The Annals of Statistics
1999, Vol. 27, No. 1, 204–236

NONPARAMETRIC MODEL CHECKS FOR TIME SERIES

BY HIRA L. KOUL[1] AND WINFRIED STUTE

Michigan State University and Justus-Liebig-University of Giessen

This paper studies a class of tests useful for testing the goodness-of-fit of an autoregressive model. These tests are based on a class of empirical processes marked by certain residuals. The paper first gives their large sample behavior under null hypotheses. Then a martingale transformation of the underlying process is given that makes tests based on it asymptotically distribution free. Consistency of these tests is also discussed briefly.

1. Introduction and summary. This paper studies some general methods for testing the goodness-of-fit of a parametric model for a real-valued stationary Markovian time series X_i, $i = 0, \pm 1, \pm 2, \ldots$. Much of the existing literature is concerned with the parametric modeling in terms of the conditional mean function μ of X_i, given X_{i-1}. That is, one assumes the existence of a parametric family

$$\mathscr{M} = \{m(\cdot, \theta)\colon \theta \in \Theta\}$$

of functions and then proceeds to estimate θ or test the hypothesis $\mu \in \mathscr{M}$, that is, $\mu(\cdot) = m(\cdot, \theta_0)$ for some θ_0 in Θ, where Θ is a proper subset of the q-dimensional Euclidean space $\mathbb{R}^q$. One of the reasons for this is that parametric models continue to be attractive among practitioners because the parameter θ together with the functional form of $m(\cdot, \theta)$ describes, in a concise way, the impact of the past observations on the predicted variable. Since there may be several competing models, in order to prevent wrong conclusions, every statistical inference which is based on a model $\mathscr{M}$ should be accompanied by a proper model check, that is, by a test for the hypothesis $\mu \in \mathscr{M}$. The best known example is the usual linear autoregressive model where $m(x, \theta) = x\theta$. The proposed classical tests are based on the least squares residuals. For a discussion of these types of tests, see, for example, Tong (1990), Chapter 5, and MacKinnon (1992) and references therein.

Robinson (1983), Roussas and Tran (1992), Truong and Stone (1992), among others, provide various types of nonparametric estimates of μ which in turn can be used to construct tests of the hypothesis $\mu \in \mathscr{M}$. This has been done, for example, in McKeague and Zhang (1994) and Hjellvik and Tjøstheim (1995, 1996). This approach requires smoothing of the data in addition to the estimation of the finite-dimensional parameter vector and leads to less precise fits.

The use of the conditional mean function as an autoregressive function is justified partly for historical reasons and partly for convenience. In the

Received January 1997; revised November 1998.
[1]Supported in part by the Humboldt Research Award and NSF Grant DMS-94-02904.
AMS 1991 *subject classifications*. Primary 60F17; secondary 62M10, 62M30, 62J02.
Key words and phrases. Marked empirical process, ψ-residuals, martingale transform tests, autoregressive median function.

presence of non-Gaussian innovations it is desirable to look for other dispersions that would lead to different autoregressive functions and which may well be equally appropriate to model the dynamics of the underlying process. For example, if the innovations are believed to form a white noise from a double exponential distribution, then the conditional median function would be a proper autoregressive function. The present paper discusses a class of tests for testing goodness-of-fit hypotheses pertaining to a class of implicitly defined autoregressive functions. The proposed approach avoids smoothing methodology and leads to tests that are consistent against a broad class of alternatives. Even though many of our results are extendable to higher order autoregression, our discussion is confined to the first-order autoregression partly for the sake of clarity of the exposition and partly for some technical reasons. See Remark 2.4 below.

Now, to describe these procedures, let ψ be a nondecreasing real-valued function such that $\mathbb{E}|\psi(X_1 - r)| < \infty$, for each $r \in \mathbb{R}$. Define the ψ-autoregressive function m_ψ by the requirement that

$$\mathbb{E}[\psi(X_1 - m_\psi(X_0))|X_0] = 0 \quad \text{a.s.} \tag{1.1}$$

and the corresponding marked empirical process, based on a sample of size $n+1$, by

$$V_{n,\psi}(x) := n^{-1/2} \sum_{i=1}^{n} \psi(X_i - m_\psi(X_{i-1}))\, I(X_{i-1} \le x), \qquad x \in \mathbb{R}.$$

The marks, or weights at X_{i-1}, of the process $V_{n,\psi}$ are given by the ψ-innovations $\psi(X_i - m_\psi(X_{i-1}))$, $1 \le i \le n$. The process $V_{n,\psi}$ is uniquely determined by these marks and the variables $\{X_{i-1}\}$ and vice versa.

Observe that, if $\psi(x) \equiv x$, then $m_\psi = \mu$, and if $\psi(x) \equiv I(x > 0) - (1-\alpha)$ for a $0 < \alpha < 1$, then $m_\psi(x) \equiv m_\alpha(x)$, the αth quantile of the conditional distribution of X_1, given $X_0 = x$. The choice of ψ is up to the practitioner. If the desire is to have a goodness-of-fit procedure that is less sensitive to outliers in the innovations $X_i - m_\psi(X_{i-1})$, then one may choose a bounded ψ. The motivation for the above definition of m_ψ comes from Huber (1981). Its existence for a general ψ and its kernel type estimators in the time series context have been discussed by Robinson (1984). In the sequel, m_ψ is assumed to exist uniquely.

The process $V_{n,\psi}$ takes its value in the Skorokhod space $D(-\infty, \infty)$. Extend it continuously to $\pm\infty$ by putting

$$V_{n,\psi}(-\infty) = 0 \quad \text{and} \quad V_{n,\psi}(+\infty) = n^{-1/2} \sum_{i=1}^{n} \psi(X_i - m_\psi(X_{i-1})).$$

Then $V_{n,\psi}$ becomes a process in $D[-\infty, \infty]$ which, modulo a continuous transformation, is the same as the more familiar $D[0, 1]$. Throughout we shall assume that the underlying process is ergodic, the stationary distribution (d.f.) G of the X's is continuous and that

$$\mathbb{E}\psi^2(X_1 - m_\psi(X_0)) < \infty. \tag{1.2}$$

It readily follows from (1.1) and (1.2) that $V_{n,\psi}(x)$ is a mean zero square integrable martingale for each $x \in \mathbb{R}$. Thus one obtains from the martingale central limit theorem [Hall and Heyde (1980), Corollary 3.1] and the Cramér–Wold device that all finite-dimensional distributions of $V_{n,\psi}$ converge weakly to a multivariate normal distribution with mean vector zero and covariance matrix given by the covariance function

$$K_{\psi}(x, y) = \mathbb{E}\psi^2(X_1 - m_{\psi}(X_0))\, I(X_0 \le x \wedge y), \qquad x, y \in \mathbb{R}.$$

Under some additional assumptions, Theorem 2.1 below establishes the weak convergence of $V_{n,\psi}$ to a continuous Gaussian process V_{ψ} with the covariance function K_{ψ}. Since the function

$$\tau_{\psi}^2(x) := K_{\psi}(x, x) = \mathbb{E}\psi^2(X_1 - m_{\psi}(X_0))\, I(X_0 \le x)$$

is nondecreasing and nonnegative, V_{ψ} admits a representation

$$V_{\psi}(x) = B(\tau_{\psi}^2(x)) \quad \text{in distribution,} \tag{1.3}$$

where B is a standard Brownian motion on the positive real line. Note that the continuity of the stationary d.f. G implies that of τ_{ψ} and hence that of $B(\tau_{\psi}^2)$. The representation (1.3), Theorem 2.1 and the continuous mapping theorem yield

$$\sup_{x\in\mathbb{R}} |V_{n,\psi}(x)| \Rightarrow \sup_{0\le t\le \tau_{\psi}^2(\infty)} |B(t)| = \tau_{\psi}(\infty) \sup_{0\le t\le 1} |B(t)| \quad \text{in law.}$$

Here and in the sequel, $\Rightarrow$ denotes the convergence in distribution.

The above observations are useful for testing the simple hypothesis $\tilde{H}_0$: $m_{\psi} = m_0$, where m_0 is a known function. Estimate (under $m_{\psi} = m_0$) the variance $\tau_{\psi}^2(x)$ by

$$\tau_{n,\psi}^2(x) := n^{-1} \sum_{i=1}^{n} \psi^2(X_i - m_0(X_{i-1})) I(X_{i-1} \le x), \qquad x \in \mathbb{R},$$

and replace m_{ψ} by m_0 in the definition of $V_{n,\psi}$. Write $s_{n,\psi}^2$ for $\tau_{n,\psi}^2(\infty)$. Then, for example, the Kolmogorov–Smirnov (K–S) test based on $V_{n,\psi}$ of the given asymptotic level would reject the hypothesis $\tilde{H}_0$ if $\sup\{s_{n,\psi}^{-1}|V_{n,\psi}(x)|: x \in \mathbb{R}\}$ exceeds an appropriate critical value obtained from the boundary crossing probabilities of a Brownian motion which are readily available on the unit interval. More generally, in view of Remark 4.1 below, the asymptotic level of any test based on a continuous function of $s_{n,\psi}^{-1} V_{n,\psi}((\tau_{n,\psi}^2)^{-1})$ can be obtained from the distribution of the corresponding function of B on $[0, 1]$, where $(\tau_{n,\psi}^2)^{-1}(t) := \inf\{x \in \mathbb{R}: \tau_{n,\psi}^2(x) \ge t\}$, $t \ge 0$.

Theorem 2.1 is useful for testing the simple hypothesis $\tilde{H}_0$; for testing a composite parametric hypothesis, the process $V_{n,\psi}$ requires some modification. Consider the null hypothesis

$$H_0\colon m_{\psi}(\cdot) = m(\cdot, \theta_0) \quad \text{for some } \theta_0 \in \Theta.$$

Let θ_n be a consistent estimator of θ_0 under H_0 based on $\{X_i,\ 0 \le i \le n\}$. Define

$$V^1_{n,\psi}(x) = n^{-1/2} \sum_{i=1}^{n} \psi(X_i - m(X_{i-1}, \theta_n)) I(X_{i-1} \le x), \qquad x \in \mathbb{R}.$$

The process $V^1_{n,\psi}$ is a marked empirical process, where the marks, or the weights at X_{i-1}, are now given by the ψ-residuals $\psi(X_i - m(X_{i-1}, \theta_n))$. It is uniquely determined by the $\{X_{i-1}\}$ and these residuals and vice versa. Tests for H_0 can be based on an appropriately scaled discrepancy of this process. For example, an analogue of the K–S test would reject H_0 in favor of H_1 if $\sup\{\sigma^{-1}_{n,\psi}|V^1_{n,\psi}(x)|\colon x \in \mathbb{R}\}$ is too large, where $\sigma^2_{n,\psi} := n^{-1}\sum_{i=1}^{n}\psi^2(X_i - m(X_{i-1}, \theta_n))$. These tests, however, are not generally asymptotically distribution free (ADF).

The main focus of the present paper is to construct a transform of the $V^1_{n,\psi}$ process whose limiting distribution is known so that tests based on it will be ADF. To do this, one first needs to establish the weak convergence of these processes. As indicated earlier, Theorem 2.1 below obtains the weak convergence of $V_{n,\psi}$, for a general ψ, to a continuous Brownian motion with respect to the time τ^2_ψ, under some moment assumptions on $\psi(X_1 - m_\psi(X_0))$ and X_1 and under the assumption that the conditional d.f.'s F_y of $X_1 - m_\psi(X_0)$, given $X_0 = y$, have uniformly bounded Lebesgue densities. Theorem 2.2 and Corollary 2.1 present similar results for the $V^1_{n,\psi}$ processes when H_0 is satisfied, under the same moment condition and under some additional conditions on $\mathscr{M}$ and ψ. For technical reasons, the cases of an absolutely continuous ψ and a general bounded ψ are handled separately. The estimator θ_n is assumed to be asymptotically linear.

Now assume that $\sigma^2_\psi(x) := \mathbb{E}[\psi^2(X_1 - m(X_0, \theta_0))|X_0 = x] \equiv \sigma^2_\psi$ does not depend on x. In this case $\tau^2_\psi(x) \equiv \sigma^2_\psi G(x)$, for all real x. Using ideas of Khmaladze (1981), a linear transformation T of the underlying processes is then constructed so that for a general ψ, $TV_\psi(\cdot)/\sigma_\psi$ is a transformed Brownian motion $B \circ G$ and $TV^1_{n,\psi}/\sigma_{n,\psi}$ converges in distribution to TV_ψ/σ_ψ. This is done in Theorem 2.3 below. Informally speaking, T maps $V^1_{n,\psi}$ into the (approximate) martingale part of its Doob–Meyer decomposition.

The transformation T depends on the unknown entities θ_0 and G. An estimator T_n of T is then constructed and it is shown that $T_n V^1_{n,\psi}(\cdot)/\sigma_{n,\psi}$ also converges in distribution to the process $B \circ G$. This is done in Theorem 2.4. Because the transformation is generally unstable in the extreme right tails, this weak convergence can be proved only in $D[-\infty, x_0]$, for an $x_0 < \infty$. Consequently, the tests based on the transformed process $T_n(V^1_{n,\psi}/\sigma_{n,\psi})/G_n(x_0)$ are ADF for testing that $m_\psi(x) = m(x, \theta_0)$, $-\infty < x \le x_0$, where G_n is the empirical d.f. based on $\{X_0, X_1, \dots, X_{n-1}\}$. A similar approach has been also used in an analogous problem that arises in model checking for a regression model by Stute, Thies and Zhu (1998). The restriction of the testing domain to the interval $(-\infty, x_0]$ is not a serious constraint from the applications point of view. Often one tests the more restrictive hypothesis that $m_\psi(x) = m(x, \theta_0)$, for all x in a bounded interval $[a, b]$ of $\mathbb{R}$. Modification of the above tests for this problem is given in Remark 2.3 below. They continue to be ADF.

None of the proofs require the underlying process $\{X_i\}$ to have any type of mixing property. Rather our arguments are based on a general invariance principle for marked empirical processes, which may be of interest on its own and which is formulated and proved in Section 3. The other proofs are deferred to Section 4.

