Lack-of-fit testing of the conditional mean function in a class of Markov duration models

Hira L. Koul

Michigan State University

Indeewara Perera

Monash University

Mervyn J. Silvapulle

Monash University

Hira Koul was partially supported by the NSF DMS Grant 0704130, and Mervyn Silvapulle and Indeewara Perera were partially supported by the 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. Please 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@buseco.monash.edu.au.

Running head: Lack-of-fit test for an ACD model

Corresponding author: Indeewara Perera, Department of Econometrics and Business Statistics, Monash University, P.O. Box 197, Caulfield East, Australia 3145; e-mail:indeewara.perera@buseco.monash.edu.au.

Abstract

The family of autoregressive conditional duration models, introduced by Engle and Russell (1998), plays a central role in modeling data on duration, such as the duration between trades at a stock exchange. This paper develops a new method for testing the lack-of-fit of a given parametric autoregressive conditional duration model having 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, consistent against a large class of fixed alternatives and has non-trivial 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 significantly better than the general purpose Ljung-Box Q-test. The new test is an important addition to the literature because, in empirical studies, one would want to consider more complicated non-Markov models only if a simpler Markov model does not fit well.

1 INTRODUCTION

The class of autoregressive conditional duration (ACD) models, introduced by Engle and Russell (1998), is widely used in financial econometrics for modeling durations, such as the duration between consecutive trades at a stock exchange and waiting time for the change in price of an asset to exceed a threshold level. In financial econometrics, they are used for studying market microstructure. They could also be used in other areas, for example, for modeling waiting time for service at a queue. A large proportion of the class of ACD models have a complicated probabilistic structure. Consequently, assessing the goodness-of-fit of such models is a non-trivial task. In empirical studies one would want to consider such more 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 goodness of fit of a given model from the simpler subclass of Markov ACD models.

To introduce the ACD class of models, let t_i denote the time of the ith trade and let $Y_i = t_i - t_{i-1}$, for $i = 1, \ldots, n$. Thus, Y_i, the ith duration, is the duration between the $(i-1)$th and ith trades. Let $\mathcal{H}_i$ denote the σ-field generated by $\{Y_i, Y_{i-1}, ...\}$, and $\nu_i = E[Y_i|\mathcal{H}_{i-1}]$. An ACD model for Y_i takes the form,

$$Y_i = \nu_i \varepsilon_i, \quad i \in \mathbb{Z} := \{0, \pm 1, \ldots\}, \tag{1}$$

where $\{\varepsilon_i, i \in \mathbb{Z}\}$ is a sequence of positive independent and identically distributed (i.i.d.) random variables with $E(\varepsilon_0) = 1$, $0 < \text{var}(\varepsilon_0) < \infty$, and ε_i is stochastically independent of $\{(\nu_s, Y_{s-1}), s \le i\}$.

Parametric modeling of ν_i has attracted considerable attention in the recent literature; see Pacurar (2008), for a recent survey. Relatively, literature is scant on testing for the lack-of-fit of a parametric ACD model. A common practice for evaluating an ACD model appears to be 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 *et al.* (2004), Bauwens and Veredas (2004), Luca and Gallo (2004) and Bauwens (2006). The approach employed by Engle and Russell (1998), and the most common to be seen in subsequent studies, is to apply the Ljung-Box Q-test. However, see Pacurar (2008) for a discussion on some issues related to this test.

Some authors examine the moment restrictions of the standardized durations implied by the ACD model. Engle and Russell (1998) introduce a test for no excess dispersion of the estimated residuals, paying particular attention on checking the first and second moments of the residuals when the error distribution is assumed to be either exponential or Weibull. Meitz and Teräsvirta (2006) propose Lagrange multiplier type tests for specification testing.

The focus of the present paper is to introduce a new test that is asymptotically distribution free for testing the goodness of fit of a given Markov ACD model. To be more specific, let Y_i, $i \in \mathbb{Z}$, be a stationary and ergodic Markov process that follows model (1) with $\nu_i = \tau(Y_{i-1})$ for some positive measurable function $\tau(\cdot)$, defined on $\mathbb{R}^+ := [0, \infty)$. Accordingly,

$$Y_i = \tau(Y_{i-1})\varepsilon_i, \quad \text{where } \tau(y) = E(Y_i \mid Y_{i-1} = y), \quad y \geq 0,\ i \in \mathbb{Z}. \tag{2}$$

Let $\Theta \subseteqq \mathbb{R}^q$ for some positive integer q, $\Psi(y, \theta)$ be a given positive function of (y, θ) where $y \geq 0$ and $\theta \in \Theta$, and let $\mathcal{M} = \{\Psi(\cdot, \theta) : \theta \in \Theta\}$ denote the corresponding parametric family. The objective of this paper is to propose an asymptotically distribution free test of

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

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 says that the asymptotic null distribution of the test statistic is that of the supremum of standard Brownian motion on $[0, 1]$. Therefore, the test is asymptotically distribution free, and a set of asymptotic critical values are 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. Perhaps it is worth mentioning that the latter result about local power is completely new and has not been discussed in any existing papers in the context of time series analysis. Section 4 contains a simulation study. An illustrative example is discussed in section 5. In the simulation study, the proposed test performed significantly better than the Ljung-Box Q-test. 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, subsection 2.1 provides a motivation for the test and a brief indication of the approach adopted in constructing the test statistic. Then, subsection 2.2 introduces the regularity conditions, defines the test statistic and states the main result on its asymptotic null distribution.

