{-1}\sum_{i=1}^n g_{i-1}g_{i-1}^T\mu_i(y) - C_y\big\|$. The first term on the right-hand side is $o_p(1)$ by arguing as for (A.13), $j=3$. A Glivenko–Cantelli type argument and the ergodic theorem imply that the second term is also $o_p(1)$. Thus, $\sup_{y\ge0}\|\widehat{C}_y - C_y\| = o_p(1)$. Consequently, the positive definiteness of C_y for all $y \in [0,\infty)$ implies that

$$\sup_{0\le y\le y_0} \big\|\widehat{C}_y^{-1} - C_y^{-1}\big\| = o_p(1). \tag{A.17}$$

Condition (C5) and the ergodic theorem imply $n^{-1}\sum_{i=1}^n \|\widehat{g}_{i-1} - g_{i-1}\| = o_p(1)$. Hence, (A.17), (C9), and a routine argument yield $n^{-1}\sum_{i=1}^n \widehat{g}_{i-1}^T \widehat{C}_{Y_{i-1}}^{-1} I(Y_{i-1}\le y) = O_p(1)$, uniformly over $0\le y\le y_0$. Upon combining these facts with (A.6), (A.12), and (A.16), we obtain

$$\sup_{0\le y\le y_0} |\widehat{\mathcal{W}}_n(y) - \widetilde{\mathcal{W}}_n(y)| = o_p(1). \tag{A.18}$$

Next, we shall show

$$\sup_{0\le y\le y_0} |\widetilde{\mathcal{W}}_n(y) - \mathcal{W}_n(y)| = o_p(1). \tag{A.19}$$

First observe that $\mathcal{W}_n(y) - \widetilde{\mathcal{W}}_n(y) = D_{1n}(y) + D_{2n}(y) + D_{3n}(y) + D_{4n}(y)$, where

$$\begin{aligned}
D_{1n}(y) &= \int_0^y g^T(x) C_x^{-1}\Big\{\int_x^\infty g(z)\, d\mathcal{U}_n(z)\Big\}\,[dG_n(x) - dG(x)],\\
D_{2n}(y) &= \int_0^y \Big[\widehat{g}^T(x)\big(\widehat{C}_x^{-1} - C_x^{-1}\big)\Big\{\int_x^\infty \widehat{g}(z)\, d\mathcal{U}_n(z)\Big\}\Big]\, dG_n(x),\\
D_{3n}(y) &= \int_0^y \Big[\widehat{g}^T(x) C_x^{-1}\Big\{\int_x^\infty \big(\widehat{g}(z) - g(z)\big)\, d\mathcal{U}_n(z)\Big\}\Big]\, dG_n(x),\\
D_{4n}(y) &= \int_0^y \Big[\big(\widehat{g}^T(x) - g^T(x)\big) C_x^{-1}\Big\{\int_x^\infty g(z)\, d\mathcal{U}_n(z)\Big\}\Big]\, dG_n(x).
\end{aligned}$$

Note that because g_0 and $\varepsilon_1 - 1$ are square integrable, uniformly in $y \geq 0$,

$$\int_y^\infty g(z)d\mathcal{U}_n(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n g_{i-1}(\varepsilon_i - 1)I(Y_{i-1} \geq y)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n g_{i-1}(\varepsilon_i - 1) - \frac{1}{\sqrt{n}}\sum_{i=1}^n g_{i-1}(\varepsilon_i - 1)I(Y_{i-1} \leq y) + o_p(1).$$

By the martingale CLT, the first term is bounded in probability. Lemma 1 together with (C1), (C2), (C3)(a), (7), and the continuous mapping theorem implies that the second term is $O_p(1)$, uniformly over $y \geq 0$. Hence,

$$\sup_{y\geq 0} \| \int_y^\infty g d\mathcal{U}_n \| = O_p(1). \tag{A.20}$$

By (C8) and (C7), $\sup_{0\leq y\leq y_0} \|g(y)^T C_y^{-1}\| < \infty$. These facts together with Lemma 4 yield $\sup_{0\leq y\leq y_0} \|D_{1n}(y)\| = o_p(1)$.