We now summarize our results for the two interesting ψ's, namely, $\psi(x) \equiv x$ and $\psi(x) \equiv I(x > 0) - (1-\alpha)$ in the case the innovations $\{\varepsilon_i := X_i - m_\psi(X_{i-1})\}$ are assumed to be i.i.d. according to some d.f. having a uniformly bounded Lebesgue density. Let $V_{n,I}(V^1_{n,I})$ and $V_{n,\alpha}(V^1_{n,\alpha})$ denote the corresponding $V_{n,\psi}(V^1_{n,\psi})$ processes, respectively. Then, $\mathbb{E}\varepsilon_1^{4(1+\delta)} < \infty$ and $\mathbb{E}X_0^{2(1+\delta)} < \infty$, for some $\delta > 0$, suffice for the weak convergence of $V_{n,I}$ while $\mathbb{E}|X_0|^{1+\delta} < \infty$, for some $\delta > 0$, suffices for that of $V_{n,\alpha}$. See also Theorem 2.1(ii) and Remark 2.1 for some alternate conditions for the latter result. Moreover, if θ_n is taken to be the least square estimator, then any test based on a continuous function of $T_n(V^1_{n,I})/\{\sigma_{n,I} G_n(x_0)\}$ is ADF for testing that the first-order autoregressive mean function is $m(\cdot, \theta_0)$ on the interval $(-\infty, x_0]$. Similarly, tests based on $T_n(V^1_{n,0.5})/\{\sigma_{n,0.5} G_n(x_0)\}$ with θ_n equal to the least absolute deviation estimator are ADF for testing that the first-order autoregressive median function is $m(\cdot, \theta_0)$ on the interval $(-\infty, x_0]$. Of course, all these results hold under some additional smoothness conditions on $\mathscr{M}$ as given in Theorem 2.4 below.

We end this section with some historical remarks. An and Bing (1991) have proposed the K–S test based on $V^1_{n,I}$ and a half sample splitting technique a la Rao (1972) and Durbin (1973) to make it ADF for testing that a time series is linear autoregressive. This method typically leads to a loss of power. Su and Wei (1991) proposed the K–S test based on the $V^1_{n,I}$-process to test for fitting a generalized linear regression model. Delgado (1993) constructed two sample-type tests based on the $V_{n,I}$ for comparing two regression models. Diebolt (1995) has obtained the Hungarian-type strong approximation result for the analogue of $V_{n,I}$ in a special regression setting. Stute (1997) has investigated the large sample theory of the analogue of $V^1_{n,I}$ for model checking in a general regression setting. He also gave a nonparametric principal component analysis of the limiting process in a linear regression setup similar to the one given by Durbin, Knott and Taylor (1975) in the one-sample setting.

2. Main results. This section discusses the asymptotic behavior of the processes introduced in the previous section. Then a transformation T and its estimate T_n are given so that the processes $TV^1_{n,\psi}$ and $T_n V^1_{n,\psi}$ have the same weak limit with a known distribution. Consequently, the tests based on the processes $T_n V^1_{n,\psi}$ are ADF. This section also discusses some applications and provides an argument for the consistency of the proposed tests. Recall that

$$F_y(x) := \mathbb{P}(X_1 - m_\psi(X_0) \le x \,|\, X_0 = y), \qquad x, y \in \mathbb{R}.$$

Let $\varepsilon_i := X_i - m_\psi(X_{i-1}),\ i = 0, \pm 1, \pm 2, \ldots$. We are ready to state our first result.

THEOREM 2.1. *Assume that* (1.1) *and* (1.2) *hold. Then all finite-dimensional distributions of* $V_{n,\psi}$ *converge weakly to those of a centered continuous Gaussian process* V_ψ *with the covariance function* K_ψ.

(i) *Suppose, in addition, that for some* $\eta > 0$, $\delta > 0$,

$$\mathbb{E}\psi^4(\varepsilon_1) < \infty, \tag{2.1)(a}$$

$$\mathbb{E}\psi^4(\varepsilon_1)|X_0|^{1+\eta} < \infty, \tag{2.1)(b}$$

$$\mathbb{E}\{\psi^2(\varepsilon_2)\psi^2(\varepsilon_1)|X_1|\}^{1+\delta} < \infty \tag{2.1)(c}$$

and that the family of d.f.'s $\{F_y,\ y \in \mathbb{R}\}$ *have Lebesgue densities* $\{f_y,\ y \in \mathbb{R}\}$ *that are uniformly bounded*,

$$\sup_{x,y} f_y(x) < \infty. \tag{2.2}$$

Then

$$V_{n,\psi} \Rightarrow V_\psi \quad \textit{in the space } D[-\infty, \infty]. \tag{2.3}$$

(ii) *Instead of* (2.1) *and* (2.2), *suppose that* ψ *is bounded and the family of d.f.'s* $\{F_y,\ y \in \mathbb{R}\}$ *have Lebesgue densities* $\{f_y,\ y \in \mathbb{R}\}$ *satisfying*

$$\int \left[\mathbb{E}\{f_{X_0}^{1+\delta}(x - m_\psi(X_0))\}\right]^{1/(1+\delta)} dx < \infty, \tag{2.4}$$

for some $\delta > 0$. *Then also* (2.3) *holds*.

REMARK 2.1. Conditions (2.1) and (2.2) are needed to ensure the uniform tightness in the space $D[-\infty, \infty]$ for a general ψ while (2.4) suffices for a bounded ψ. Condition (2.1) is satisfied when, as is assumed in most standard time series models, the innovations are independent of the past, and when for some $\delta > 0$, $\mathbb{E}\psi^{4(1+\delta)}(\varepsilon_1)$ and $\mathbb{E}X_1^{2(1+\delta)}$ are finite. Moreover, in this situation the conditional distributions do not depend on y, so that (2.2) amounts to assuming that the density of ε_1 is bounded. In the case of bounded ψ, $\mathbb{E}|X_1|^{1+\delta} < \infty$, for some $\delta > 0$, implies (2.1).

Now consider the assumption (2.4). Note that the stationary distribution G has Lebesgue density $g(x) \equiv \mathbb{E}f_{X_0}(x - m_\psi(X_0))$. This fact together with (2.2) implies that the left-hand side of (2.4) is bounded from the above by a constant $C := [\sup_{x,y} f_y(x)]^{\delta/(1+\delta)}$ times

$$\int \left[\mathbb{E}f_{X_0}(x - m_\psi(X_0))\right]^{1/(1+\delta)} dx = \int g^{1/(1+\delta)}(x)\, dx.$$

Thus, (2.4) is implied by assuming

$$\int g^{1/(1+\delta)}(x)\, dx < \infty.$$

Alternately, suppose m_ψ is bounded and that $f_y(x) \le f(x)$, $\forall\ x, y \in \mathbb{R}$, where f is a bounded and unimodal Lebesgue density on $\mathbb{R}$. Then also the left-hand side of (2.4) is finite. One thus sees that in the particular case of the i.i.d. errors, (2.4) is satisfied for either all bounded error densities and for all stationary densities that have an exponential tail or for all bounded unimodal error densities in the case of bounded m_ψ. Summarizing, we see that (2.1), (2.2) and (2.4) are fulfilled in many models under standard assumptions on the relevant densities and moments.

We shall now turn to the asymptotic behavior of the $V^1_{n,\psi}$ process under H_0. To that effect, the following minimal additional regularity conditions on the underlying entities will be needed. The introduced quantities will be part of the approximating process and its limit covariance and are therefore indispensable. For technical reasons, the case of a smooth ψ [see (Ψ_1) below] and a nonsmooth ψ [see (Ψ_2) below] are dealt with separately. All probability statements in these assumptions are understood to be made under H_0. The d.f. of X_0 will be now denoted by G_{θ_0}.

(A1) The estimator θ_n satisfies

$$n^{1/2}(\theta_n - \theta_0) = n^{-1/2}\sum_{i=1}^{n} l(X_{i-1}, X_i, \theta_0) + o_{\mathbb{P}}(1)$$

for some q-vector valued function l such that $\mathbb{E}\{l(X_0, X_1, \theta_0)|X_0\} = 0$ and

$$L(\theta_0) := \mathbb{E}\{l(X_0, X_1, \theta_0) l^T(X_0, X_1, \theta_0)\}$$

exists and is positive definite.

(A2) There exists a function $\dot m$ from $\mathbb{R}\times\Theta$ to $\mathbb{R}^q$ such that $\dot m(\cdot, \theta_0)$ is measurable and satisfies the following: for all $k < \infty$,

$$\sup_{1\le i\le n,\, n^{1/2}\|t-\theta_0\|\le k} n^{1/2}\big|m(X_{i-1}, t) - m(X_{i-1}, \theta_0) - (t-\theta_0)^T \dot m(X_{i-1}, \theta_0)\big| = o_{\mathbb{P}}(1) \tag{2.5}$$

and

$$\mathbb{E}\|\dot m(X_0, \theta_0)\|^2 < \infty. \tag{2.6}$$

(Ψ_1) (Smooth ψ). The function ψ is absolutely continuous with its almost everywhere derivative $\dot\psi$ bounded and having right and left limits.

(Ψ_2) (Nonsmooth ψ). The function ψ is nondecreasing, right continuous, bounded and such that the function

$$x \mapsto \mathbb{E}\big[\psi(X_1 - m(X_0, \theta_0) + x) - \psi(X_1 - m(X_0, \theta_0))\big]^2$$

is continuous at 0.

(F) The family of d.f.'s $\{F_y,\ y \in \mathbb{R}\}$ has Lebesgue densities $\{f_y,\ y \in \mathbb{R}\}$ that are equicontinuous: for every $\alpha > 0$ there exists a $\delta > 0$ such that

$$\sup_{y\in\mathbb{R},\, |x-z|<\delta} |f_y(x) - f_y(z)| \le \alpha.$$

Note that (2.6) and (Ψ_1) or (2.6), (Ψ_2) and (F) imply that the vectors of functions

$$\dot{M}(x,\theta_0) = (\dot{M}_1(x,\theta_0),\dots,\dot{M}_q(x,\theta_0))^T,$$
$$\dot{\Gamma}(x,\theta_0) = (\dot{\Gamma}_1(x,\theta_0),\dots,\dot{\Gamma}_q(x,\theta_0))^T,$$

with

$$\dot{M}_j(x,\theta_0) = \mathbb{E}\dot{m}_j(X_0,\theta_0)\dot{\psi}(X_1 - m(X_0,\theta_0))I(X_0 \le x),$$
$$\dot{\Gamma}_j(x,\theta_0) = \mathbb{E}\dot{m}_j(X_0,\theta_0)\int f_{X_0}\,d\psi\, I(X_0 \le x), \qquad 1 \le j \le q,\ x \in \mathbb{R},$$

are well defined. We are now ready to formulate an expansion of $V^1_{n,\psi}$, which is crucial for the subsequent results.

THEOREM 2.2. *Assume that* (1.1), (A1), (A2) *and* H_0 *hold.*

(a) *If, in addition* (Ψ_1) *holds, then*

$$\sup_{x\in\mathbb{R}}\left|V^1_{n,\psi}(x) - V_{n,\psi}(x) + \dot{M}^T(x,\theta_0)\,n^{-1/2}\sum_{i=1}^n l(X_{i-1},X_i,\theta_0)\right| = o_{\mathbb{P}}(1). \tag{2.7}$$

(b) *Assume, in addition, that* (Ψ_2) *and* (F) *hold and that either* $\mathbb{E}|X_0|^{1+\delta} < \infty$, *for some* $\delta > 0$ *and* (2.2) *holds or* (2.4) *holds. Then the conclusion* (2.7) *with* $\dot{M}$ *replaced by* $\dot{\Gamma}$ *continues to hold.*

REMARK 2.2. The assumption (Ψ_1) covers many interesting ψ's including the least square score $\psi(x) \equiv x$ and the Huber score $\psi(x) \equiv xI(|x| \le c) + c\,\mathrm{sign}(x)I(|x| > c)$, where c is a real constant, while (Ψ_2) covers the α-quantile score $\psi(x) \equiv I(x > 0) - (1-\alpha)$.

The following corollary is an immediate consequence of Theorems 2.1 and 2.2. We shall state it for the smooth ψ-case only. The same holds in the nonsmooth case with $\dot{M}$ replaced by $\dot{\Gamma}$.