Let $\{Y_0, Y_1, \ldots, Y_n\}$ be observations of a positive, strictly stationary and ergodic process $\{Y_i\}$ that obeys the model (2). Let G denote the stationary distribution function of Y_0 and $\sigma^2 := \mathrm{var}(\varepsilon_1)$. Let τ, Ψ and the testing problem be as in (2) and (3). Let θ denote the true parameter value under H_0 and let 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 subsection 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 the next subsection. Let $T(y,\vartheta) = \int_0^y [\{\tau(x)/\Psi(x;\vartheta)\} - 1] dG(x)$, and

$$\mathcal{U}_n(y,\vartheta) = n^{-1/2} \sum_{i=1}^{n} \Big\{ \frac{Y_i}{\Psi(Y_{i-1},\vartheta)} - 1 \Big\} I(Y_{i-1} \le y), \quad y \ge 0, \vartheta \in \Theta, \tag{4}$$

where we have assumed that τ, Ψ and G are continuous.

First, consider the special case when the true parameter value θ in H_0 is given. 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, $T(y,\theta) = EI(Y_0 \le y)[\{Y_1/\Psi(Y_0,\theta)\} - 1] = 0$, for all $y \ge 0$. Further, an unbiased estimator of $T(y,\theta)$ is $n^{-1/2}\mathcal{U}_n(y,\theta)$. It is shown later that, under H_0, $\mathcal{U}_n(y,\theta)$ converges weakly to $W \circ G$, where W is standard Brownian motion on $[0,\infty)$. 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 for $\tau(y)$. Let $\widehat{\theta}$ be a $n^{1/2}$-consistent estimator of θ. 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. To construct such a test we appeal to a martingale transform method that has been successfully applied to location, regression and autoregressive models in Stute et al. (1998) and Koul and Stute 1999). This approach yields a functional of $\mathcal{U}_n(y,\widehat{\theta})$ that converges weakly, under H_0, to $W \circ G$. Constructing such a functional and establishing its weak converges is the focus of the next subsection.

The process $\mathcal{U}_n(y,\vartheta)$ is an extension of the so called cumulative sum process for the one sample setting to the current set up. The use of cumulative sum process for testing the lack-of-fit of a given regression function goes back to von Neumann (1941) where he proposed a test of constant regression based on an analog of this process. More recently, analogs of this process have been used by several authors to propose asymptotically distribution free lack-of-fit tests of hypotheses similar to (3) in additive regression type models. More specifically, tests have been developed when the null hypothesis specifies a parametric family for the mean function of a regression model, the mean function of the autoregressive model, and the conditional variance function in a regression model; for example, see Stute *et al.* (1998), Koul and Stute (1999), Dette and Hetzler (2009), Koul and Song (2010). A common feature of all these studies is that they are all for additive models. The ACD model studied in this paper is multiplicative and hence is structurally different.

2.2 The test and the main results

Let F denote the cumulative distribution function [*cdf*] of ε_1. In the sequel, $\|a\|$ denotes Euclidean norm for any vector $a \in \mathbb{R}^q$ and $\|D\| := \sup\{\|a^T D\|;\ a \in \mathbb{R}^q,\ \|a\| = 1\}$, for a $q \times q$ real matrix D. Now let us introduce a set of regularity conditions.

(C1). The *cdf* G is continuous, $G(y) > 0$ for $y > 0$, and $EY_0^4 < \infty$. The sequence of random variables $\{\varepsilon_i\}$ is positive and i.i.d. with $E(\varepsilon_1) = 1$, $0 < \sigma^2 < \infty$ 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 $\int_0^\infty |\Psi(y,\theta)|^2\, dG(y) < \infty$. Moreover, 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 $\dot{\Psi}(y,\vartheta) = \Big[(\partial/\partial\vartheta_1)\Psi(y,\vartheta), \cdots, (\partial/\partial\vartheta_q)\Psi(y,\vartheta)\Big]^T$,

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

(C4). $\sup \sqrt{n} |\Psi(Y_{i-1},\vartheta) - \Psi(Y_{i-1},\theta) - (\vartheta - \theta)' \dot{\Psi}(Y_{i-1},\theta)| = o_p(1)$, where the sup is taken over $\{1 \leq i \leq n,\ \sqrt{n}\|\vartheta - \theta\| \leq K\}$ and K is a given arbitrary positive number.

(C5). There exists 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: $\forall\, \delta > 0, \exists\, \eta > 0$ such that $\|\vartheta - \theta\| \leq \eta$ implies

$$\|g(y,\vartheta) - g(y,\theta) - \dot{g}(y,\theta)(\vartheta - \theta)\| \leq \delta h(y,\theta)\|\vartheta - \theta\|, \quad \forall\, y \geq 0,$$
$$Eh^2(Y_0,\theta) \ < \ \infty, \quad E\|\dot{g}(Y_0,\theta)\|\|g(Y_0,\theta)\|^j < \infty, \qquad j = 0, 1.$$

(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 $n^{1/2}\|\widehat{\theta}_n - \theta\| = O_p(1)$.