We shall next prove that $\sup_{0\leq y\leq y_0} |D_{jn}(y)| = o_p(1)$ for $j = 2, 3, 4$. Toward this end we make use of the following fact.

$$\sup_{y\geq 0} \left\| n^{-1/2} \sum_{i=1}^n \left(\widehat{g}_{i-1} - g_{i-1}\right)(\varepsilon_i - 1)\mu_i(y) \right\| = o_p(1). \tag{A.21}$$

The proofs of this fact will be given shortly.

Arguing as in the proof of (A.11), by (C5), (C10), and the ergodic theorem, we obtain that $n^{-1}\sum_{i=1}^n \|\widehat{g}_{i-1} - g_{i-1}\| = o_p(1)$ and $n^{-1}\sum_{i=1}^n \|g_{i-1}\| = O_p(1)$. Because for each $0 \leq y \leq y_0$, $C_y - C_{y_0}$ is positive semidefinite, we also have $\sup_{0\leq y\leq y_0} \|C_y^{-1}\| < \infty$. Hence, (A.17), (A.20), (A.21), and a routine argument yield $\sup_{y\in[0,y_0]} |D_{2n}(y)| = o_p(1)$. Similarly, by (A.21), it follows that $\sup_{y\in[0,y_0]} |D_{3n}(y)| = o_p(1)$, and by (A.20), it yields $\sup_{y\in[0,y_0]} |D_{4n}(y)| = o_p(1)$. This completes the proof of $\sup_{0\leq y\leq y_0} |D_{jn}(y)| = o_p(1)$ for $j = 2, 3, 4$, and hence of (A.19).

Consequently, in view of (A.18), we obtain

$$\sup_{0\leq y\leq y_0} |\widehat{\mathcal{W}}_n(y) - \mathcal{W}_n(y)| = o_p(1). \tag{A.22}$$

This fact, together with consistency of $\hat{\sigma}$ for $\sigma > 0$ and Lemma 3, completes the proof of Theorem 1.

We shall now prove (A.21). Again, for the sake of brevity, write $\xi_{i-1} = (\widehat{g}_{i-1} - g_{i-1} - \dot{g}_{i-1}(\widehat{\theta}_n - \theta))$. Observe that the left-hand side of (A.21) is bounded above by

$$\sup_{y\geq 0} \left\| n^{-1/2} \sum_{i=1}^n \xi_{i-1}(\varepsilon_i - 1)\mu_i(y) \right\| + \sup_{y\geq 0} \left\| n^{-1} \sum_{i=1}^n \dot{g}_{i-1}(\varepsilon_i - 1)\mu_i(y) \right\| \|\Delta_n\|. \tag{A.23}$$

The following argument is similar to that in the proof of (4.19) of Koul and Stute (1999). Condition (C5) implies that, on the set $\{\|\widehat{\theta}_n - \theta\| \leq \eta\}$, where η is as in (C5),

$$\sup_{y\geq 0} \left\| n^{-1/2} \sum_{i=1}^n \xi_{i-1}(\varepsilon_i - 1)\mu_i(y) \right\| \leq \delta \Delta_n n^{-1} \sum_{i=1}^n h(Y_{i-1})|\varepsilon_i - 1| = O_p(\delta).$$

Because $\delta > 0$ is arbitrarily chosen, this implies that the first term in (A.23) is $o_p(1)$. On the other hand, as ε_i is independent of $\{Y_{i-1}, Y_{i-2}, \dots, Y_0\}$, $\mathbb{E}(\varepsilon_1) = 1$, and $\mathbb{E}\|\dot{g}_0\| < \infty$. For the proof of Theorem 1 we shall make use of the following lemma which follows from Lemma 3.1 of Chang (1990). ■

Proof of Theorem 2. It suffices to show that $n^{-1/2}|\widehat{\mathcal{W}}_n(y)| = O_p(1)$, for some $0 < y < \infty$ satisfying (12). Fix such a y. Under H_a, $\varepsilon_i = Y_i/v(Y_{i-1})$. Write $v_i := v(Y_i)$, $\Psi_i := \Psi(Y_i, \theta)$, and $\widehat{\Psi}_i := \Psi(Y_i, \widehat{\theta}_n)$. Then, with θ as in (C11),

$$n^{-1/2}\widehat{\mathcal{U}}_n(y) = n^{-1}\sum_{i=1}^{n} \varepsilon_i v_{i-1}\big[\widehat{\Psi}_{i-1}^{-1} - \Psi_{i-1}^{-1}\big] I(Y_{i-1} \le y) + n^{-1}\sum_{i=1}^{n} \big\{\varepsilon_i\big(v_{i-1}/\Psi_{i-1}\big) - 1\big\} I(Y_{i-1} \le y). \tag{A.24}$$