COROLLARY 2.1. *Under the assumptions of Theorems* 2.1 *and* 2.2(a),

$$V^1_{n,\psi} \Rightarrow V^1_{\psi} \quad \textit{in the space } D[-\infty,\infty],$$

where V^1_ψ *is a centered continuous Gaussian process with the covariance function*

$$\begin{aligned}K^1_\psi(x,y) = K_\psi(x,y) &+ \dot{M}^T(x,\theta_0)L(\theta_0)\dot{M}(y,\theta_0)\\ &- \dot{M}^T(x,\theta_0)\mathbb{E}\{I(X_0 \le y)\psi(X_1 - m(X_0,\theta_0))l(X_0,X_1,\theta_0)\}\\ &- \dot{M}^T(y,\theta_0)\mathbb{E}\{I(X_0 \le x)\psi(X_1 - m(X_0,\theta_0))l(X_0,X_1,\theta_0)\}.\end{aligned}$$

The above complicated-looking covariance function can be further simplified if we choose θ_n to be related to the function ψ in the following fashion. Recall from the previous section that $\sigma_\psi^2(x) = \mathbb{E}[\psi^2(\varepsilon_1)|X_0 = x]$ and let, for $x \in \mathbb{R}$,

$$\gamma_\psi(x) := \mathbb{E}[\dot\psi(X_1 - m(X_0, \theta_0))|X_0 = x] \quad \text{for smooth } \psi,$$
$$:= \int f_x(y)\,\psi(dy) \quad \text{for nonsmooth } \psi.$$

Assume that

$$\begin{aligned} \sigma_\psi^2(x) &= \sigma_\psi^2 \quad \text{a positive constant in } x, \text{ a.s.,} \\ \gamma_\psi(x) &= \gamma_\psi \quad \text{a positive constant in } x, \text{ a.s.,} \end{aligned} \tag{2.8}$$

and that θ_n satisfies (A1) with

$$l(x, y, \theta_0) = \gamma_\psi^{-1}(\Sigma_{\theta_0})^{-1}\dot m(x, \theta_0)\psi(y - m(x, \theta_0)), \qquad x, y \in \mathbb{R}, \tag{2.9}$$

where $\Sigma_{\theta_0} := \mathbb{E}\dot m(X_0, \theta_0)\dot m^T(X_0, \theta_0)$ so that $L(\theta_0) = \tau\Sigma_{\theta_0}^{-1}$, with $\tau := \sigma_\psi^2/\gamma_\psi^2$. Then direct calculations show that the above covariance function simplifies to

$$K_\psi^1(x, y) = \mathbb{E}\psi^2(X_1 - m(X_0, \theta_0))\,[G_{\theta_0}(x \wedge y) - \nu^T(x)\Sigma_{\theta_0}^{-1}\nu(y)], \tag{2.10}$$

with

$$\nu(x) = \mathbb{E}\dot m(X_0, \theta_0)\,I(X_0 \le x), \qquad x, y \in \mathbb{R}.$$

Suppose $\{\varepsilon_i = X_i - m(X_{i-1}, \theta_0)\}$ are i.i.d. and ε_i is independent of X_{i-1} for all i. Then (2.8) is satisfied a priori and Koul (1996) gives a set of sufficient conditions on the model $\mathscr{M}$ under which a class of M-estimators of θ_0 corresponding to a bounded ψ defined by the relation

$$\theta_{n,\psi} := \arg\min_t \left\| n^{-1/2}\sum_{i=1}^n \dot m(X_{i-1}, t)\psi(X_i - m(X_{i-1}, t))\right\|$$

satisfies (2.9). See also Tjøstheim (1986) for a similar result for the least square estimator.

Unlike (1.3), the structure of K_ψ^1 given at (2.10) still does not allow for a simple representation of V_ψ^1 in terms of a process with known distribution. The situation is similar to the model checking for the underlying distribution in the classical i.i.d. setup. We shall now describe a transformation so that the limiting distribution of the transformed $V_{n,\psi}^1$-process is known. The exposition here is based on the ideas of Khmaladze (1981) and Nikabadze and Stute (1996).

Throughout the rest of the section we shall assume that (2.8) holds. To simplify the exposition further, assume without loss of generality that $\mathbb{E}\psi^2(\varepsilon_1) = 1$. Write $\dot m(\cdot) = \dot m(\cdot, \theta_0)$, $G = G_{\theta_0}$. Set

$$A(x) = \int \dot m(y)\dot m^T(y)\; I(y \ge x)G(dy), \qquad x \in \mathbb{R}.$$

Assume that

$$A(x_0) \quad \text{is nonsingular for some } x_0 < \infty. \tag{2.11}$$

This and the nonnegative definiteness of $A(y) - A(x_0)$ implies that $A(y)$ is nonsingular for all $y \le x_0$. Write $A^{-1}(y)$ for $(A(y))^{-1}$, $y \le x_0$, and define

$$Tf(x) = f(x) - \int_{-\infty}^{x} \dot m^T(y) A^{-1}(y) \left[\int \dot m(z)\ I(z \ge y)\ f(dz) \right] G(dy), \qquad x \le x_0.$$

We will apply T to functions f which are either of bounded variation or Brownian motion. In the latter case the inner integral needs to be interpreted as a stochastic integral. Since T is a linear operator, $T(V_\psi)$ is a centered Gaussian process. Moreover we have the following fact.

LEMMA 2.1.

$$\mathrm{Cov}[TV_\psi(x), TV_\psi(y)] = G(x \wedge y), \qquad x, y \in \mathbb{R},$$

that is, TV_ψ *is a Brownian motion with respect to time* G.

The proof uses the independence of increments of the Brownian motion V_ψ and properties of stochastic integrals and is similar to that of Lemma 3.1 of Stute, Thies and Zhu (1998). We are now ready to state our next result.

THEOREM 2.3. (a) *Assume, in addition to the assumptions of Theorem* 2.2(a), *that* (2.8) *and* (2.11) *hold. Then*

$$\sup_{x \le x_0} \left| TV^1_{n,\psi}(x) - TV_{n,\psi}(x) \right| = o_{\mathbb{P}}(1). \tag{2.12}$$

If in addition, (1.2), (2.1) *and* (2.2) *hold, then*

$$TV_{n,\psi} \Rightarrow TV_\psi \quad \text{and} \quad TV^1_{n,\psi} \Rightarrow TV_\psi \text{ in } D[-\infty, x_0]. \tag{2.13}$$

(b) *The above claims continue to hold under the assumptions of Theorem* 2.2(b), (2.8) *and* (2.11).

The usefulness of the above theorem in statistics is limited because T is known only in theory. For statistical applications to the goodness-of-fit testing, one needs to obtain an analogue of the above theorem where T is replaced by an estimator T_n. Let, for $x \in \mathbb{R}$,

$$G_n(x) := n^{-1} \sum_{i=1}^{n} I(X_{i-1} \le x)$$

and

$$A_n(x) := \int \dot m(y, \theta_n) \dot m^T(y, \theta_n)\ I(y \ge x)\ G_n(dy).$$

Define an estimator of T to be

$$T_n f(x) = f(x) - \int_{-\infty}^{x} \dot m^T(y, \theta_n) A_n^{-1}(y) \times \left[\int \dot m(z, \theta_n)\ I(z \ge y)\ f(dz) \right] G_n(dy), \qquad x \le x_0.$$

The next result, one of the main results of the paper, proves the consistency of $T_n V^1_{n,\psi}$ for $TV^1_{n,\psi}$ under the following additional smoothness condition on $\dot{m}$. For some $q \times q$ square matrix $\ddot{m}(x, \theta_0)$ and a nonnegative function $K_1(x, \theta_0)$, both measurable in the x-coordinate, the following holds:

$$\begin{aligned} &\mathbb{E}\|\dot{m}(X_0, \theta_0)\|^j K_1(X_0, \theta_0) < \infty, \\ &\mathbb{E}\|\ddot{m}(X_0, \theta_0)\| \|\dot{m}(X_0, \theta_0)\|^j < \infty, \qquad j = 0, 1, \end{aligned} \tag{2.14}$$

and $\forall\, \epsilon > 0$, there exists a $\delta > 0$ such that $\|\theta - \theta_0\| < \delta$ implies

$$\begin{aligned} &\|\dot{m}(x, \theta) - \dot{m}(x, \theta_0) - \ddot{m}(x, \theta_0)(\theta - \theta_0)\| \\ &\qquad \le \epsilon\ K_1(x, \theta_0)\, \|\theta - \theta_0\| \quad \text{for } G\text{-almost all } x. \end{aligned}$$

Here, and in the sequel, for a $q \times q$ real matrix D, $\|D\| := \sup\{(a' DD' a)^{1/2};\ a \in \mathbb{R}^q,\ \|a\| = 1\}$. The assumption (2.14) is neccessary because we are now approximating the entities involving the estimator θ_n. We are ready to state the theorem.

THEOREM 2.4. (a) *Suppose, in addition to the assumptions of Theorem* 2.3(a), (2.14) *holds and that* (2.1) *with* $\psi(\varepsilon_1)$, $\psi(\varepsilon_2)$ *replaced by* $\|\dot{m}(X_0, \theta_0)\|\ \psi(\varepsilon_1)$, $\|\dot{m}(X_1, \theta_0)\|\ \psi(\varepsilon_2)$, *respectively*, *holds*. *Then*

$$\sup_{x \le x_0} |T_n V^1_{n,\psi}(x) - TV_{n,\psi}(x)| = o_{\mathbb{P}}(1), \tag{2.15}$$

and consequently,

$$\sigma^{-1}_{n,\psi}\, T_n V^1_{n,\psi}(\cdot) \Rightarrow B \circ G \quad \text{in } D[-\infty, x_0]. \tag{2.16}$$

(b) *The same continues to hold under the assumptions of Theorem* 2.3(b) *and* (2.14).

REMARK 2.3. By (2.11), $\lambda_1 := \inf\{a' A(x_0) a;\ a \in \mathbb{R}^q,\ \|a\| = 1\} > 0$ and $A(x)$ is positive definite for all $x \le x_0$, Hence, $\|A^{-1/2}(x)\|^2 \le \lambda_1^{-1} < \infty$, for all $x \le x_0$, and (2.6) implies

$$\mathbb{E}\|\dot{m}^T(X_0) A^{-1}(X_0)\| I(X_0 \le x_0) \le \mathbb{E}\|\dot{m}(X_0)\|\, \lambda_1^{-1} < \infty. \tag{2.17}$$

This fact is used in the proofs repeatedly.

Now, let $a < b$ be given real numbers and suppose one is interested in testing the hypothesis

$$H\colon m_\psi(x) = m(x, \theta_0) \quad \text{for all } x \in [a, b] \text{ and for some } \theta_0 \in \Theta.$$

Assume the support of G is $\mathbb{R}$, and $A(b)$ is positive definite. Then, $A(x)$ is nonsingular for all $x \le b$, continuous on $[a, b]$ and $A^{-1}(x)$ is continuous on $[a, b]$ and

$$\mathbb{E}\|\dot{m}^T(X_0) A^{-1}(X_0)\| I(a < X_0 \le b) < \infty.$$

Thus, under the conditions of Theorem 2.4, $\sigma_{n,\psi}^{-1} T_n V_{n,\psi}^1(\cdot) \Rightarrow B\circ G$, in $D[a,b]$ and we obtain

$$\sigma_{n,\psi}^{-1}\{T_n V_{n,\psi}^1(\cdot) - T_n V_{n,\psi}^1(a)\} \Rightarrow B(G(\cdot)) - B(G(a)) \quad \text{in } D[a,b].$$

The stationarity of the increments of the Brownian motion then readily implies that

$$D_n := \sup_{a\le x\le b} \frac{|T_n V_{n,\psi}^1(x) - T_n V_{n,\psi}^1(a)|}{\{G_n(b)-G_n(a)\}\,\sigma_{n,\psi}} \Rightarrow \sup_{0\le u\le 1}|B(u)|.$$

Hence, any test of H based on D_n is ADF.

Applications. Here we discuss some examples of nonlinear time series to which the above results may be applied. It may be useful for computational purposes to rewrite $T_n V_{n,\psi}^1$ as follows: for all $x \le x_0$,

$$\begin{aligned} T_n V_{n,\psi}^1(x) = n^{-1/2}\sum_{i=1}^n \Big[I(X_{i-1}\le x) - n^{-1}\sum_{j=1}^n \dot m^T(X_{j-1},\theta_n) \\ \times A_n^{-1}(X_{j-1})\dot m(X_{i-1},\theta_n)\, I(X_{j-1}\le X_{i-1}\wedge x)\Big] \\ \times \psi(X_i - m(X_{i-1},\theta_n)). \end{aligned} \tag{2.18}$$

Now, let $g_1,\ldots,g_q$ be known real-valued G-square integrable functions on $\mathbb{R}$ and consider the class of models $\mathscr{M}$ with

$$m(x,\theta) = g_1(x)\theta_1 + \cdots + g_q(x)\theta_q.$$

Then (A1), (2.14) are trivially satisfied with $\dot m(x,\theta) \equiv (g_1(x),\ldots,g_q(x))^T$ and $\ddot m(x,\theta)\equiv 0 \equiv K_1(x,\theta)$. Besides including the first-order linear autoregressive [AR(1)] model where $q=1$, $g_1(x)\equiv x$, this class also includes some nonlinear autoregressive models. For example, the choice of $q=2$, $g_1(x)=x$, $g_2(x) = x\exp(-x^2)$ gives an exponential-amplitude dependent AR(1) [EXPAR(1)] model of Ozaki and Oda (1978) [Tong (1990), pages 129 and 130]. In the following discussion, the innovations $\varepsilon_i = X_i - m(X_{i-1},\theta_0)$ are assumed to be i.i.d. according to a d.f. F with a bounded Lebesgue density f.

In the linear AR(1) model $\dot m(x,\theta)\equiv x$ and $A(x)\equiv \mathbb{E}X_0^2 I(X_0\ge x)$ is positive for all real x, uniformly continuous and decreasing on $\mathbb{R}$, and thus trivially satisfies (2.11). A uniformly a.s. consistent estimator of A is

$$A_n(x) \equiv n^{-1}\sum_{k=1}^n X_{k-1}^2\, I(X_{k-1}\ge x).$$

Thus a test of the hypothesis that the first-order autoregressive mean function is linear AR(1) on the interval $(-\infty, x_0]$ can be based on

$$\sup_{x\le x_0} |T_n V_{n,I}^1(x)|/\{\sigma_{n,I} G_n(x_0)\},$$

where

$$T_n V^1_{n,I}(x) = n^{-1/2} \sum_{i=1}^{n} \Bigg[I(X_{i-1} \le x) - n^{-1} \sum_{j=1}^{n} \frac{X_{j-1} X_{i-1} I(X_{j-1} \le X_{i-1} \wedge x)}{n^{-1} \sum_{k=1}^{n} X^2_{k-1} I(X_{k-1} \ge X_{j-1})} \Bigg] \times (X_i - X_{i-1}\theta_n),$$

$$\sigma^2_{n,I} = n^{-1} \sum_{i=1}^{n} (X_i - X_{i-1}\theta_n)^2.$$

Similarly, a test of the hypothesis that the first-order autoregressive median function is AR(1) can be based on $\sup_{x \le x_0} |T_n V^1_{n,0.5}(x)|/\{\sigma_{n,0.5} G_n(x_0)\}$, where

$$T_n V^1_{n,0.5}(x) = n^{-1/2} \sum_{i=1}^{n} \Bigg[I(X_{i-1} \le x) - n^{-1} \sum_{j=1}^{n} \frac{X_{j-1} X_{i-1} I(X_{j-1} \le X_{i-1} \wedge x)}{n^{-1} \sum_{k=1}^{n} X^2_{k-1} I(X_{k-1} \ge X_{j-1})} \Bigg] \times \{I(X_i - X_{i-1}\theta_n > 0) - 0.5\}$$

and

$$\sigma^2_{n,0.5} = n^{-1} \sum_{i=1}^{n} \{I(X_i - X_{i-1}\theta_n > 0) - 0.5\}^2.$$

By Theorem 2.4, both of these tests are ADF as long as the estimator θ_n is the least square (LS) estimator in the former test and the least absolute deviation (LAD) estimator in the latter. For the former test we additionally require $\mathbb{E}\varepsilon_1^{4(1+\delta)} < \infty$, for some $\delta > 0$, while for the latter test, $\mathbb{E}\varepsilon_1^2 < \infty$ and f being uniformly continuous and positive suffice.