Remark 1. An example of $\widehat{\theta}_n$ satisfying Condition (C10) is the quasi maximum likelihood (QML) estimator of θ given by

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

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.

Now, let $\widehat{\mathcal{U}}_n(y) := \mathcal{U}(y, \widehat{\theta}_n)$, $\widehat{g}(y) := g(y, \widehat{\theta}_n)$, $G_n(y) := n^{-1}\sum_{i=1}^n I(Y_{i-1} \le y)$, and $\widehat{C}_y := \int_{x\ge y} \widehat{g}(x)\widehat{g}^T(x)\, dG_n(x)$. The proposed test is to be based on the following analog of the Stute-Thies-Zhu's 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 in turn has roots in the work of Khmaladze (1981).

The next theorem provides the required weak convergence result, where W is standard Brownian motion on $[0, \infty)$. Recall from Stone (1963) that the weak convergence in $D[0, \infty)$ means the weak convergence in $D[0, y]$, for every $0 \le y < \infty$. Here, and in the sequel, the symbol "$\Longrightarrow$" denotes weak convergence.

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

$$(a) \quad E\|g(Y_0,\theta)\|^4 < \infty, \qquad (b) \quad E\{\|g(Y_0,\theta)\|^4 |Y_0|^{1+\beta}\} < \infty,$$

$$(c) \quad E\{(\|g(Y_1,\theta)\|^2 \|g(Y_0,\theta)\|^2 |\varepsilon_1 - 1|^2 |Y_1|\}^{1+\gamma} < \infty. \tag{7}$$

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

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

Let $0 < y_0 < \infty$. For rest of this section, let us 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, let us define

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

Now, by arguments similar to those in Stute *et al.* (1998), we have that $T_n \xrightarrow{d} \sup_{0\le t\le 1}|W(t)|$. Therefore, an asymptotic level-α test rejects H_0 if $T_n > c_\alpha$ where $P(\sup_{0\le t\le 1}|W(t)| > c_\alpha) = \alpha$. While 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 99-th 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:

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

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

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

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

Observe that, under H_0, (C3)(b), (C6) and (C10) imply that (A.15) below holds true, which together with the Law of Large Numbers imply $\widehat{\sigma}^2 \to_p \sigma^2$.

3 Asymptotic Power

In this section we show, under some regularity conditions, that the above test is consistent against certain fixed alternatives, and that it has nontrivial asymptotic power against a large class of $n^{-1/2}$-local nonparametric alternatives.

3.1 Consistency

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

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

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) $\inf_{y\in\mathbb{R}^+} v(y) > 0$. (c) $E[v(Y_0)/\Psi(Y_0, \theta)] \neq 1$ and $Ev^2(Y_0) < \infty$ 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 exists a $d > 0$ and a nonnegative function $t(y, \theta)$, measurable in the y-coordinate, such that, $\|\Psi(y, \vartheta) - \Psi(y, \theta)\| \leq t(y, \theta)\|\vartheta - \theta\|$ and $Et^2(Y_0, \theta) < \infty$, for $y \geq 0$ and $\|\vartheta - \theta\| \leq d$.

$$(e) \qquad E\Big(\Big[\frac{v(Y_0)}{\Psi(Y_0,\theta)} - 1\Big] I(Y_0 \leq y)\Big) - B(y,\theta) \neq 0, \quad \text{for some } \ y > 0, \tag{12}$$

where $D(x, \theta) := E\big([v(Y_0)/\Psi(Y_0, \theta) - 1]g(Y_0, \theta)I(Y_0 \geq x)\big)$, 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 (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, $P(T_n > c_\alpha) \to 1$. That is, the test that rejects H_0 whenever $T_n > c_\alpha$, is consistent for H_a.*

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

3.2 Local Power

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

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

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

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

Then we have the following theorem.

Theorem 3. *Assume that (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 $E[\gamma^2(Y_0)] < \infty$. Then, for all $y_0 > 0$,*

$$\lim_{n\to\infty} P(T_n > c_\alpha) = P\Big(\sup_{0\leq y\leq y_0} |W \circ G(y) + \sigma^{-2}M(y)| \geq c_\alpha\Big),$$

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

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

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

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

One can verify, under $H_{n\gamma}$ and the conditions (C3)(b), (C6) and (C10), that (A.15) below holds true. Thus, the first term in the right hand side of the last equality, by a routine argument, converges in probability to σ^2. Since, $E[\gamma^2(Y_0)] < \infty$, Ψ is bounded away from zero and ε_1 is independent of Y_0, with the aid of the Ergodic Theorem, the second term is $o_p(1)$. Since, by (C3)(b), $n^{-1/2} \max_{1 \le i \le n} |\Psi(Y_{i-1}, \theta)| = o_p(1)$, a routine argument together with (A.15) yields that the last two terms are also $o_p(1)$.