By (C11)(d), for d and $t(\cdot,\theta)$ as in (C11), on the set $\|\widehat{\theta}_n - \theta\| \le d$, the first term on the right-hand side of (A.24) is bounded from above by $\kappa^2 n^{-1}\sum_{i=1}^n \varepsilon_i v_{i-1} t(Y_{i-1},\theta)\|\widehat{\theta}_n - \theta\| = o_p(1)$, by the ergodic theorem and because $\widehat{\theta}_n \to_p \theta$. Hence, by an extended Glivenko–Cantelli type argument,

$$n^{-1/2}\widehat{\mathcal{U}}_n(y) = \mathbb{E}\big([v(Y_0)/\Psi(Y_0,\theta) - 1] I(Y_0 \le y)\big) + o_p(1), \quad \text{under } H_a. \tag{A.25}$$

Recall under (C11), $\mathbb{E}(v(Y_0)/\Psi(Y_0,\theta)) \neq 1$.

Next, let $\widehat{\mathcal{L}}_n(y)$ denote the second term in $\widehat{\mathcal{W}}_n(y)$ and

$$\widehat{K}_n(x) := n^{-1/2}\int_{z \ge x} \widehat{g}(z)\, d\widehat{\mathcal{U}}_n(z), \qquad K_n(x) := n^{-1/2}\int_{z \ge x} g(z)\, d\mathcal{U}_n(z).$$

Recall $\widehat{g}(z) = g(z,\widehat{\theta})$, $g(z) = g(z,\theta)$, and $\mu_i(x) = I(Y_{i-1} \ge x)$. Also, observe that $K_n(x) = n^{-1}\sum_{i=1}^n [\varepsilon_i(v_{i-1}/\Psi_{i-1}) - 1] g_{i-1}\mu_i(x)$, and $\mathbb{E}K_n(x) = \mathbb{E}\big([v(Y_0)/\Psi(Y_0,\theta) - 1]\, g(Y_0,\theta) I(Y_0 \ge x)\big) = D(x,\theta)$. Hence, an adaptation of the Glivenko–Cantelli argument yields

$$\sup_x \|K_n(x) - D(x,\theta)\| = o_p(1). \tag{A.26}$$

Moreover,

$$\begin{aligned}\widehat{K}_n(x) - K_n(x) = {} & n^{-1}\sum_{i=1}^{n} \varepsilon_i v_{i-1}\Big[\frac{1}{\widehat{\Psi}_{i-1}} - \frac{1}{\Psi_{i-1}}\Big][\widehat{g}_{i-1} - g_{i-1}]\mu_i(x) \\ & + n^{-1}\sum_{i=1}^{n} \varepsilon_i v_{i-1}\Big[\frac{1}{\widehat{\Psi}_{i-1}} - \frac{1}{\Psi_{i-1}}\Big] g_{i-1}\mu_i(x) \\ & + n^{-1}\sum_{i=1}^{n} \varepsilon_i \frac{v_{i-1}}{\Psi_{i-1}}\big[\widehat{g}_{i-1} - g_{i-1}\big]\mu_i(x).\end{aligned}$$

Then using the same arguments as before we see that under the assumed conditions,

$$\sup_{x \in [0,\infty]} \|\widehat{K}_n(x) - K_n(x)\| = o_p(1), \quad \text{under } H_a. \tag{A.27}$$

Now,

$$n^{-1/2}\widehat{\mathcal{L}}_n(y) = \int_0^y \widehat{g}^T(x)\big(\widehat{C}_x^{-1} - C_x^{-1}\big)\widehat{K}_n(x)\,dG_n(x) + \int_0^y \widehat{g}^T(x)C_x^{-1}\big(\widehat{K}_n(x) - K_n(x)\big)dG_n(x) + \int_0^y \widehat{g}^T(x)C_x^{-1}K_n(x)dG_n(x) = S_1(y) + S_2(y) + S_3(y), \quad \text{say.}$$