In the EXPAR(1) model, $\dot m(x, \theta_0) \equiv (x,\ x\exp(-x^2))^T$ and $A(x)$ is the 2×2 symmetric matrix

$$A(x) = \mathbb{E} I(X_0 \ge x) \begin{pmatrix} X_0^2 & X_0^2 \exp(-X_0^2) \\ X_0^2 \exp(-X_0^2) & X_0^2 \exp(-2X_0^2) \end{pmatrix}.$$

From Theorem 4.3 of Tong [(1990), page 128], if $\mathbb{E}\varepsilon_1^4 < \infty$, f is absolutely continuous and positive on $\mathbb{R}$, then the above EXPAR(1) process is stationary, ergodic, the corresponding stationary d.f. G is strictly increasing on $\mathbb{R}$, and $\mathbb{E}X_0^4 < \infty$. Moreover, one can directly verify that $\mathbb{E}X_0^2 < \infty$ implies $A(x)$ is nonsingular for every real x and A^{-1} and A are continuous on $\mathbb{R}$. The matrix

$$A_n(x) = n^{-1} \sum_{i=1}^{n} I(X_{i-1} \ge x) \begin{pmatrix} X_{i-1}^2 & X_{i-1}^2 \exp(-X_{i-1}^2) \\ X_{i-1}^2 \exp(-X_{i-1}^2) & X_{i-1}^2 \exp(-2X_{i-1}^2) \end{pmatrix}$$

provides a uniformly a.s. consistent estimator of $A(x)$.

Thus one may use $\sup_{x \le x_0} |T_n V^1_{n,I}(x)|/\{\sigma_{n,I} G_n(x_0)\}$ to test the hypothesis that the autoregressive mean function is given by an EXPAR(1) function on an interval $(-\infty, x_0]$. Similarly, one can use the test statistic $\sup_{x\le x_0}|T_n V^1_{n,0.5}(x)|/\{\sigma_{n,0.5}G_n(x_0)\}$ to test the hypothesis that the autoregressive median function is given by an EXPAR(1) function. In both cases, A_n is as above and one should now use the general formula (2.18) to compute these statistics. Again, from Theorem 2.4 it readily follows that the asymptotic levels of both of these tests can be computed from the distribution of $\sup_{0\le u\le 1}|B(u)|$, provided the estimator θ_n is taken to be, respectively, the LS and the LAD. Again one needs the $(4+\delta)$th moment assumption for the former test and the uniform continuity of f for the latter test. The relevant asymptotics of the LS-estimator and a class of M-estimators with bounded ψ in a class of nonlinear time series models is given in Tjøstheim (1986) and Koul (1996), respectively. In particular, these papers include the above EXPAR(1) model.

REMARK 2.4. Theorems 2.1 and 2.2 can be extended to the case where X_{i-1} is replaced by a q-vector $(X_{i-1},\dots,X_{i-q})^T$ in the definitions of the $V_{n,\psi}$ and $V^1_{n,\psi}$. In this case, the time parameter of these processes is a q-dimensional vector. The difficulty in transforming such processes to obtain a limiting process that has a known limiting distribution is similar to that faced in transforming the multivariate empirical process in the i.i.d. setting. This, in turn, is related to the difficulty of having a proper definition of a multitime parameter martingale. See Khmaladze (1988, 1993) for a discussion on the issues involved. For these reasons, we restricted our attention here to the one-dimensional case only.

Consistency. Here we shall give sufficient conditions that will imply the consistency of goodness-of-fit tests based on $V_{n,\psi}$ for a simple hypothesis $m_\psi = m_0$ against the fixed alternative $m_\psi \ne m_0$, where m_0 is a known function. By the statement $m_\psi \ne m_0$ it should be understood that the G-measure of the set $\{y\in\mathbb{R};\ m_\psi(y)\ne m_0(y)\}$ is positive. Let $\lambda(y,z) := \mathbb{E}\{\psi(X_1 - m_\psi(X_0) + z)|X_0 = y\}$, $y, z\in\mathbb{R}$. Note that ψ nondecreasing implies that $\lambda(y,z)$ is nondecreasing in z, for each real y. Assume that for every $y\in\mathbb{R}$,

$$\lambda(y,z) = 0 \quad \text{if and only if } z = 0. \tag{2.19}$$

Let $d(x) := m_\psi(x) - m_0(x)$, $x\in\mathbb{R}$ and

$$D_n(x) := n^{-1/2}\sum_{i=1}^n \lambda(X_{i-1}, d(X_{i-1}))\, I(X_{i-1}\le x), \qquad x\in\mathbb{R}.$$

An adaptation of the Glivenko–Cantelli arguments known for the i.i.d. case to the strictly stationary case [see (4.1) below] yields

$$\sup_{x\in\mathbb{R}}|n^{-1/2}D_n(x) - \mathbb{E}\lambda(X_0, d(X_0))I(X_0\le x)| \to 0 \quad \text{a.s.},$$

where the $\mathbb{E}$ is computed under the alternative m_ψ. Moreover, by Lemma 3.1 below, we have

$$V_{n,\psi} \Rightarrow \text{a continuous Gaussian process.}$$

These facts together with the assumption (2.19) and a routine argument yield the consistency of the K–S and Cramér–von Mises tests based on $V_{n,\psi}$.

The study of the asymptotic power of the above tests against a sequence of alternatives $m_{n,\psi}(x) \to m_0(x)$ at a $n^{-1/2}$ rate deserves serious attention and will be discussed elsewhere.

3. Weak convergence of a marked empirical process. Let X_i, $i=0, \pm 1, \dots$ be a strictly stationary ergodic Markov process with stationary d.f. G, and let $\mathscr{F}_i = \sigma(X_i, X_{i-1}, \dots)$ be the σ-field generated by the observations obtained up to time i. Furthermore, let, for each $n \ge 1$, $\{Z_{n,i},\ 1 \le i \le n\}$ be an array of r.v.'s adapted to $\{\mathscr{F}_i\}$ such that $(Z_{n,i}, X_i)$ is strictly stationary in i, for each $n \ge 1$, and satisfying

$$\mathbb{E}\{Z_{n,i} | \mathscr{F}_{i-1}\} = 0, \qquad 1 \le i \le n. \tag{3.1}$$

Our goal here is to establish the weak convergence of the process

$$\alpha_n(x) = n^{-1/2} \sum_{i=1}^{n} Z_{n,i} I(X_{i-1} \le x), \qquad x \in \mathbb{R}. \tag{3.2}$$

The process α_n constitutes a marked empirical process of the X_{i-1}'s, the marks being given by the martingale difference array $\{Z_{n,i}\}$. An example of this process is the $V_{n,\psi}$ process where $Z_{n,i} \equiv Z_i = \psi(X_i - m_\psi(X_{i-1}))$. Further examples appear in the proofs in Section 4 below.

We now formulate the assumptions that guarantee the weak convergence of α_n to a continuous limit in $D[-\infty, \infty]$. To this end, let

$$L_y(x) := \mathbb{P}(X_1 - \varphi(X_0) \le x | X_0 = y), \qquad x, y \in \mathbb{R},$$

where φ is a real-valued measurable function. Because of the Markovian assumption, L_{X_0} is the d.f. of $X_1 - \varphi(X_0)$, given $\mathscr{F}_0$. For example, if $\{X_i\}$ is integrable, we may take $\varphi(x) \equiv \mathbb{E}[X_{i+1} \,|\, X_i = x]$, so that $\varepsilon_{i+1} := X_{i+1} - \varphi(X_i)$ are just the innovations generating the process $\{X_i\}$. As another example, for the $V_{n,\psi}$, we may take $\varphi = m_\psi$. In the context of time series analysis the innovations are often i.i.d., in which case L_y does not depend on y. However, for our general result of this section, we may let L_y depend on y.

This section contains two lemmas. Lemma 3.1 deals with the general marks and Lemma 3.2 with the bounded marks. The following assumptions will be needed in Lemma 3.1.

(A) For some $\eta > 0$, $\delta > 0$, $K < \infty$ and all n sufficiently large,

$$\mathbb{E} Z_{n,1}^4 \le K, \tag{a}$$

$$\mathbb{E} Z_{n,1}^4 |X_0|^{1+\eta} \le K \tag{b}$$

and

$$\mathbb{E}\big\{Z_{n,2}^2 Z_{n,1}^2 |X_1|\big\}^{1+\delta} \le K. \tag{c}$$

(B) There exists a function φ from $\mathbb{R}$ to $\mathbb{R}$ such that the corresponding family of functions $\{L_y,\ y \in \mathbb{R}\}$ admit Lebesgue densities l_y which are uniformly bounded,

$$\sup_{x,\,y} l_y(x) \le K < \infty.$$

(C) There exists a continuous nondecreasing function τ^2 on $\mathbb{R}$ to $[0, \infty)$ such that

$$n^{-1} \sum_{i=1}^{n} \mathbb{E}[Z_{n,i}^2 | \mathscr{F}_{i-1}] I(X_{i-1} \le x) = \tau^2(x) + o_{\mathbb{P}}(1) \qquad \forall\ x \in [-\infty, \infty].$$

Assumptions (A) and (B) are needed to guarantee the continuity of the weak limit and the tightness of the process α_n, while (C) is needed to identify the weak limit. As before, B denotes the Brownian motion on $[0, \infty)$.

LEMMA 3.1. *Under* (A)–(C),

$$\alpha_n \Rightarrow B \circ \tau^2 \quad \textit{in the space } D[-\infty, \infty], \tag{3.3}$$

where $B \circ \tau^2$ is a continuous Brownian motion on $\mathbb{R}$ with respect to time τ^2.

PROOF. For convenience, we shall not now exhibit the dependence of $Z_{n,i}$ on n. Apply the CLT for martingales [Hall and Heyde (1980), Corollary 3.1] to show that the fidis tend to the right limit, under (A)(a) and (C). □

As to tightness, fix $-\infty \le t_1 < t_2 < t_3 \le \infty$ and assume, without loss of generality, that the moment bounds of (A) hold for all $n \ge 1$ with $K \ge 1$. Then we have

$$\begin{aligned}
&[\alpha_n(t_3) - \alpha_n(t_2)]^2 [\alpha_n(t_2) - \alpha_n(t_1)]^2 \\
&\quad = n^{-2} \left[\sum_{i=1}^{n} Z_i I(t_2 < X_{i-1} \le t_3) \right]^2 \left[\sum_{i=1}^{n} Z_i I(t_1 < X_{i-1} \le t_2) \right]^2 \\
&\quad = n^{-2} \sum_{i,j,k,l} U_i U_j V_k V_l,
\end{aligned}$$

where

$$U_i = Z_i I(t_2 < X_{i-1} \le t_3) \quad \text{and} \quad V_i = Z_i I(t_1 < X_{i-1} \le t_2).$$

Now, if the largest index among i, j, k, l is not matched by any other, then $\mathbb{E}\{U_i U_j V_k V_l\} = 0$. Also, since the two intervals $(t_2, t_3]$ and $(t_1, t_2]$ are disjoint, $U_i V_i \equiv 0$. We thus obtain

$$\begin{aligned}
\mathbb{E}\left\{ n^{-2} \sum_{i,\,j,\,k,\,l} U_i U_j V_k V_l \right\} &= n^{-2} \sum_{i,\,j<k} \mathbb{E}\{V_i V_j U_k^2\} \\
&\quad + n^{-2} \sum_{i,\,j<k} \mathbb{E}\{U_i U_j V_k^2\}.
\end{aligned} \tag{3.4}$$

Note that the moment assumption (A)(a) guarantees that the above expectations exist. In this proof, the constant K is a generic constant, which may vary from expression to expression but never depends on n or the chosen t's.