4 A simulation study

A simulation study was carried out to compare the new test introduced in this paper with the Ljung-Box Q-test, which is perhaps one of the more commonly used one in this context. The data generating process [DGP] is defined by

$$\mathcal{M}(m): \quad Y_i = \tau_i \varepsilon_i, \quad \text{where } \tau_i = 0.2 + 0.5 Y_{i-1} + m\tau_{i-1},\ i = 1, \ldots, n \tag{14}$$

where m is a nonnegative number. In our study, we considered the values $0, 0.2$ and 0.4. The null and alternative hypotheses are

$$H_0 : \Psi(y, \vartheta) = 0.2 + \vartheta y, \quad \text{and} \quad H_1 : \text{ Not } H_0 \tag{15}$$

respectively, where it is assumed that $y \ge 0$ and $\vartheta > 0$. Thus, $\mathcal{M}(0)$ is the null model, and $\mathcal{M}(0.2)$ and $\mathcal{M}(0.4)$ are two models under the alternative hypothesis. For the error term, we considered the following Weibull [W], Generalized Gamma [GG] and Burr [B] distributions:

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

For each of these, the scale parameter b was chosen so that $E(\varepsilon_1) = 1$. For each error distribution and the model $\mathcal{M}(m)$, the sample sizes $n = 500$ and $n = 1000$

were considered. Thus, the design has a $2 \times 3 \times 3$ factorial structure with 2 sample sizes, 3 error distributions, and 3 specifications for τ_i. To start the recursive data generating process, the initial value of τ was set equal to $\{0.2/(0.5-m)\}$ which is the unconditional mean of Y under $\mathcal{M}(m)$. 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. All the simulation estimates are based on 1000 repetitions.

It follows from the null hypothesis in (15) that the parametric family to be fitted is $\mathcal{M} = \{\Psi(\cdot,\vartheta) : \Psi(y,\vartheta) = 0.2 + \vartheta y, \vartheta > 0,\, y \geq 0\}$. Let $\widehat{\theta}_n$ denote the quasi-maximum likelihood estimator (5) of θ and let $\widehat{\sigma}^2$ be given by (10). Then, we have that $\dot{\Psi}(y,\vartheta) = y$, $g(y,\vartheta) = \dot{\Psi}(y,\vartheta)/\Psi(y,\vartheta) = y/(0.2+\vartheta y)$, $y \geq 0, \vartheta > 0$,

$$\widehat{g}(y) = g(y,\widehat{\theta}_n) = \frac{y}{0.2+\widehat{\theta}_n y}, \quad r_i = \frac{Y_i}{\Psi(Y_{i-1},\widehat{\theta}_n)} - 1 = \frac{Y_i}{0.2+\widehat{\theta}_n Y_{i-1}} - 1 \quad \text{and}$$

$$\widehat{C}_y = n^{-1}\sum_{i=1}^{n}\widehat{g}(Y_{i-1})\widehat{g}^T(Y_{i-1})I(Y_{i-1}>y) = n^{-1}\sum_{i=1}^{n}\Big(\frac{Y_{i-1}}{0.2+\widehat{\theta}_n Y_{i-1}}\Big)^2 I(Y_{i-1}>y).$$

With the forgoing choices, and y_0 as the 99.5% quantile, we have that

$$T_n = \{\widehat{\sigma}\sqrt{0.995}\}^{-1}\sup_{0\leq y\leq y_0}|\widehat{\mathcal{W}}_n(y)| = \{\widehat{\sigma}\sqrt{0.995}\}^{-1}\max_{1\leq i\leq [n0.995]}|\widehat{\mathcal{W}}_n(Y_{(i-1)})|,$$

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

The large sample level-α critical value c_α of T_n 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.

To compare the performance of the new test statistic T_n with a competitor, we considered the Ljung-Box Q-statistic applied to the estimated residuals as in Engle and Russell (1998). This test appears to be the one that is commonly used in empirical studies involving ACD models. The critical values for a $Q(k)$ with lag length k are obtained from the χ^2 distribution with k degrees of freedom.

The results are presented in Tables 1−3. Each entry in these tables is the proportion of times H_0 was rejected out of the 1000 repetitions. For each entry p in the table, a corresponding standard error could be computed as $\{p(1-p)/1000\}^{-1/2}$. These tables show that the estimated sizes are close to the nominal levels. Therefore, the estimated rejection rates in these tables can be used to compare the performance of the two tests. Let us recall that $\mathcal{M}(0.2)$ and $\mathcal{M}(0.4)$ are two specific models under the alternative hypothesis. The rejection rates in Tables 1−3 for these two models show that, the new test performed substantially better than the general purpose Ljung-Box Q-test.

5 An example

In this section, we shall briefly discuss an example to illustrate the testing procedure, using NYSE price duration data. Price durations from NYSEs Trade and Quote (TAQ) database were studied in detail by Giot (2000), Bauwens and Giot (2003) and Fernandes and Grammig (2005). The data for this example were downloaded from the home page of Dr. Joachim Grammig, who in turn acknowledges Bauwens and Giot for providing the data. We refer to Fernandes and Grammig (2005) and Giot (2000) for a detailed description of the data. The sample consists of the first 1017 of the seasonally adjusted Exxon price durations for the period September to November of 1996. The price duration is defined as the waiting time to witness a cumulative change in the mid-quote price of at least \$0.125.