Let $H_n(z) := \int_0^z \|\widehat{g}(x)\|dG_n(x)$. Arguing as before, we see that uniformly in $z \in [0,\infty]$, $H_n(z) = \mathbb{E}\|g(Y_0)\|I(Y_0 \le z) + o_p(1)$. Hence, by (A.17), (A.26), and (A.27), it follows that $|S_1(y)| \le \sup_{x\le y}\|\widehat{C}_x^{-1} - C_x^{-1}\|\sup_x\|\widehat{K}_n(x)\|H_n(y) = o_p(1)$. Similarly, $|S_2(y)| = o_p(1)$, and $S_3(y) = B(y) + o_p(1)$. These facts combined with (A.25) yield

$$n^{-1/2}\widehat{\mathcal{W}}_n(y) = n^{-1/2}\widehat{\mathcal{U}}_n(y) - n^{-1/2}\widehat{\mathcal{L}}_n(y) = \mathbb{E}\big([\frac{\upsilon(Y_0)}{\Psi(Y_0,\theta)} - 1]I(Y_0 \le y)\big) - B(y,\theta) + o_p(1), \quad \text{under } H_a.$$

In view of (12), this completes the proof of Theorem 2. ■

Proof of Theorem 3. Many details of the proof are similar to that of Theorem 1, and so we shall be brief at times. Fix a $y_0 > 0$. We shall shortly show that under the assumptions of Theorem 3, (A.22) continues to hold. Consequently, by the consistency of $\hat{\sigma}$ for $\sigma > 0$ under $H_{n\gamma}$, the weak limit of $\widehat{\sigma}^{-1}\widehat{\mathcal{W}}_n(y)$ is as same as that of $\sigma^{-1}\mathcal{W}_n(y)$. Let $\overline{U}_n(y) := n^{-1/2}\sum_{i=1}^n(\varepsilon_i - 1)I(Y_{i-1} \le y)$ and

$$\overline{\mathcal{W}}_n(y) := \overline{U}_n(y) - \int_0^y g^T(x)C^{-1}(x)\Big[\int_x^\infty g(z)\,d\overline{U}_n(y)\Big]\,dG(x), \qquad y \ge 0.$$

Then $\mathcal{W}_n(y) = \overline{\mathcal{W}}_n(y) + M_n(y)$, $y \ge 0$, where

$$M_n(y) := n^{-1}\sum_{i=1}^n \frac{\gamma(Y_{i-1})}{\Psi(Y_{i-1},\theta)}\varepsilon_i\mu_i(y) - \int_0^y g^T(x)C^{-1}(x)\Big[n^{-1}\sum_{i=1}^n \frac{g_{i-1}\gamma(Y_{i-1})}{\Psi(Y_{i-1},\theta)}\varepsilon_i\mu_i(x)\Big]\,dG(x).$$

Proceeding as in the proof of Lemma 3 we obtain that $\sigma^{-1}\overline{\mathcal{W}}_n(y) \Longrightarrow W \circ G(y)$ in $D[0,\infty)$ and uniform metric. By the ergodic theorem and an extended Glivenko–Cantelli type argument, $\sup_{y\ge 0}|M_n(y) - M(y)| = o_p(1)$, where M is as in Theorem 3. Now, these facts, together with consistency of $\widehat{\sigma}$ for $\sigma > 0$, Slutsky's theorem, and the continuous mapping theorem, complete the proof of Theorem 3.

We shall now prove that (A.22) holds under the conditions of Theorem 3. For brevity, write $\Psi_{i-1}^* := \Psi(Y_{i-1},\theta_n^*)$, $\dot{\Psi}_{i-1}^* := \dot{\Psi}(Y_{i-1},\theta_n^*)$, and $g_{i-1}^* := g(Y_{i-1},\theta_n^*)$. Arguing as for (A.7), we obtain

$$\widehat{\mathcal{U}}_n(y) - \mathcal{U}_n(y) = -\Delta_n^T n^{-1}\sum_{i=1}^n \varepsilon_i\{\Psi_{i-1}\dot{\Psi}_{i-1}^*/\Psi_{i-1}^{*2}\}I(Y_{i-1} \le y) + \Delta_n^T\widetilde{R}_n(y),$$

where $\{\theta_n^*\} \in \Theta$ satisfies $\|\theta_n^* - \theta\| \le \|\widehat{\theta}_n - \theta\|$ and

$$\widetilde{R}_n(y) := -n^{-1}\sum_{i=1}^n \varepsilon_i n^{-1/2}\{\gamma(Y_{i-1})\dot{\Psi}_{i-1}^*/\Psi_{i-1}^{*2}\}I(Y_{i-1} \le y).$$