We shall only bound the first sum in (3.4), the second being dealt with similarly. To this end, fix a $2 \le k \le n$ for the time being and write

$$\begin{aligned}
\sum_{i,\, j<k} \mathbb{E}\{V_i V_j U_k^2\} &= \mathbb{E}\biggl\{\biggl(\sum_{i=1}^{k-1} V_i\biggr)^2 U_k^2\biggr\} \\
&= \mathbb{E}\biggl\{\biggl(\sum_{i=1}^{k-1} V_i\biggr)^2 \mathbb{E}(U_k^2|\mathscr{F}_{k-1})\biggr\} \\
&\le 2\mathbb{E}\biggl\{\biggl(\sum_{i=1}^{k-2} V_i\biggr)^2 \mathbb{E}(U_k^2|\mathscr{F}_{k-1})\biggr\} \\
&\quad + 2\mathbb{E}\{V_{k-1}^2 E(U_k^2|\mathscr{F}_{k-1})\}.
\end{aligned} \tag{3.5}$$

The first expectation equals

$$\mathbb{E}\biggl\{\biggl(\sum_{i=1}^{k-2} V_i\biggr)^2 I(t_2 < X_{k-1} \le t_3)\mathbb{E}(Z_k^2|\mathscr{F}_{k-1})\biggr\}. \tag{3.6}$$

Write

$$\mathbb{E}\bigl(Z_k^2|\mathscr{F}_{k-1}\bigr) = r(X_{k-1}, X_{k-2}, \dots)$$

for an appropriate function r. Note that due to stationarity r is the same for each k. Condition on $\mathscr{F}_{k-2}$ and use the Markov property and Fubini's theorem to show that (3.6) is the same as

$$\begin{aligned}
&\mathbb{E}\biggl\{\biggl(\sum_{i=1}^{k-2} V_i\biggr)^2 \int_{t_2}^{t_3} r(x, X_{k-2}, \dots)\, l_{X_{k-2}}(x - \varphi(X_{k-2}))\, dx\biggr\} \\
&\quad = \int_{t_2}^{t_3} \mathbb{E}\biggl\{\biggl(\sum_{i=1}^{k-2} V_i\biggr)^2 r(x, X_{k-2}, \dots)\, l_{X_{k-2}}(x - \varphi(X_{k-2}))\biggr\}\, dx \\
&\quad \le \int_{t_2}^{t_3} \biggl\{\mathbb{E}\biggl(\sum_{i=1}^{k-2} V_i\biggr)^4\biggr\}^{1/2} \bigl\{\mathbb{E}\bigl(r(x, X_{k-2}, \dots)\, l_{X_{k-2}}(x - \varphi(X_{k-2}))\bigr)^2\bigr\}^{1/2}\, dx,
\end{aligned}$$

where the last inequality follows from the Cauchy–Schwarz inequality. Since the V_i's form a centered martingale difference array, Burkholder's inequality [Chow and Teicher (1978), page 384] and the moment inequality yield

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

We also have

$$\mathbb{E}V_1^4 = \mathbb{E}\bigl(Z_1^4\, I(t_1 < X_0 \le t_2)\bigr) = [L_1(t_2) - L_1(t_1)],$$

where

$$L_1(t) = \mathbb{E} Z_1^4 I(X_0 \le t), \qquad -\infty \le t \le \infty.$$

Let

$$L_2(t) = \int_{-\infty}^{t} \left\{ \mathbb{E}\big(r(x, X_{k-2}, \ldots)\; l_{X_{k-2}}(x - \varphi(X_{k-2}))\big)^2 \right\}^{1/2} dx, \qquad -\infty \le t \le \infty.$$

Note that due to stationarity, L_2 is the same for each k. It thus follows that (3.6) is bounded from the above by

$$K\ (k-2)\ [L_1(t_2) - L_1(t_1)]^{1/2}\ [L_2(t_3) - L_2(t_2)]. \tag{3.7}$$

The assumption (A)(a) implies that L_1 is a continuous nondecreasing bounded function on the real line. Clearly, L_2 is also nondecreasing and continuous. We shall now show that $L_2(\infty)$ is finite. For this, let h be a strictly positive continuous Lebesgue density on the real line such that $h(x) \sim |x|^{-1-\eta}$ as $x \to \pm\infty$, where η is as in (A)(b). By Hölder's inequality,

$$L_2(\infty) \le \left[\int_{-\infty}^{\infty} \mathbb{E}\big(r(x, X_{k-2}, \ldots)\; l_{X_{k-2}}(x - \varphi(X_{k-2}))\big)^2 h^{-1}(x)\, dx \right]^{1/2}.$$

Use the assumption (B) to bound one power of $l_{X_{k-2}}$ from the above so that the last integral in turn is less than or equal to

$$K\ \mathbb{E}\big\{ r^2(X_{k-1}, X_{k-2}, \ldots)\ h^{-1}(X_{k-1}) \big\} \le K\ \mathbb{E}\big\{ Z_1^4\ |h^{-1}(X_0)| \big\}.$$

The finiteness of the last expectation follows, however, from assumption (A)(b).

We now bound the second expectation in (3.5). Since Z_{k-1} is measurable w.r.t. $\mathscr{F}_{k-1}$, there exists some function s such that

$$Z_{k-1}^2 = s(X_{k-1}, X_{k-2}, \ldots).$$

Put $u = rs$ with r as before. Then we have, with the δ as in (A)(c),

$$\begin{aligned}
&\mathbb{E}\big\{ V_{k-1}^2 \mathbb{E}(U_k^2 | \mathscr{F}_{k-1}) \big\} \\
&\quad = \mathbb{E}\big\{ I(t_2 < X_{k-1} \le t_3) I(t_1 < X_{k-2} \le t_2)\ u(X_{k-1}, X_{k-2}, \ldots) \big\} \\
&\quad = \int_{t_2}^{t_3} \mathbb{E}\big\{ I(t_1 < X_{k-2} \le t_2)\ u(x, X_{k-2}, \ldots)\ l_{X_{k-2}}(x - \varphi(X_{k-2})) \big\}\, dx \\
&\quad \le [G(t_2) - G(t_1)]^{\delta/(1+\delta)} \\
&\qquad \times \int_{t_2}^{t_3} \mathbb{E}^{1/(1+\delta)} \big\{ u^{1+\delta}(x, X_{k-2}, \ldots) l_{X_{k-2}}^{1+\delta}(x - \varphi(X_{k-2})) \big\}\, dx.
\end{aligned}$$

Put

$$L_3(t) = \int_{-\infty}^{t} \mathbb{E}^{1/(1+\delta)} \big\{ u^{1+\delta}(x, X_{k-2}, \ldots) l_{X_{k-2}}^{1+\delta}(x - \varphi(X_{k-2})) \big\}\, dx, \qquad -\infty \le t \le \infty.$$

Arguing as above, now let q be a positive continuous Lebesgue density on the real line such that $q(x) \sim |x|^{-1-1/\delta}$ as $x \to \pm\infty$. By Hölder's inequality,

$$\begin{aligned} L_3(\infty) &\le \left[\int_{-\infty}^{\infty} \mathbb{E}\{u^{1+\delta}(x, X_{k-2}, \dots)\, l_{X_{k-2}}^{1+\delta}(x - \varphi(X_{k-2}))\} q^{-\delta}(x)\, dx\right]^{1/(1+\delta)} \\ &\le K\, \left\{\mathbb{E}\big(u^{1+\delta}(X_{k-1}, \dots)\, q^{-\delta}(X_{k-1})\big)\right\}^{1/(1+\delta)} \\ &\le K\, \left\{\mathbb{E}\big(Z_2^{2(1+\delta)} Z_1^{2(1+\delta)} q^{-\delta}(X_1)\big)\right\}^{1/(1+\delta)}, \end{aligned}$$

where the last inequality follows from Hölder's inequality applied to conditional expectations. The last expectation is, however, finite by assumption (A)(c). Thus L_3 is also a nondecreasing continuous bounded function on the real line and we obtain

$$\mathbb{E}\big\{V_{k-1}^2 \mathbb{E}\{U_k^2 | \mathscr{F}_{k-1}\}\big\} \le K\, [G(t_2) - G(t_1)]^{\delta/(1+\delta)} [L_3(t_3) - L_3(t_2)].$$

Upon combining this with (3.5) and (3.7) and summing over $k = 2$ to $k = n$, we obtain

$$\begin{aligned} n^{-2} \sum_{i,\, j<k} \mathbb{E}\{V_i V_j U_k^2\} \le K\, \Big\{&[L_1(t_2) - L_1(t_1)]^{1/2}\, [L_2(t_3) - L_2(t_2)] \\ &+ [G(t_2) - G(t_1)]^{\delta/(1+\delta)}\, [L_3(t_3) - L_3(t_2)]\Big\}. \end{aligned}$$

One has a similar bound for the second sum in (3.4). Thus, summarizing, we see that the sums in (3.4) satisfy Chentsov's criterion for tightness. For relevant details see Billingsley (1968), Theorem 15.6. This completes the proof of Lemma 3.1. □

The next lemma covers the case of uniformly bounded $\{Z_{n,i}\}$. In this case we can avoid the moment conditions (A)(b) and (A)(c) and replace the condition (B) by a weaker condition.

LEMMA 3.2. *Suppose the r.v.'s* $\{Z_{n,i}\}$ *are uniformly bounded and* (C) *holds. In addition, suppose there exists a measurable function* φ *from* $\mathbb{R}$ *to* $\mathbb{R}$ *such that the corresponding family of functions* $\{L_y,\ y \in \mathbb{R}\}$ *has Lebesgue densities* $\{l_y, y \in \mathbb{R}\}$ *satisfying*

$$\int \big[\mathbb{E}\, l_{X_0}^{1+\delta}(x - \varphi(X_0))\big]^{1/(1+\delta)}\, dx < \infty, \tag{3.8}$$

for some $\delta > 0$. *Then also the conclusion of Lemma* 3.1 *holds.*

PROOF. Proceed as in the proof of Lemma 3.1 up to (3.6). Now use the boundedness of $\{Z_k\}$ and argue as for (3.7) to conclude that (3.6) is bounded from the above by

$$
\begin{aligned}
&K \int_{t_2}^{t_3} \mathbb{E}\left\{\left(\sum_{i=1}^{k-2} V_i\right)^2 l_{X_{k-2}}(x-\varphi(X_{k-2}))\right\} dx \\
&\quad \le K \int_{t_2}^{t_3} \left(\mathbb{E}\left|\sum_{i=1}^{k-2} V_i\right|^{2((1+\delta)/\delta)}\right)^{\delta/(1+\delta)} \left\{\mathbb{E}\, l_{X_0}^{1+\delta}(x-\varphi(X_0))\right\}^{1/(1+\delta)} dx \\
&\quad \le K\,(k-2)\,[G(t_2)-G(t_1)]^{\delta/(1+\delta)}\,[\Gamma(t_3)-\Gamma(t_2)],
\end{aligned}
$$

with δ as in (3.8). Here

$$
\Gamma(t) = \int_{-\infty}^{t} \left\{\mathbb{E}\, l_{X_0}^{1+\delta}(x-\varphi(X_0))\right\}^{1/(1+\delta)} dx, \qquad -\infty \le t \le \infty.
$$

Note that (3.8) implies that Γ is strictly increasing continuous and bounded on $\mathbb{R}$. Similarly,

$$
\begin{aligned}
\mathbb{E}\left\{V_{k-1}^2 \mathbb{E}(U_k^2|\mathscr{F}_{k-1})\right\} &\le K\,\mathbb{E}\left\{V_{k-1}^2\, I(t_2 < X_{k-1} \le t_3)\right\} \\
&= \mathbb{E}\left\{I(t_1 < X_{k-2} \le t_2)\mathbb{E}\left[Z_{k-1}^2\, I(t_2 < X_{k-1} \le t_3)\middle|\mathscr{F}_{k-2}\right]\right\} \\
&\le K\,\mathbb{E}\left\{I(t_1 < X_{k-2} \le t_2)\left[\int_{t_2}^{t_3} l_{X_{k-2}}(x-\varphi(X_{k-2}))\,dx\right]\right\} \\
&= K\int_{t_2}^{t_3} \mathbb{E}\left\{I(t_1 < X_{k-2} \le t_2)\, l_{X_{k-2}}(x-\varphi(X_{k-2}))\right\} dx \\
&\le K\,[G(t_2)-G(t_1)]^{\delta/(1+\delta)}\,[\Gamma(t_3)-\Gamma(t_2)].
\end{aligned}
$$

Upon combining the above bounds we obtain that (3.5) is bounded from the above by

$$
K(k-1)[G(t_2)-G(t_1)]^{\delta/(1+\delta)}\,[\Gamma(t_3)-\Gamma(t_2)].
$$

Summation from $k=2$ to $k=n$ thus yields

$$
n^{-2} \sum_{i,\, j<k} \mathbb{E}\{V_i V_j U_k^2\} \le K\,[G(t_2)-G(t_1)]^{\delta/(1+\delta)}\,[\Gamma(t_3)-\Gamma(t_2)].
$$

The rest of the details are as in the proof of Lemma 3.1. □

4. Proofs. In this section we present the proofs of various results stated in Sections 1 and 2.

PROOF OF THEOREM 2.1. Part (i) follows from Lemma 3.1 while part (ii) follows from Lemma 3.2 upon choosing $\varphi = m_\psi$, $Z_i = \psi(X_i - m_\psi(X_{i-1}))$, $L_y \equiv F_y$ and $l_y = f_y$ in there. □

Before proceeding further, we state two facts that will be used below repeatedly. Let $\{\xi_i\}$ be r.v.'s with finite first moment such that $\{(\xi_i,\ X_{i-1})\}$ is strictly stationary and ergodic and let ζ_i be stationary square integrable r.v.'s. Then $\max_{1\le i\le n} n^{-1/2}|\zeta_i| = o_{\mathbb{P}}(1)$ and

$$\sup_{y\in\mathbb{R}}\left|n^{-1}\sum_{i=1}^{n}\xi_i\, I(X_{i-1}\le y) - E\xi_1\, I(X_0\le y)\right| \to 0 \quad \text{a.s.} \tag{4.1}$$

The ET (ergodic theorem) implies the pointwise convergence in (4.1). The uniformity is obtained with the aid of the triangle inequality and by decomposing each ξ_i into its negative and positive part and applying a Glivenko–Cantelli type argument to each part.