We employ the proposed test for testing the adequacy of the following Markov ACD model: $\Psi(y,\vartheta) = \vartheta_1 + \vartheta_2 y, \vartheta = (\vartheta_1,\vartheta_2)^T \in (\mathbb{R}^+)^2$. In the standard notation, this is ACD(1,0) model. We used the QML estimator (5) to estimate the model. This yields $\Psi(y, \widehat{\theta}) = 1.034 + 0.067y$, with standard errors 0.054 and 0.037 for the estimates 1.034 and 0.067, respectively. Figure 1 provides a plot of $\{\widehat{\sigma}\sqrt{0.995}\}^{-1}|\widehat{\mathcal{W}}_n(y)|$ against y. From this graph, we have $T_n = 1.4758$, which is the supremum of this graph. From the tables of Dr. R. Brownrigg available at http://homepages.mcs.vuw.ac.nz/∼ray/Brownian, the 22% and 46% critical values are 1.6 and 1.2 respectively. Therefore, the large sample p-value for $T_n = 1.4758$ is between 0.22 and 0.46. Therefore, the test indicates that there is no evidence that the model $\Psi(y, \vartheta) = \vartheta_1 + \vartheta_2 y$, does not fit the data.

6 Conclusion

The contribution of this paper has methodological and theoretical components. We developed a new lack-of-fit test for a given ACD model having a Markov structure. The family of such Markov ACD models is a simple subfamily of the well-known ACD family introduced by Engle and Russell (1998). For example, such a Markov ACD model does not have infinite memory. In empirical studies, one would want to consider a general non-Markov ACD model only if a simpler Markov ACD model does not fit. Because the test makes use of the specific structure of the Markov processes, in contrast to the general purpose ones such as the Ljung-Box Q-test, there are some grounds to conjecture that the test is likely to perform well. In fact, the new test performed better than the Ljung-Box Q-test in a simulation study. Therefore, the indications are that the new test would be useful in empirical studies involving ACD models.

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 asymptotically distribution free test, is fairly recent. At this stage, this method has been developed for location, regression and AR(1) models. This paper is the first one to develop the method for multiplicative time series models.

The ideas that underlie this approach are nontrivial. It is likely to suit only special classes of models. Therefore, the details 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. & Giot, P. (2001) *Econometric Modeling of Stock Market Intraday Activity*. Kluwer Academic Publishers, Boston, Dordrecht, London. Advanced Studies in Theoretical and Applied Econometrics, vol. 38.

Bauwens, L. & Giot, P. (2003) Asymmetric acd models: Introducing price information in acd models. *Empirical Economics* **28**, 709 – 731.

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

Billingsley, P. (1968) *Convergence of probability measures*. John Wiley & Sons Inc., New York.

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

Chow, Y. S. & Teicher, H. (1978) *Probability theory*. Springer-Verlag, New York. Independence, interchangeability, martingales.

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

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

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

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

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

Hall, P. & Heyde, C. C. (1980) *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics.

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. *Teor. Veroyatnost. i Primenen.* **26**, 246–265.

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

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

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

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

Meitz, M. & Teräsvirta, T. (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. (1997) Nonparametric model checks for regression. *The Annals of Statistics* **25**, 613–641.

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

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

APPENDIX: PROOFS

In this section we present the proofs of Theorem 1 and Theorem 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 (2), (C1) and (C2) hold, and that* $\inf_{y \in \mathbb{R}^+} \tau(y) > 0$. *Suppose, in addition, that for some* $\beta > 0$, $\gamma > 0$,

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

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

Remark 3. The above 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 while the one considered in Koul and Stute (1999) is an additive.

Proof of Lemma 1. The convergence of finite dimensional distributions of $\alpha_n(\cdot)$ follows by an application of the CLT for martingales [Hall and Heyde (1980), Corollary 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,

$$[\alpha_n(t_3) - \alpha_n(t_2)]^2 [\alpha_n(t_2) - \alpha_n(t_1)]^2 \;=\; n^{-2} \sum_{i,\, j,\, k,\, l} U_i U_j V_k V_l,$$

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

$$E\Big\{n^{-2} \sum_{i,\ j,\ k,\ l} U_i U_j V_k V_l\Big\} = n^{-2} \sum_{i,\ j<k} E\{V_i V_j U_k^2\} + n^{-2} \sum_{i,\ j<k} E\{U_i U_j V_k^2\}. \qquad \text{(A.2)}$$

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

We shall now find bounds for the two sums in 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 $\{Y_i\}$,

$$\begin{aligned} \sum_{i,\ j<k} E\{V_i V_j U_k^2\} \ &\le\ 2\sigma^2 \Big[E\Big\{ \Big(\sum_{i=1}^{k-2} V_i \Big)^2 \ell^2(Y_{k-1}) I(t_2 < Y_{k-1} \le t_3) \Big\} \\ &\qquad + E\{V_1^2 \ell^2(Y_1) I(t_2 < Y_1 \le t_3)\} \Big]. \qquad \text{(A.3)} \end{aligned}$$

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

$$\begin{aligned} &E\Big\{ \Big(\sum_{i=1}^{k-2} V_i \Big)^2 \int_{t_2}^{t_3} \frac{\ell^2(y)}{\tau(Y_{k-2})} f\Big(\frac{y}{\tau(Y_{k-2})}\Big)\, dy \Big\} \\ &\le\ \int_{t_2}^{t_3} \Big\{ E\Big(\sum_{i=1}^{k-2} V_i \Big)^4 \Big\}^{1/2} \Big\{ \ell^4(y) E\Big[\frac{1}{\tau^2(Y_0)} f^2\Big(\frac{y}{\tau(Y_0)}\Big) \Big] \Big\}^{1/2} dy. \end{aligned}$$