By the triangle inequality, $n^{-1}\sum_{i=1}^n \varepsilon_i \|\dot{\Psi}_{i-1}^*/\Psi_{i-1}^{*2}\| \le S_n + n^{-1}\sum_{i=1}^n \varepsilon_i \|g_{i-1}/\Psi_{i-1}\|$, where $S_n := n^{-1}\sum_{i=1}^n \varepsilon_i \|\{g_{i-1}^*/\Psi_{i-1}^*\} - \{g_{i-1}/\Psi_{i-1}\}\| \le \max_{1\le i\le n}\|\Psi_{i-1} - \Psi_{i-1}^*\|\kappa^2 \big(n^{-1}\sum_{i=1}^n \|g_{i-1}^*\|\varepsilon_i\big) + \kappa\Big(n^{-1}\sum_{i=1}^n \big\|g_{i-1}^* - g_{i-1}\big\|\varepsilon_i\Big)$. Proceeding as in the proof of Theorem 1, one can obtain that $\max_{1\le i\le n}\|\Psi_{i-1} - \Psi_{i-1}^*\|(n^{-1}\sum_{i=1}^n \|g_{i-1}^*\|\varepsilon_i) = o_p(1)$ and that $(n^{-1}\sum_{i=1}^n \|g_{i-1}^* - g_{i-1}\|\varepsilon_i) = o_p(1)$, under $H_{n\gamma}$. Hence, $S_n = o_p(1)$. Also note that $n^{-1/2}\max_{1\le i\le n}|\gamma(Y_{i-1})| = o_p(1)$ and clearly $n^{-1}\sum_{i=1}^n \varepsilon_i\|g_{i-1}/\Psi_{i-1}\| = O_p(1)$. Thus, $\sup_{y\ge 0}\|\widetilde{R}_n(y)\| \le n^{-1/2}\max_{1\le i\le n}|\gamma(Y_{i-1})|\big(S_n + n^{-1}\sum_{i=1}^n \varepsilon_i\|g_{i-1}/\Psi_{i-1}\|\big) = o_p(1)$. Consequently, under $H_{n\gamma}$, uniformly in $y \in [0,\infty]$,

$$\widehat{\mathcal{U}}_n(y) - \mathcal{U}_n(y) = -\Delta_n^T n^{-1}\sum_{i=1}^n \varepsilon_i \frac{\Psi_{i-1}\dot{\Psi}_{i-1}^*}{\Psi_{i-1}^{*2}} I(Y_{i-1} \le y) + o_p(1).$$

Thus, by proceeding as for the proof of (A.11) we obtain that, under $H_{n\gamma}$, uniformly in $y \in [0,\infty]$, $\widehat{\mathcal{U}}_n(y) - \mathcal{U}_n(y) = -\Delta_n^T n^{-1}\sum_{i=1}^n g_{i-1} I(Y_{i-1} \le y) + o_p(1)$. Then, in view of (A.6), under $H_{n\gamma}$, uniformly in $0 \le y \le y_0$,

$$\begin{aligned}
&\widehat{\mathcal{W}}_n(y) - \widetilde{\mathcal{W}}_n(y)\\
&\qquad = -\Delta_n^T \frac{1}{n}\sum_{i=1}^n g_{i-1} I(Y_{i-1} \le y) - \int_0^y \widehat{g}^T(x)\widehat{C}_x^{-1}\widetilde{J}_n(x)\,dG_n(x) + o_p(1),\\
&\widetilde{J}_n(y) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{g}_{i-1}\Big(\frac{Y_i}{\widehat{\Psi}_{i-1}} - 1\Big)\mu_i(y) - \frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{g}_{i-1}\Big(\frac{Y_i}{\Psi_{i-1}} - 1\Big)\mu_i(y)\\
&\qquad = \Big[-\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{g}_{i-1}\Big(\frac{\widehat{\Psi}_{i-1} - \Psi_{i-1}}{\widehat{\Psi}_{i-1}}\Big)\varepsilon_i\mu_i(y)\Big] + \widetilde{S}_n(y),
\end{aligned}$$