REMARK 4.1. We are now ready to sketch an argument for the weak convergence of $V_{n,\psi}((\tau^2_{n,\psi})^{-1})$ to B under the hypothesis $m_\psi = m_0$. For the sake of brevity, let $b_n := \tau^2_{n,\psi}(\infty)$, $b := \tau^2_\psi(\infty)$. First, note that

$$\sup_{0\le t\le b_n}|\tau^2_{n,\psi}((\tau^2_{n,\psi})^{-1}(t)) - t| \le \max_{1\le i\le n} n^{-1}\psi^2(X_i - m_0(X_{i-1})) = o_{\mathbb{P}}(1)$$

by (1.2). Next, fix an $\epsilon > 0$ and let $\mathscr{A}_n := [|b_n - b| \le \epsilon]$ and $c_\epsilon := 1/[1-\epsilon/b]$. On $\mathscr{A}_n$,

$$1\Big/\left[1+\frac{\epsilon}{b}\right] \le \frac{b}{b_n} \le 1\Big/\left[1-\frac{\epsilon}{b}\right] = c_\epsilon$$

and

$$\begin{aligned}\sup_{0\le t\le b}|\tau^2_{n,\psi}((\tau^2_{n,\psi})^{-1}(t)) - t| &\le \sup_{0\le t\le b_n}|\tau^2_{n,\psi}((\tau^2_{n,\psi})^{-1}(t)) - t| \\ &\quad + \sup_{b_n< t\le b_n c_\epsilon}|\tau^2_{n,\psi}((\tau^2_{n,\psi})^{-1}(t)) - t|.\end{aligned}$$

The second term is further bounded from the above, on $\mathscr{A}_n$, by $((b+\epsilon)/(b-\epsilon))\epsilon$. However, by the ET, $\mathbb{P}(\mathscr{A}_n) \to 1$. The arbitrariness of ϵ thus readily implies that

$$\sup_{0\le t\le \tau^2_\psi(\infty)}|\tau^2_{n,\psi}((\tau^2_{n,\psi})^{-1}(t)) - t| = o_{\mathbb{P}}(1).$$

We thus obtain, in view of (4.1),

$$\begin{aligned}\sup_{0\le t\le \tau^2_\psi(\infty)}|\tau^2_{\psi}((\tau^2_{n,\psi})^{-1}(t)) - t| &\le \sup_{x\in\mathbb{R}}|\tau^2_\psi(x) - \tau^2_{n,\psi}(x)| \\ &\quad + \sup_{0\le t\le \tau^2_\psi(\infty)}|\tau^2_{n,\psi}((\tau^2_{n,\psi})^{-1}(t)) - t| \\ &= o_{\mathbb{P}}(1).\end{aligned}$$

These observations together with the continuity of the weak limit of $V_{n,\psi}$ imply that

$$\sup_{0\le t\le \tau^2_\psi(\infty)\vee\tau^2_{n,\psi}(\infty)}|V_{n,\psi}((\tau^2_{n,\psi})^{-1}(t)) - V_{n,\psi}((\tau^2_\psi)^{-1}(t))| = o_{\mathbb{P}}(1).$$

Therefore, by Theorem 2.1, $s_{n,\psi}^{-1} V_{n,\psi}((\tau_{n,\psi}^2)^{-1}) \Rightarrow B$ and the limiting distribution of any continuous functional of $s_{n,\psi}^{-1} V_{n,\psi}((\tau_{n,\psi}^2)^{-1})$ can be obtained from the distribution of the corresponding functional of B. In particular, the asymptotic level of the test based on the Cramér–von Mises-type statistic

$$s_{n,\psi}^{-2} \int_0^{s_{n,\psi}^2} V_{n,\psi}^2((\tau_{n,\psi}^2)^{-1}(t))\, dH(t/s_{n,\psi}^2)$$

can be obtained from the distribution of $\int_0^1 B^2\, dH$, where H is a d.f. function on $[0, 1]$.

For our next lemma, put $\varepsilon_{n,i} = X_i - m(X_{i-1}, \theta_n)$ and recall $\varepsilon_i = X_i - m(X_i, \theta_0)$. Let

$$\mathscr{D}_n = n^{-1/2} \sum_{i=1}^{n} \big|\psi(\varepsilon_{n,i}) - \psi(\varepsilon_i) - (\varepsilon_{n,i} - \varepsilon_i)\dot{\psi}(\varepsilon_i)\big| \|r(X_{i-1})\|,$$

where r is a measurable vector-valued function such that

$$\mathbb{E}\|r(X_0)\|^2 < \infty. \tag{4.2}$$

Let $\Delta_n = n^{1/2}(\theta_n - \theta_0)$ and write $m(\cdot)$, $\dot{m}(\cdot)$ for $m(\cdot, \theta_0)$, $\dot{m}(\cdot, \theta_0)$, respectively.

LEMMA 4.1. *Under the assumptions of Theorem* 2.2(a) *and* (4.2), $\mathscr{D}_n = o_{\mathbb{P}}(1)$.

PROOF. Fix an $\alpha > 0$. Then by assumption (A1) and (2.5) there exists a large $k < \infty$ and an integer N such that

$$\mathbb{P}(B_n) > 1 - \alpha \quad \text{for all } n > N,$$

where

$$B_n = \Big\{ \|\Delta_n\| \le k; \max_i \big| m(X_{i-1}, \theta_n) - m(X_{i-1}) - \dot{m}^T(X_{i-1})(\theta_n - \theta_0) \big| \le \alpha/n^{1/2} \Big\}.$$

Now, on B_n, we obtain for $1 \le i \le n$, under H_0,

$$|m(X_{i-1}, \theta_n) - m(X_{i-1})| \le n^{-1/2}[\alpha + k\|\dot{m}(X_{i-1})\|] =: n^{-1/2} h(X_{i-1}).$$

From (2.6) and the stationarity of the process $\{X_i\}$,

$$\max_{1 \le i \le n} n^{-1/2} h(X_{i-1}) = o_{\mathbb{P}}(1).$$

Furthermore, by the absolute continuity of ψ, on B_n,

$$\mathscr{D}_n \le n^{-1/2} \sum_{i=1}^{n} \|r(X_{i-1})\| \int_{-h(X_{i-1})/n^{1/2}}^{h(X_{i-1})/n^{1/2}} |\dot{\psi}(\varepsilon_i - z) - \dot{\psi}(\varepsilon_i)|\, dz.$$

Assuming for the moment that $\dot\psi$ is uniformly continuous, we obtain uniformly in $1 \le i \le n$,

$$\int_{-h(X_{i-1})/n^{1/2}}^{h(X_{i-1})/n^{1/2}} |\dot\psi(\varepsilon_i - z) - \dot\psi(\varepsilon_i)|\, dz = n^{-1/2} h(X_{i-1}) \times o_{\mathbb{P}}(1)$$

so that by (2.6), (4.2) and the ET, the lemma holds.

For an arbitrary $\dot\psi$ satisfying (Ψ_1), we shall use the fact that continuous functions with compact support are dense in $L^1(\mu)$, where μ may be any finite Borel measure [Wheeden and Zygmund (1977), Problem 27, page 192]. In our case $\mu(A) := \mathbb{E}\|r(X_0)\|\ h(X_0) I(\varepsilon_1 \in A)$, for any Borel subset A of $\mathbb{R}$. Again, by (2.6) and (4.2), this is a finite measure. Accordingly, given $\gamma > 0$, there is a continuous function g with compact support such that

$$\text{(4.3)} \qquad \int \|r(X_0)\| h(X_0)\big[|\dot\psi - g|(\varepsilon_1) + |\dot\psi - g|(\varepsilon_1 -) + |\dot\psi - g|(\varepsilon_1 +)\big]\, d\mathbb{P} \le \gamma.$$

We also have

$$\begin{aligned}\mathscr{D}_n \le n^{-1/2} \sum_{i=1}^n \|r(X_{i-1})\| & \\ \times \bigg\{ & \int_{-h(X_{i-1})/n^{1/2}}^{h(X_{i-1})/n^{1/2}} |g(\varepsilon_i - z) - g(\varepsilon_i)|\, dz \\ & + \int_{-h(X_{i-1})/n^{1/2}}^{h(X_{i-1})/n^{1/2}} |\dot\psi(\varepsilon_i - z) - g(\varepsilon_i - z)|\, dz \bigg\} \\ + n^{-1} \sum_{i=1}^n \|r(X_{i-1})\| h(X_{i-1}) |\dot\psi(\varepsilon_i) - g(\varepsilon_i)|. &\end{aligned}$$

The uniform continuity of g implies that the first term is $o_{\mathbb{P}}(1)$ while by the ET and (4.3), the third term is $O_{\mathbb{P}}(\gamma)$. It remains to bound the second term. Its expectation equals $\mathbb{E} f_n(\varepsilon_1, X_0)$, where

$$f_n(\varepsilon_1, X_0) := n^{1/2} \|r(X_0)\| \int_{\varepsilon_1 - h(X_0)/n^{1/2}}^{\varepsilon_1 + h(X_0)/n^{1/2}} |g(z) - \dot\psi(z)|\, dz.$$

Since g and $\dot\psi$ are bounded, $f_n(\varepsilon_1, X_0) \le C\|r(X_0)\| h(X_0)$, which is integrable by (2.6) and (4.2). Moreover, we also have that for each value of (ε_1, X_0),

$$f_n(\varepsilon_1, X_0) \to \|r(X_0)\| h(X_0) \{ |g - \dot\psi|(\varepsilon_1 +) + |g - \dot\psi|(\varepsilon_1 -)\}$$

so that by (4.3) and the dominated convergence theorem $\mathbb{E} f_n(\varepsilon_1, X_0) = O(\gamma)$. This concludes the proof of Lemma 4.1. $\square$

PROOF OF THEOREM 2.2. Put

$$R_n(x) := V_n^1(x) - V_n(x) = n^{-1/2} \sum_{i=1}^n [\psi(\varepsilon_{n,i}) - \psi(\varepsilon_i)] I(X_{i-1} \le x), \qquad x \in \mathbb{R}.$$

Decompose R_n as

$$
\begin{aligned}
R_n(x) &= n^{-1/2}\sum_{i=1}^{n}[\psi(\varepsilon_{n,i})-\psi(\varepsilon_i)-(\varepsilon_{n,i}-\varepsilon_i)\dot\psi(\varepsilon_i)]I(X_{i-1}\le x)\\
&\qquad - n^{-1/2}\sum_{i=1}^{n}[m(X_{i-1},\theta_n)-m(X_{i-1})-\dot m^T(X_{i-1})(\theta_n-\theta_0)]\\
&\qquad\qquad \times \dot\psi(\varepsilon_i)\, I(X_{i-1}\le x)\\
&\qquad - n^{-1/2}\sum_{i=1}^{n}\dot m^T(X_{i-1})\,\dot\psi(\varepsilon_i) I(X_{i-1}\le x)(\theta_n-\theta_0)\\
&= R_{n1}(x)+R_{n2}(x)+R_{n3}(x)\, n^{1/2}(\theta_n-\theta_0)\quad \text{say.}
\end{aligned}
$$

The term $R_{n3}(x)$ is equal to

$$
n^{-1}\sum_{i=1}^{n}\dot m^T(X_{i-1})\,\dot\psi(\varepsilon_i) I(X_{i-1}\le x).
$$

By an application of (4.1) we readily obtain that

$$
\sup_{x\in\mathbb{R}}\|R_{n3}(x)-\dot M^T(x,\theta_0)\| = o_{\mathbb{P}}(1).
$$

Due to (A1), it thus remains to show that R_{n1} and R_{n2} tend to zero in probability uniformly in x. The assertion for R_{n1} follows immediately from Lemma 4.1, because it is uniformly bounded by the $\mathscr{D}_n$ with $r\equiv 1$. As to R_{n2}, recall the event B_n from the proof of Lemma 4.1 and note that on B_n,

$$
\sup_x |R_{n2}(x)| \le \alpha n^{-1}\sum_{i=1}^{n}|\dot\psi(\varepsilon_i)| = O(\alpha)\quad \text{a.s.,}
$$

by the ET. Since $\alpha>0$ is arbitrarily chosen, this completes the proof of part (a).

As to the proof of part (b), put

$$
\begin{aligned}
d_{n,i}(t) &:= m(X_{i-1},\theta_0+n^{-1/2}t)-m(X_{i-1},\theta_0);\\
\gamma_{n,i} &:= n^{-1/2}(2\alpha+\delta\|\dot m(X_{i-1},\theta_0)\|),\qquad \alpha>0,\ \delta>0;\\
\mu_n(X_{i-1},t,a) &:= \mathbb{E}[\psi(\varepsilon_i-d_{n,i}(t)+a\gamma_{n,i})\,|\,X_{i-1}],\qquad 1\le i\le n,\ t\in\mathbb{R}^q.
\end{aligned}
$$

Define, for $a,\ x\in\mathbb{R}$ and $t\in\mathbb{R}^q$,

$$
\begin{aligned}
D_n(x,t,a) := n^{-1/2}\sum_{i=1}^{n}\big[&\psi(\varepsilon_i-d_{n,i}(t)+a\gamma_{n,i})\\
&-\mu_n(X_{i-1},t,a)-\psi(\varepsilon_i)\big]\, I(X_{i-1}\le x).
\end{aligned}
$$

Write $D_n(x,t)$ and $\mu_n(X_{i-1},t)$ for $D_n(x,t,0)$ and $\mu_n(X_{i-1},t,0)$, respectively.

Note that the summands in $D_n(x, t, a)$ form mean zero bounded martingale differences, for each x, t and a. Thus

$$\begin{aligned}\mathrm{Var}(D_n(x, t, a)) &\le \mathbb{E}[\psi(\varepsilon_1 - d_{n,1}(t) + a\gamma_{n,1}) - \mu_n(X_0, t, a) - \psi(\varepsilon_1)]^2\\ &\le \mathbb{E}[\psi(\varepsilon_1 - d_{n,1}(t) + a\gamma_{n,1}) - \psi(\varepsilon_1)]^2 \to 0,\end{aligned}$$

by assumption (A2) and (Ψ_2). Upon an application of Lemma 3.2 with $Z_{n,i} = \psi(\varepsilon_i - d_{n,i}(t) + a\gamma_{n,i}) - \mu_n(X_{i-1}, t, a) - \psi(\varepsilon_i)$ we readily obtain that

$$\sup_{x\in\mathbb{R}} |D_n(x, t, a)| = o_{\mathbb{P}}(1) \qquad \forall\ a \in \mathbb{R},\ t \in \mathbb{R}^q. \tag{4.4}$$

The assumption (C) of Lemma 3.2 with these $\{Z_{n,i}\}$ and $\tau^2 \equiv 0$ is implied by (Ψ_2) while (2.4) implies (3.8) here.