Since the V_i's form a centered martingale difference array, by the Burkholder's inequality [Chow and Teicher (1978), page 384] and the fact $(\sum_{i=1}^{k-2} V_i^2)^2 \le (k-2)(\sum_{i=1}^{k-2} V_i^4)$,

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

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

$$\begin{aligned} F_1(t) \ &:=\ E(\varepsilon_1 - 1)^4 E\big(\ell^4(Y_0) I(Y_0 \le t)\big), \qquad 0 \le t \le \infty, \\ F_2(t) \ &:=\ \int_0^t \Big\{ \ell^4(y) E\Big[\frac{1}{\tau^2(Y_0)} f^2\Big(\frac{y}{\tau(Y_0)}\Big) \Big] \Big\}^{1/2} dy, \qquad 0 \le t \le \infty. \end{aligned}$$

Then, we obtain, $EV_1^4 = E(\varepsilon_1 - 1)^4 E\big(\ell^4(Y_0)I(t_1 < Y_0 \le t_2)\big) = [F_1(t_2) - F_1(t_1)]$ and $\int_{t_2}^{t_3} \{\ell^4(y)E\big[f^2(y/\tau(Y_0))/\tau^2(Y_0)\big]\}^{1/2}\, dy = [F_2(t_3) - F_2(t_2)]$. Hence, the first expectation inside brackets in (A.3) is bounded from the above by

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

Since, $EY_1^4 < \infty$, we have that $E(\varepsilon_1 - 1)^4 < \infty$. 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, we have that $0 < f/\tau$ is uniformly bounded,

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

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

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

Thus,

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

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

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

Clearly, F_3 is a nondecreasing and continuous function on $\mathbb{R}^+$. For the boundedness, we shall show that $F_3(\infty)$ is finite. Towards 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

$$\begin{aligned}
&\Big[\int_0^\infty E\Big\{\ell^2(Y_0)\ell^2(y)\Big(\frac{y}{\tau(Y_0)}-1\Big)^2\frac{1}{\tau(Y_0)}f\Big(\frac{y}{\tau(Y_0)}\Big)\Big\}^{1+\gamma}s^{-\gamma}(y)\,dy\Big]^{1/(1+\gamma)} \\
&\le K\big[E\big\{\ell^2(Y_0)\ell^2(Y_1)(\varepsilon_1-1)^2 s^{-\gamma/(1+\gamma)}(Y_1)\big\}^{1+\gamma}\big]^{1/(1+\gamma)} < \infty,
\end{aligned}$$

This yields that F_3 is also a continuous nondecreasing and bounded function on $\mathbb{R}^+$. Now, by (A.3), (A.4) and (A.5) and summing from $k=2$ to $k=n$ we obtain

$$\begin{aligned}
n^{-2}\sum_{i,\ j<k} E\{V_iV_jU_k^2\} \;\le\; & K\Big\{[F_1(t_2)-F_1(t_1)]^{1/2}[F_2(t_3)-F_2(t_2)] \\
& + n^{-1}[G(t_2)-G(t_1)]^{\gamma/(1+\gamma)}[F_3(t_3)-F_3(t_2)]\Big\}.
\end{aligned}$$

By similar arguments, the second sum in 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 more additional results. The next lemma gives the needed weak convergence result for $\mathcal{U}_n(y,\theta)$.

Lemma 2. *Suppose (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. Since, by (C1), $EY_0^4 < \infty$ 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

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

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

Lemma 3. *Under (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 $\{\varepsilon_i, \varepsilon_{i-1}, \cdots, Y_i, Y_{i-1}, \cdots\}, i \in \mathbb{Z}$ and

$$h_i(y) = \sigma^{-1}(\varepsilon_i - 1)\big\{I(Y_{i-1} \le y) - \int_0^{y \wedge Y_{i-1}} g^T(x) C_x^{-1} g(Y_{i-1})\, dG(x)\big\},\ i = 1, \cdots, n.$$

Note that $E\big(h_i(y)|\mathcal{F}_{i-1}\big) = 0$, 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 CLT for martingales, e.g., cf. Corollary 3.1 of Hall and Heyde (1980), 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 strictly increasing continuous distribution function on $[0,\infty]$. Moreover, using the fact $\|C_x\| \le \int \|g\|^2 dG$, for all $0 \le x \le \infty$, and by the Fubini Theorem, for $y_1 < y < y_2$,

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

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

For the proof of Theorem 1 we also make use of Lemma 3.4 of Stute *et al.* (1998) which in turn is a generalization of Lemma 3.2 of Chang (1990). For the sake of completeness we reproduce it here.