and $\widetilde{S}_n(y) := -n^{-1}\sum_{i=1}^n \big\{\widehat{g}_{i-1}\gamma(Y_{i-1})(\widehat{\Psi}_{i-1}\Psi_{i-1})^{-1}(\widehat{\Psi}_{i-1} - \Psi_{i-1})\varepsilon_i\mu_i(y)\big\}$. Because we assume $\mathbb{E}\gamma^2(Y_{i-1}) < \infty$, by (C5), the ergodic theorem, and a routine argument, $n^{-1}\sum_{i=1}^n \|\widehat{g}_{i-1}\|\gamma(Y_{i-1})\varepsilon_i = O_p(1)$. One can verify, under the assumptions of Theorem 3, that (A.15) continues to hold true. Because Ψ is bounded below by κ^{-1}, then it follows that $\sup_{y\ge 0}\|\widetilde{S}_n(y)\| \le \kappa^2 \max_{1\le i\le n}|\widehat{\Psi}_{i-1} - \Psi_{i-1}|\big[n^{-1}\sum_{i=1}^n \|\widehat{g}_{i-1}\|\gamma(Y_{i-1})\varepsilon_i\big] = o_p(1)$. Consequently, uniformly in $y \ge 0$, $\widetilde{J}_n(y) = -n^{-1/2}\sum_{i=1}^n \widehat{g}_{i-1}\big((\widehat{\Psi}_{i-1} - \Psi_{i-1})/\widehat{\Psi}_{i-1}\big)\varepsilon_i\mu_i(y) + o_p(1)$. Thus, proceeding as in the proof of Theorem 1, we obtain that $\sup_{y\ge 0}\|\widetilde{J}_n(y) + \widehat{C}_y\Delta_n\| = o_p(1)$. This fact and a routine argument yield that (A.18) continues to hold under the assumptions of Theorem 3.

Next we shall show that (A.19) also holds under the assumptions of Theorem 3. First observe that

$$\mathcal{U}_n(y) = \overline{U}_n(y) + n^{-1}\sum_{i=1}^n \Big\{\frac{\gamma(Y_{i-1})}{\Psi_{i-1}}\varepsilon_i I(Y_{i-1} \le y)\Big\}, \qquad y \ge 0.$$

Let $e_n(y) := n^{-1}\sum_{i=1}^n g_{i-1}\frac{\gamma(Y_{i-1})}{\Psi_{i-1}}\varepsilon_i\mu_i(y)$, $\qquad \widetilde{e}_n(y) := n^{-1}\sum_{i=1}^n \widehat{g}_{i-1}\frac{\gamma(Y_{i-1})}{\Psi_{i-1}}\varepsilon_i\mu_i(y)$.

Then, under $H_{n\gamma}$, $\mathcal{W}_n(y) - \widetilde{\mathcal{W}}_n(y) = L_n(y) + \ell_{1n}(y) + \ell_{2n}(y) + \ell_{3n}(y) + \ell_{4n}(y)$, where

$$
\begin{aligned}
L_n(y) &= \int_0^y \widehat{g}^T(x)\widehat{C}_x^{-1}\Big[\int_x^\infty \widehat{g}(z)\, d\overline{U}_n(z)\Big]\, dG_n(x) \\
&\quad - \int_0^y g^T(x)C_x^{-1}\Big[\int_x^\infty g(z)\, d\overline{U}_n(z)\Big]\, dG(x), \\
\ell_{1n}(y) &= \int_0^y g^T(x)C_x^{-1}e_n(x)\, [dG_n(x) - dG(x)], \\
\ell_{2n}(y) &= \int_0^y \Big[\widehat{g}^T(x)\Big(\widehat{C}_x^{-1} - C_x^{-1}\Big)\widetilde{e}_n(x)\Big]\, dG_n(x), \\
\ell_{3n}(y) &= \int_0^y \Big[\widehat{g}^T(x)C_x^{-1}\big(\widetilde{e}_n(x) - e_n(x)\big)\Big]\, dG_n(x), \\
\ell_{4n}(y) &= \int_0^y \Big[\big(\widehat{g}^T(x) - g^T(x)\big)C_x^{-1}e_n(x)\Big]\, dG_n(x).
\end{aligned}
$$

Proceeding as in the proof of Theorem 1, one can show that under $H_{n\gamma}$ and the assumed conditions on $\mathcal{M}$ and γ, $\sup_{0\le y\le y_0}|L_n(y)| = o_p(1) = \sup_{0\le y\le y_0}|\ell_{jn}(y)|$, $j = 1, 2, 3, 4$. ■