Let $\mathscr{N}_b := \{t \in \mathbb{R}^q;\ \|t\| \le b\},\ 0 < b < \infty$. We need to prove that for every $b < \infty$,

$$\sup_{x\in\mathbb{R},\, t\in\mathscr{N}_b} |D_n(x, t)| = o_{\mathbb{P}}(1). \tag{4.5}$$

To that effect, let $C_n := \{\sup_{t\in N_b} |d_{n,i}(t)| \le n^{-1/2}(\alpha + b\|\dot m(X_{i-1})\|),\ 1 \le i \le n\}$, and for an $s \in \mathscr{N}_b$, let

$$A_n := \Big\{ \sup_{t\in\mathscr{N}_b,\, \|t-s\|\le\delta} |d_{n,i}(t) - d_{n,i}(s)| \le \gamma_{n,i},\ 1 \le i \le n \Big\} \cap C_n.$$

By assumption (A2), there is an $N < \infty$, depending only on α, such that $\forall\, b < \infty$ and $\forall\ s \in \mathscr{N}_b$,

$$\mathbb{P}(A_n) > 1 - \alpha \qquad \forall\ n > N. \tag{4.6}$$

Now, by the monotonicity of ψ one obtains that on A_n, for each fixed $s \in \mathscr{N}_b$ and $\forall\ t \in \mathscr{N}_b$ with $\|t - s\| \le \delta$,

$$\begin{aligned}|D_n(x, t)| &\le |D_n(x, s, 1)| + |D_n(x, s, -1)|\\ &\quad + \Big| n^{-1/2} \sum_{i=1}^{n} [\mu_n(X_{i-1}, s, 1) - \mu_n(X_{i-1}, s, -1)]\, I(X_{i-1} \le x) \Big|.\end{aligned}$$

By (4.4), the first two terms converge to zero uniformly in x, in probability, while the last term is bounded above by

$$n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} |F_{X_{i-1}}(y + d_{n,i}(s) + \gamma_{n,i}) - F_{X_{i-1}}(y + d_{n,i}(s) - \gamma_{n,i})|\, \psi(dy).$$

Observe that for every $s \in \mathscr{N}_b$, on A_n, $|d_{n,i}(s)| + \gamma_{n,i} \le a_n$, for all $1 \le i \le n$, where $a_n := \max_{1\le i\le n} n^{-1/2}[3\alpha + (b + \delta)\|\dot m(X_{i-1})\|]$. By (Ψ_2), the above bound in turn is bounded from above by

$$n^{-1/2} \sum_{i=1}^{n} \gamma_{n,i} \Big[\sup_{y\in\mathbb{R},\, |x-z|\le a_n} |f_y(x) - f_y(z)| + 2 \int_{-\infty}^{\infty} f_{X_{i-1}}(y)\, \psi(dy) \Big]. \tag{4.7}$$

Now, by the ET, (2.6) and (F),

$$n^{-1/2}\sum_{i=1}^{n}\gamma_{n,i} = n^{-1}\sum_{i=1}^{n}(2\alpha+\delta\|\dot m(X_{i-1},\theta_0)\|) = O_{\mathbb{P}}(1)$$

and

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n}\gamma_{n,i}&\int_{-\infty}^{\infty} f_{X_{i-1}}(y)\,\psi(dy)\\ &\to 2\alpha\int_{-\infty}^{\infty}\mathbb{E}f_{X_0}(y)\,\psi(dy)+\delta\int_{-\infty}^{\infty}\mathbb{E}\|\dot m(X_0)\|f_{X_0}(y)\,\psi(dy). \end{aligned}$$

Observe that, by (2.6), the functions $q_j(y) := \mathbb{E}\|\dot m(X_0)\|^j f_{X_0}(y)$, $j=0,1$, $y\in\mathbb{R}$, are Lebesgue integrable on $\mathbb{R}$. By (F), they are also uniformly continuous and hence bounded on $\mathbb{R}$ so that $\int_{-\infty}^{\infty} q_j(y)\,\psi(dy)<\infty$ for $j=0,1$. We thus obtain that the bound in (4.7) converges in probability to

$$4\,\alpha\int_{-\infty}^{\infty} q_0(y)\,\psi(dy)+2\,\delta\int_{-\infty}^{\infty} q_1(y)\,\psi(dy),$$

which can be made less than δ by the choice of α. This together with (4.4) applied with $a=0$ and the compactness of $\mathcal{N}_b$ proves (4.5).

Next, by (1.1) and Fubini's theorem, we have

$$\begin{aligned} n^{-1/2}&\sum_{i=1}^{n}\mu_n(X_{i-1},t)\,I(X_{i-1}\le x)\\ &= n^{-1/2}\sum_{i=1}^{n}\big[\mu_n(X_{i-1},t)-\mu_n(X_{i-1},0)\big]\,I(X_{i-1}\le x)\\ &= -n^{-1/2}\sum_{i=1}^{n} I(X_{i-1}\le x)\int_{-\infty}^{\infty}\big[F_{X_{i-1}}(y+d_{n,i}(t))-F_{X_{i-1}}(y)\big]\,\psi(dy)\\ &= -n^{-1}\sum_{i=1}^{n}\dot m^T(X_{i-1},\theta_0)\,t\,I(X_{i-1}\le x)\int_{-\infty}^{\infty} f_{X_{i-1}}(y)\,\psi(dy)+o_{\mathbb{P}}(1)\\ &= -\mathbb{E}\dot m^T(X_0,\theta_0)\,I(X_0\le x)\int_{-\infty}^{\infty} f_{X_0}(y)\,\psi(dy)\,t+o_{\mathbb{P}}(1), \end{aligned}$$

uniformly in $x\in\mathbb{R}$ and $t\in\mathcal{N}_b$. In the above, the last equality follows from (4.1) while the one before that follows from the assumptions (A2), (Ψ_2) and (F). This together with (4.5), (4.6) and the assumption (A1) proves (2.7) and hence the part (b) of Theorem 2.2. □

PROOF OF THEOREM 2.3. Details will be given for part (a) only, being similar to part (b). We shall first prove (2.12). From the definitions of T, we obtain that

$$TV^1_{n,\psi}(x) = V^1_{n,\psi}(x)-\int_{-\infty}^{x}\dot m^T(y)\,A^{-1}(y)\bigg[\int_{y}^{\infty}\dot m(t)\,V^1_{n,\psi}(dt)\bigg]G(dy), \tag{4.8}$$

$$TV_{n,\psi}(x) = V_{n,\psi}(x) - \int_{-\infty}^{x} \dot{m}^T(y)\, A^{-1}(y)\left[\int_y^{\infty} \dot{m}(t)\, V_n(dt)\right] G(dy). \tag{4.9}$$

As before, set $\Delta_n := n^{1/2}(\theta_n - \theta_0)$. From (2.7) we obtain, uniformly in $x \in \mathbb{R}$,

$$V^1_{n,\psi}(x) = V_{n,\psi}(x) - \gamma\, \nu^T(x)\, \Delta_n + o_{\mathbb{P}}(1). \tag{4.10}$$

Recall $\varepsilon_{n,i} := X_i - m(X_{i-1}, \theta_n)$, $1 \le i \le n$. Then the two integrals in (4.8) and (4.9) differ by

$$\int_{\infty}^{x} \dot{m}^T(y)\, A^{-1}(y)\, D_n(y) G(dy),$$

where

$$D_n(y) := n^{-1/2} \sum_{i=1}^{n} \dot{m}(X_{i-1})\, [\psi(\varepsilon_i) - \psi(\varepsilon_{n,i})] I(X_{i-1} \ge y).$$

This process is similar to the process R_n as studied in the proof of Theorem 2.2(a). Decompose D_n as

$$\begin{aligned} D_n(y) &= n^{-1/2} \sum_{i=1}^{n} \dot{m}(X_{i-1})\, [\psi(\varepsilon_i) - \psi(\varepsilon_{n,i}) - (\varepsilon_i - \varepsilon_{n,i})\dot{\psi}(\varepsilon_i)] I(X_{i-1} \ge y) \\ &\quad + n^{-1/2} \sum_{i=1}^{n} \dot{m}(X_{i-1}) \big[m(X_{i-1}, \theta_n) - m(X_{i-1}) \\ &\qquad\qquad - \dot{m}^T(X_{i-1})\, (\theta_n - \theta_0)\big]\, \dot{\psi}(\varepsilon_i)\, I(X_{i-1} \ge y) \\ &\quad + n^{-1/2} \sum_{i=1}^{n} \dot{m}(X_{i-1})\, \dot{m}^T(X_{i-1})\, \dot{\psi}(\varepsilon_i) I(X_{i-1} \ge y)\, (\theta_n - \theta_0) \\ &= D_{n1}(y) + D_{n2}(y) + D_{n3}(y)\, \Delta_n \quad \text{say.} \end{aligned} \tag{4.11}$$

Apply Lemma 4.1, with $r = \dot{m}$ and the triangle inequality to readily obtain

$$\sup_{y\in\mathbb{R}} \|D_{n1}(y)\| = o_{\mathbb{P}}(1).$$

This fact together with (2.17) yields

$$\sup_{x \le x_0} \left| \int_{-\infty}^{x} \dot{m}^T(y)\, A^{-1}(y)\, D_{n1}(y)\, G(dy) \right| = o_{\mathbb{P}}(1). \tag{4.12}$$

Recall B_n from the proof of Lemma 4.1. Then, on B_n,

$$\sup_{y\in\mathbb{R}} \|D_{n2}(y)\| \le \alpha\, n^{-1} \sum_{i=1}^{n} \|\dot{m}(X_{i-1})\|\, |\dot{\psi}(\varepsilon_i)| = O(\alpha) \quad \text{a.s.,} \tag{4.13}$$

by the ET. Arbitrariness of α and (2.17) yield

$$\sup_{x \le x_0} \left| \int_{-\infty}^{x} \dot{m}^T(y)\, A^{-1}(y)\, D_{n2}(y)\, G(dy) \right| = o_{\mathbb{P}}(1). \tag{4.14}$$

Now consider the third term. We have

$$D_{n3}(y) = n^{-1} \sum_{i=1}^{n} \dot{m}(X_{i-1})\, \dot{m}^T(X_{i-1})\, \dot{\psi}(\varepsilon_i)\, I(X_{i-1} \ge y).$$

An application of (4.1) together with (2.8) yields that $\sup_{y\in\mathbb{R}} \|D_{n3}(y) - \gamma A(y)\| \to 0$, a.s. This together with the fact $\|\Delta_n\| = O_{\mathbb{P}}(1)$ entails that

$$\sup_{x \le x_0} \left| \left[\int_{-\infty}^{x} \dot{m}^T(y)\, A^{-1}(y)\, D_{n3}(y)\, G(dy) - \gamma\, \nu^T(x) \right] \Delta_n \right| = o_{\mathbb{P}}(1). \tag{4.15}$$

The proof of the claim (2.12) is complete upon combining (4.12)–(4.15) with (4.8)–(4.10).

Next, we turn to the proof of (2.13). In view of (2.12), it suffices to prove $TV_{n,\psi} \Rightarrow TV_{\psi}$. To this effect, note that for each real x, $TV_{n,\psi}(x)$ is a mean zero square integrable martingale. The convergence of the finite-dimensional distributions thus follows from the martingale CLT.

To verify the tightness, because $V_{n,\psi}$ is tight and has a continuous limit by Theorem 2.1, it suffices to prove the same for the second term in (4.9). To that effect, let

$$L(x) := \int_{-\infty}^{x} \|\dot{m}^T A^{-1}\|\, dG, \qquad x \le x_0.$$

Note that L is nondecreasing, continuous and $L(x_0) < \infty$ [see (2.17)]. Now, rewrite the relevant term as

$$K_n(x) := n^{-1/2} \sum_{i=1}^{n} \psi(\varepsilon_i) \int_{-\infty}^{x} \dot{m}^T(y)\, A^{-1}(y) \dot{m}(X_{i-1}) I(X_{i-1} \ge y)\, G(dy).$$

Because the summands are martingale differences and because of (2.8) with $\sigma_\psi^2 = 1$, we obtain, with the help of Fubini's theorem, that for $x < y$,

$$\mathbb{E}[K_n(y) - K_n(x)]^2 = \int_x^y \int_x^y \dot{m}^T(s)\, A^{-1}(s)\, A(s \vee t)\, A^{-1}(t)\, \dot{m}(t)\, G(dt)\, G(ds).$$

By (2.6), $\|A\|_\infty := \sup_{x\in\mathbb{R}} \|A(x)\| \le \int_{-\infty}^{\infty} \|\dot{m}\|^2\, dG < \infty$. We thus obtain that

$$\mathbb{E}[K_n(y) - K_n(x)]^2 \le \|A\|_\infty \left[\int_x^y \|\dot{m}^T A^{-1}\|\, dG \right]^2 = \|A\|_\infty [L(y) - L(x)]^2.$$

This then yields the tightness of the second term in (4.9) in a standard fashion and also completes the proof of Theorem 2.3(a). □

For the proof of Theorem 2.4 the following lemma will be crucial.

LEMMA 4.2. *Let $\mathscr{U}$ be a relatively compact subset of $D[-\infty, x_0]$. Let L, L_n be a sequence of random distribution functions on $\mathbb{R}$ such that*

$$\sup_{t \le x_0} |L_n(t) - L(t)| \to 0 \quad a.s.$$

Then

$$\sup_{t \le x_0,\, \alpha \in \mathscr{U}} \left| \int_{-\infty}^{t} \alpha(x)[L_n(dx) - L(dx)] \right| = o_{\mathbb{P}}(1).$$

This lemma is an extension of Lemma 3.1 of Chang (1990) to the dependent setup, where the strong law of large numbers is replaced by the ET. Its proof is similar and uses the fact that the uniform convergence over compact families of functions follows from the uniform convergence over intervals.

In the following proofs, the above lemma is used with $L_n \equiv G_n$ and $L \equiv G$ and more generally, with L_n and L given by the relations $dL_n \equiv h dG_n$, $dL \equiv h dG$, where h is an G-integrable function. As to the choice of $\mathscr{U}$, let $\{\alpha_n\}$ be a sequence of stochastic processes which are uniformly tight; that is, for a given $\delta > 0$ there exists a compact set $\mathscr{U}$ such that $\alpha_n \in \mathscr{U}$ with probability at least $1 - \delta$. Apply Lemma 4.2 with this $\mathscr{U}$ and observe that $\alpha_n \notin \mathscr{U}$, with small probability to finally get uniformly in t,

$$\left| \int_{-\infty}^{t} \alpha_n(x)[L_n(dx) - L(dx)] \right| = o_{\mathbb{P}}(1).$$

As will be seen below, these types of integrals appear in the expansion of $T_n V^1_{n,\psi}$.