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 $\widehat{\mathcal{U}}_n(y) = \mathcal{U}_n(u, \widehat{\theta}_n)$ and 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

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

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

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

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

Since, 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 the above by

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

By condition (C4),

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

Since (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

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

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

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

Since $E(\varepsilon_1)=1$, and ε_1 is independent of Y_0, by the ET and (C6), the first term in the right hand side converges almost surely (a.s.) to $E\|g(Y_0)\|<\infty$. 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 ET imply

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

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, $\sup_{y\ge 0}\|R_n(y)\| = o_p(1)$, and uniformly over $0 \le y \le \infty$,

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

The last claim is proved as follows. Since ε_i is independent of $\{Y_{i-1}, Y_{i-2}, \cdots, Y_0\}$, $E(\varepsilon_1) = 1$ and, by (C6), $E\|g(Y_0)\| < \infty$, ET implies the point wise 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 $\varepsilon_i = Y_i/\Psi(Y_{i-1},\theta)$,

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

where

$$\begin{aligned}J_{1n}(y) &= -\frac{1}{n}\sum_{i=1}^n \widehat{g}_{i-1}\widehat{g}_{i-1}^T\mu_i(y), \quad J_{2n}(y) = \frac{1}{n}\sum_{i=1}^n g_{i-1}g_{i-1}^T(1-\varepsilon_i)\mu_i(y),\\ J_{3n}(y) &= \frac{1}{n}\sum_{i=1}^n \big(\widehat{g}_{i-1}\widehat{g}_{i-1}^T - g_{i-1}g_{i-1}^T\big)(1-\varepsilon_i)\mu_i(y),\\ J_{4n}(y) &= \frac{-1}{\sqrt{n}}\sum_{i=1}^n \big[\Psi(Y_{i-1},\widehat{\theta}) - \Psi(Y_{i-1},\theta) - (\widehat{\theta}-\theta)^T\dot{\Psi}(Y_{i-1},\theta)\big]\frac{\widehat{g}_{i-1}\varepsilon_i}{\Psi(Y_{i-1},\widehat{\theta})}\mu_i(y),\\ J_{5n}(y) &= \frac{1}{n}\sum_{i=1}^n \widehat{g}_{i-1}\big(\widehat{g}_{i-1} - g_{i-1}\big)^T\varepsilon_i\mu_i(y),\\ J_{6n}(y) &= \frac{1}{n}\sum_{i=1}^n \widehat{g}_{i-1}g_{i-1}^T\bigg(\frac{\Psi(Y_{i-1},\widehat{\theta}) - \Psi(Y_{i-1},\theta)}{\Psi(Y_{i-1},\widehat{\theta})}\bigg)\varepsilon_i\mu_i(y).\end{aligned}$$

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

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

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 implies that, on the set $\{\|\gamma_n\| \leq \eta\}$, where η is as in (C5),

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

Then by (C5), ET 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

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

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

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

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

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

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

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

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

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

Condition (C5) and ET 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 $n^{-1}\sum_{i=1}^n\widehat{g}_{i-1}^T\widehat{C}_{Y_{i-1}}^{-1}I(Y_{i-1}\leq y) = O_p(1)$, uniformly over $0\leq y\leq y_0$. Upon combining these facts with (A.6), (A.12) and (A.16), we obtain

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

Next, we shall show

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

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 $y\geq 0$,

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

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, imply 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 $j=2,3,4$. Towards this end we make use of the following fact.

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

The proofs of this fact will be given shortly.

Arguing as in the proof of (A.11), by (C5), (C10) and ET, we obtain that $n^{-1}\sum_{i=1}^n\|\widehat{g}_{i-1}-g_{i-1}\|=o_p(1)$ and $n^{-1}\sum_{i=1}^n\|g_{i-1}\|=O_p(1)$. Since, for each $0\leq y\leq y_0$, $C_y-C_{y_0}$ is positive semi-definite, 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

$$\sup_{y\ge 0}\Big\|n^{-1/2}\sum_{i=1}^n \xi_{i-1}(\varepsilon_i - 1)\mu_i(y)\Big\| + \sup_{y\ge 0}\Big\|n^{-1}\sum_{i=1}^n \dot g_{i-1}(\varepsilon_i - 1)\mu_i(y)\Big\|\,\|\Delta_n\|. \quad \text{(A.23)}$$

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 $\{\|\widehat\theta_n - \theta\| \le \eta\}$, where η is as in (C5),

$$\sup_{y\ge 0}\Big\|n^{-1/2}\sum_{i=1}^n \xi_{i-1}(\varepsilon_i - 1)\mu_i(y)\Big\| \;\le\; \delta\Delta_n n^{-1}\sum_{i=1}^n h(Y_{i-1})|\varepsilon_i - 1| \;=\; O_p(\delta).$$

Since $\delta > 0$ 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 $\{Y_{i-1}, Y_{i-2}, \cdots, Y_0\}$, $E(\varepsilon_1) = 1$ and $E\|\dot g_0\| < \infty$, we have $E\dot g_0(\varepsilon_1 - 1)\mu_0(y) = 0$. Hence, by (C10), a Glivenko-Cantelli type argument and ET, the second term in the bound (A.23) is $o_p(1)$. ■

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),

$$\begin{aligned} n^{-1/2}\widehat{\mathcal{U}}_n(y) &= n^{-1}\sum_{i=1}^n \varepsilon_i v_{i-1}\big[\widehat\Psi_{i-1}^{-1} - \Psi_{i-1}^{-1}\big] I(Y_{i-1} \le y) \\ &\quad + n^{-1}\sum_{i=1}^n \big\{\varepsilon_i\big(v_{i-1}/\Psi_{i-1}\big) - 1\big\} I(Y_{i-1} \le y). \end{aligned} \quad \text{(A.24)}$$

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 the 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 ET, and because $\widehat\theta_n \to_p \theta$. Hence, by an extended Glivenko-Cantelli type argument,

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

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

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

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

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 $EK_n(x) = E\big([v(Y_0)/\Psi(Y_0,\theta) - 1]g(Y_0,\theta)I(Y_0 \ge x)\big) = D(x,\theta)$. 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 arguments as above we see that under the assumed conditions,

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

Now,

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

Let $H_n(z) := \int_0^z \|\widehat{g}(x)\|dG_n(x)$. Arguing as above, we see that uniformly in $z \in [0,\infty]$, $H_n(z) = E\|g(Y_0)\|I(Y_0 \le z) + o_p(1)$. Hence, by (A.17), (A.27) and (A.26), 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,

$|S_2(y)| = o_p(1)$, while $S_3(y) = B(y) + o_p(1)$. These facts combined with (A.25) yield

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

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, 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

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

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

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

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 ET 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 completes the proof of Theorem 3.