PROOF OF THEOREM 2.4. Again, the details below are given for part (a) only, and for convenience we do not exhibit θ_0 in m, $\dot{m}$, K_1 and $\ddot{m}$. First, note that the stationarity of the process and (2.6) imply that $n^{-1/2} \max_i \|\dot{m}(X_{i-1})\| = o_{\mathbb{P}}(1)$. Recall $\Delta_n = n^{1/2}(\theta_n - \theta_0)$. Then by (2.14) and the ET, on an event with probability tending to one and for a given $\varepsilon > 0$,

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} &\|\dot{m}(X_{i-1})\| \, \|\dot{m}(X_{i-1}, \theta_n) - \dot{m}(X_{i-1})\| \\ &\le n^{-1/2} \max_i \|\dot{m}(x_{i-1})\| \, \|\Delta_n\| \\ &\quad \times \left\{ \epsilon\, n^{-1} \sum_{i=1}^{n} K_1(X_{i-1}) + n^{-1} \sum_{i=1}^{n} \|\ddot{m}(X_{i-1})\| \right\} \\ &= o_{\mathbb{P}}(1). \end{aligned}$$

Similarly, one obtains

$$n^{-1} \sum_{i=1}^{n} \|\dot{m}(X_{i-1}, \theta_n) - \dot{m}(X_{i-1})\|^2 = o_{\mathbb{P}}(1).$$

These bounds in turn, together with (4.1), imply that

$$\begin{aligned}\sup_{y\in\mathbb{R}}\|A_n(y)-A(y)\| &\le 2n^{-1}\sum_{i=1}^{n}\|\dot m(X_{i-1})\|\,\|\dot m(X_{i-1},\theta_n)-\dot m(X_{i-1})\|\\ &\quad + n^{-1}\sum_{i=1}^{n}\|\dot m(X_{i-1},\theta_n)-\dot m(X_{i-1})\|^2\\ &\quad + \sup_{y\in\mathbb{R}}\left\|n^{-1}\sum_{i=1}^{n}\dot m(X_{i-1})\dot m^T(X_{i-1})I(X_{i-1}\ge y)-A(y)\right\|\\ &= o_{\mathbb{P}}(1).\end{aligned}$$

Consequently, we have

$$\sup_{y\le x_0}\|A_n^{-1}(y)-A^{-1}(y)\| = o_{\mathbb{P}}(1). \tag{4.16}$$

Next, we shall prove (2.15). For the sake of brevity, write V_n^1, V_n for $V_{n,\psi}^1$, $V_{n,\psi}$, respectively. Let

$$U_n^1(y) := \int_y^\infty \dot m(t,\theta_n)\,V_n^1(dt), \qquad U_n(y) := \int_y^\infty \dot m(t)\,V_n(dt).$$

Then we have

$$T_nV_n^1(x) = V_n^1(x) - \int_{-\infty}^{x}\dot m^T(y,\theta_n)\,A_n^{-1}(y)\,U_n^1(y)\,G_n(dy),$$

so that from (4.9) and (4.10) we obtain, uniformly in $x\in\mathbb{R}$,

$$\begin{aligned}T_nV_n^1(x)-TV_n(x) &:= -\gamma\,\nu^T(x)\,\Delta_n + o_{\mathbb{P}}(1)\\ &\qquad + \int_{-\infty}^{x}\dot m^T(y)\,A^{-1}(y)\,U_n(y)\,G(dy)\\ &\qquad - \int_{-\infty}^{x}\dot m^T(y,\theta_n)\,A_n^{-1}(y)\,U_n^1(y)\,G_n(dy)\\ &= -\gamma\,\nu^T(x)\,\Delta_n + o_{\mathbb{P}}(1) + B_{n1}(x) - B_{n2}(x) \quad \text{say.}\end{aligned} \tag{4.17}$$

We shall shortly show that

$$\sup_{x\le x_0}\|U_n^1(x)-U_n(x)+\gamma\,A(x)\,\Delta_n\| = o_{\mathbb{P}}(1). \tag{4.18}$$

Apply Lemma 3.1 k times, jth time with $Z_{n,i}\equiv \dot m_j(X_{i-1})\psi(\varepsilon_i)$, where $\dot m_j$ is the jth component of $\dot m$, $1\le j\le k$. Then under the assumed conditions it follows that U_n is tight. Using (4.16), (4.18), Lemma 4.2 and the assumptions (2.14), we obtain

$$\begin{aligned}B_{n2}(x) &= \int_{-\infty}^{x}\dot m^T A^{-1} U_n\,dG_n - \gamma\int_{-\infty}^{x}\dot m^T\,dG\,\Delta_n + o_{\mathbb{P}}(1),\\ &= \int_{-\infty}^{x}\dot m^T A^{-1} U_n\,dG - \gamma\,\nu^T(x)\,\Delta_n + o_{\mathbb{P}}(1),\end{aligned}$$

uniformly in $x\le x_0$, which in turn together with (4.17) implies (2.15).

We shall now prove (4.18). Some of the arguments are similar to the proof of Lemma 2.1. Again for the sake of brevity, write $m_i(t)$, $\dot m_i(t)$ for $m(X_{i-1}, t)$, $\dot m(X_{i-1}, t)$, respectively, with the convention that the dependence on the true θ_0 will not be exhibited. Also recall the definition of $\varepsilon_{n,i}$ from the proof of Lemma 2.1. Now, decompose U_n^1 as follows:

$$\begin{aligned}
U_n^1(y) &= n^{-1/2}\sum_{i=1}^n \dot m_i(\theta_n)\,\psi(\varepsilon_{n,i})\,I(X_{i-1}\ge y)\\
&= n^{-1/2}\sum_{i=1}^n \dot m_i(\theta_n)\,[\psi(\varepsilon_{n,i})-\psi(\varepsilon_i)-(\varepsilon_{n,i}-\varepsilon_i)\dot\psi(\varepsilon_i)]I(X_{i-1}\ge y)\\
&\quad + n^{-1/2}\sum_{i=1}^n \dot m_i(\theta_n)\,[m_i-m_i(\theta_n)-\dot m_i^T\,(\theta_0-\theta_n)]\,\dot\psi(\varepsilon_i)\,I(X_{i-1}\ge y)\\
&\quad - n^{-1}\sum_{i=1}^n \dot m_i(\theta_n)\,\dot m_i^T\,\dot\psi(\varepsilon_i)I(X_{i-1}\ge y)\,\Delta_n\\
&\quad + n^{-1/2}\sum_{i=1}^n \dot m_i(\theta_n)\,\psi(\varepsilon_i)\,I(X_{i-1}\ge y)\\
&= -T_{n1}(y)-T_{n2}(y)-T_{n3}(y)\,\Delta_n+T_{n4}(y)\quad \text{say.}
\end{aligned}$$

Observe that T_{n1}, T_{n2} are, respectively, similar to D_{n1}, D_{n2} in the proof of Lemma 2.1 except the weights $\dot m_i$ are now replaced by $\dot m_i(\theta_n)$. We shall first approximate T_{n1} by D_{n1}. We obtain, for a given $\epsilon>0$,

$$\begin{aligned}
\sup_{y\in\mathbb{R}}\|T_{n1}(y)-D_{n1}(y)\| &\le n^{-1}\sum_{i=1}^n\big[\|\ddot m(X_{i-1})\|+\epsilon\,K_1(X_{i-1})\big]\,\|\Delta_n\|\\
&\qquad\times\int_{-h(X_{i-1})/n^{1/2}}^{h(X_{i-1})/n^{1/2}}|\dot\psi(\varepsilon_i-s)-\dot\psi(\varepsilon_i)|\,ds\\
&= o_{\mathbb{P}}(1).
\end{aligned}$$

A similar, but simpler, argument using the assumption (A2) shows that $\sup_{y\in\mathbb{R}}\|T_{n2}(y)-D_{n2}(y)\| = o_{\mathbb{P}}(1)$. Since D_{n1} and D_{n2} tend to zero uniformly in y, we conclude that

$$\sup_{y\in\mathbb{R}}\{\|T_{n1}(y)\|+\|T_{n2}(y)\|\} = o_{\mathbb{P}}(1).$$

Again, using (2.14) and (4.1) we obtain

$$T_{n3}(y) = n^{-1}\sum_{i=1}^n \dot m_i\,\dot m_i^T\,\dot\psi(\varepsilon_i)I(X_{i-1}\ge y)+o_{\mathbb{P}}(1) = \gamma A(y)+o_{\mathbb{P}}(1),$$

uniformly in $y\in\mathbb{R}$. We now turn to T_{n4}. We shall prove

$$\sup_{y\in\mathbb{R}}\left\|T_{n4}(y)-n^{-1/2}\sum_{i=1}^n \dot m_i\,\psi(\varepsilon_i)\,I(X_{i-1}\ge y)\right\| = o_{\mathbb{P}}(1). \tag{4.19}$$

To that effect let $g_{ni} := \dot{m}_i(\theta_n) - \dot{m}_i(\theta_0) - \ddot{m}_i(\theta_0)(\theta_n - \theta_0)$ and

$$\Gamma_n(y) := n^{-1/2} \sum_{i=1}^{n} g_{ni}\, \psi(\varepsilon_i)\, I(X_{i-1} \geq y).$$

Clearly, by (2.14), on a large set,

$$\sup_{y \in \mathbb{R}} \|\Gamma_n(y)\| \leq \epsilon\, kn^{-1} \sum_{i=1}^{n} K_1(X_{i-1})\, |\psi(\varepsilon_i)| = O_{\mathbb{P}}(\epsilon).$$

But, because of (1.1) and (4.1), $\sup_{y \in \mathbb{R}} \|n^{-1} \sum_{i=1}^{n} \ddot{m}_i(\theta_0)\, \psi(\varepsilon_i)\, I(X_{i-1} \geq y)\| = o_{\mathbb{P}}(1)$.

The claim (4.19) thus follows from these facts and the assumption that $\Delta_n = O_{\mathbb{P}}(1)$, in a routine fashion. This also completes the proof of (4.18) and hence that of the theorem. □

Acknowledgments. The authors thank the three referees for their constructive comments and careful reading of the manuscript.

REFERENCES

AN, H.-S. and BING, C. (1991). A Kolmogorov–Smirnov type statistic with application to test for nonlinearity in time series. *Internat. Statist. Rev.* **59** 287–307.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

CHANG, N. M. (1990). Weak convergence of a self-consistent estimator of a survival function with doubly censored data. *Ann. Statist.* **18** 391–404.

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Springer, New York.

DELGADO, M. A. (1993). Testing the equality of nonparametric curves. *Statist. Probab. Lett.* **17** 199–204.

DIEBOLT, J. (1995). A nonparametric test for the regression function: asymptotic theory. *J. Statist. Plann. Inference* **44** 1–17.

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

DURBIN, J., KNOTT, M. and TAYLOR, C. C. (1975). Components of Cramér–von Mises statistics. II. *J. Roy. Statist. Soc. Ser. B* **37** 216–237.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

HJELLVIK, V. and TJØSTHEIM, D. (1995). Nonparametric test of linearity for time series. *Biometrika* **82** 351–368.

HJELLVIK, V. and TJØSTHEIM, D. (1996). Nonparametric statistics for testing linearity and serial independence. *Nonparametric Statist.* **6** 223–251.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

KHMALADZE, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240–257.

KHMALADZE, E. V. (1988). An innovation approach to goodness-of-fit tests in $\mathbb{R}^m$. *Ann. Statist.* **16** 1503–1516.

KHMALADZE, E. V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21** 798–829.

KOUL, H. L. (1996). Asymptotics of some estimators and sequential residual empiricals in nonlinear time series. *Ann. Statist.* **24** 380–404.

MACKINNON, J. G. (1992). Model specification tests and artificial regression. *J. Econom. Literature* **XXX** (March) 102–146.

McKEAGUE, I. W. and ZHANG, M. J. (1994). Identification of nonlinear time series from first order cumulative characteristics. *Ann. Statist.* **22** 495–514.
NIKABADZE, A. and STUTE, W. (1996). Model checks under random censorship. *Statist. Probab. Lett.* **32** 249–259.
OZAKI, T. and ODA, H. (1978). Non-linear time series model identification by Akaike's information criterion. In *Proceedings of the IFAC Workshop on Information and Systems*. Pergamon Press, New York.
RAO, K. C. (1972). The Kolmogorof, Cramér–von Mises chi-squares statistics for goodness-of-fit tests in the parametric case (Abstract). *Bull. Inst. Math. Statist.* **1** 87.
ROBINSON, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207.
ROBINSON, P. M. (1984). Robust nonparametric autoregression. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 247–255. Springer, Berlin.
ROUSSAS, G. G. and TRAN, L. T. (1992). Asymptotic normality of the recursive estimate under dependence condition. *Ann. Statist.* **20** 98–120.
STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641.
STUTE, W., THIES, S. and ZHU, L. X. (1998). Model checks for regression: an innovation process approach. Unpublished manuscript.
SU, J. Q. and WEI, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *J. Amer. Statist. Assoc.* **86** 420–426.
TJØSTHEIM, D. (1986). Estimation in nonlinear time series models I: stationary series. *Stochastic Process. Appl.* **21** 251–273.
TONG, H. (1990). *Non-linear Time Series Analysis: A Dynamical Approach.* Oxford Univ. Press.
TRUONG, Y. K. and STONE, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20** 77–97.
WHEEDEN, R. L. and ZYGMUND, A. (1977). *Measure and Integral: An Introduction to Real Analysis.* Dekker, New York.

DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824-1027
E-MAIL: koul@stt.msu.edu

MATHEMATICAL INSTITUTE
UNIVERSITY OF GIESSEN
ARNDTSTRASSE 2
D-35392 GIESSEN
GERMANY