We shall now prove (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

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

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.12) 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 $0 \le y \le y_0$,

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

and $\widetilde{S}_n(y) := -n^{-1} \sum_{i=1}^n \big\{ \widehat{g}_{i-1} \gamma(Y_{i-1}) (\widehat{\Psi}_{i-1} \Psi_{i-1})^{-1} (\widehat{\Psi}_{i-1} - \Psi_{i-1}) \varepsilon_i \mu_i(y) \big\}$. Since we assume $E\gamma^2(Y_{i-1}) < \infty$, by (C5), ET 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. Since Ψ is bounded below by κ^{-1}, then it follows that $\sup_{y\ge 0} \|\widetilde{S}_n(y)\|$

$\leq \kappa^2 \max_{1\leq i\leq n} |\widehat{\Psi}_{i-1} - \Psi_{i-1}| \big[n^{-1}\sum_{i=1}^n \|\widehat{g}_{i-1}\|\gamma(Y_{i-1})\varepsilon_i\big] = o_p(1)$. Consequently, uniformly in $y \geq 0$, $\widetilde{J}_n(y) = -n^{-1/2}\sum_{i=1}^n \widehat{g}_{i-1}\big((\widehat{\Psi}_{i-1}-\Psi_{i-1})/\widehat{\Psi}_{i-1}\big)\varepsilon_i\mu_i(y) + o_p(1)$. Thus, proceeding as in the proof of Theorem 1, we obtain that $\sup_{y\geq 0}\|\widetilde{J}_n(y) + \widehat{C}_y\Delta_n\| = o_p(1)$. This fact and a routine argument yield (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 $\mathcal{U}_n(y) = \overline{U}_n(y) + n^{-1}\sum_{i=1}^n \Big\{\frac{\gamma(Y_{i-1})}{\Psi_{i-1}}\varepsilon_i I(Y_{i-1}\leq y)\Big\},\ y \geq 0$. Let

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

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, $\sup_{0\leq y\leq y_0}|L_n(y)| = o_p(1) = \sup_{0\leq y\leq y_0}|\ell_{jn}(y)|$, $j = 1,2,3,4$. ■

Table 1: Proportion of times H_0 was rejected when the error distribution is Weibull.

		n: 500		n: 1000	
α	DGP	T	Q(10)	T	Q(10)
0.01	M(0)	0.015	0.015	0.013	0.012
	M(0.2)	0.440	0.105	0.85	0.276
	M(0.4)	0.938	0.365	0.98	0.664
0.05	M(0)	0.050	0.049	0.045	0.05
	M(0.2)	0.734	0.245	0.965	0.497
	M(0.4)	0.974	0.529	0.989	0.826
0.1	M(0)	0.096	0.076	0.089	0.094
	M(0.2)	0.848	0.322	0.986	0.612
	M(0.4)	0.985	0.622	0.992	0.867

Table 2: Proportion of times H_0 was rejected when the error distribution is generalized gamma.

		n: 500		n: 1000	
α	DGP	T	Q(10)	T	Q(10)
0.01	M(0)	0.009	0.011	0.007	0.006
	M(0.2)	0.577	0.157	0.927	0.491
	M(0.4)	0.981	0.594	0.998	0.925
0.05	M(0)	0.039	0.056	0.031	0.04
	M(0.2)	0.811	0.353	0.985	0.742
	M(0.4)	0.995	0.805	1	0.962
0.1	M(0)	0.082	0.103	0.076	0.079
	M(0.2)	0.888	0.499	0.993	0.83
	M(0.4)	0.996	0.869	1	0.975

Table 3: Proportion of times H_0 was rejected when the error distribution is Burr.

		n: 500		n: 1000	
α	DGP	T	Q(10)	T	Q(10)
0.01	M(0)	0.016	0.012	0.017	0.029
	M(0.2)	0.455	0.131	0.853	0.329
	M(0.4)	0.955	0.535	0.995	0.837
0.05	M(0)	0.043	0.041	0.058	0.064
	M(0.2)	0.719	0.291	0.965	0.596
	M(0.4)	0.980	0.71	0.997	0.904
0.1	M(0)	0.090	0.076	0.099	0.11
	M(0.2)	0.828	0.402	0.983	0.713
	M(0.4)	0.987	0.793	0.998	0.925

Figure 1: Plot of $\{\widehat{\sigma}\sqrt{0.995}\}^{-1}|\widehat{\mathcal{W}}_n(y)|$ against y. The observed value of the test statistic T_n and the critical values c_α for $\alpha = 0.22$ and 0.46 are also displayed.