A Specification Test based on Convolution-type Distribution Function Estimates for Non-linear Autoregressive Processes

Kun Ho Kim[1] Hira L. Koul Jiwoong Kim
Concordia University, Canada Michigan State University, USA
University of South Florida, USA

Abstract

This paper proposes a test for a parametric specification of the autoregressive function of a given stationary autoregressive time series. This test is based on the integrated square difference between the empirical distribution function estimate and a convolution-type distribution function estimate of the stationary distribution function obtained from the autoregressive residuals. Some asymptotic properties of the proposed convolution-type distribution function estimate are studied when the model's innovation density is unknown. These properties are in turn used to derive the asymptotic null distribution of the proposed test statistic. We also discuss some finite-sample properties of the test statistic based on the block bootstrap methodology. A simulation study shows that the proposed test competes favorably with some existing tests in terms of the empirical level and power.

Key words: Integrated squared difference of the two d.f.'s. Empirical and convolution d.f. estimators. Block bootstrap. Asymptotic power. Empirical level and power.

JEL Classifications: C12, C13, C14

[1]Corresponding author: kunhokim8@gmail.com

1 Introduction

Consider the autoregressive model framework in time series:

(1.1) $$X_i = m(X_{i-1}) + \varepsilon_i, \quad i \in \mathbb{Z} := \{0, \pm 1, \cdots\},$$

where X_i is a real valued stationary and ergodic process, m is a measurable function defined on the real line $\mathbb{R}$ to $\mathbb{R}$ and $\varepsilon, \varepsilon_i$ are independent and identically distributed (i.i.d.) random variables (r.v.'s) having $E\varepsilon \equiv 0, 0 < E\varepsilon^2 < \infty$, and X_{i-1} is independent of ε_i, for each $i \in \mathbb{Z}$. Hence, $m(x) = E(X_1|X_0 = x)$ is the conditional mean function. The framework (1.1) has been a popular benchmark in econometrics due to its simple but effective representation of time series in many applications.

Let q be a known positive integer, $\Theta \subset \mathbb{R}^q$, and $\{m_\vartheta(x), \vartheta \in \Theta, x \in \mathbb{R}\}$ a family of parametric functions. The problem of interest here is to test the hypothesis

(1.2) $$H_0 : m(x) = m_\theta(x), \text{ for some } \theta \in \Theta \text{ and for all } x \in \mathbb{R}$$

based on the data $X_i, 0 \le i \le n$.

Various tests for testing H_0 have been proposed in the time series literature. Aït-Sahalia (1996) proposed a parametric specification test by comparing the nonparametric kernel density estimate of the marginal density of X_0 with its closed-form density estimate under the parametric form. Unfortunately, the test proposed in Aït-Sahalia (1996) is hampered by the slow consistency rate of the kernel density estimates, which is closely examined by Pritsker (1998). In an attempt to overcome this shortcoming, Corradi and Swanson (2005) propose a test that utilizes the empirical d.f. of X_i's and the closed-form d.f. estimate under the parametric forms of the mean function because, unlike the density estimate, the d.f. estimates achieve the root-n-consistency rate. However, the critical values of their test cannot be tabulated even for the large samples because its the limiting distribution involves a functional of a Gaussian process with unknown covariance function. They employ the block bootstrap procedures (Künsch, 1989; Lahiri, 1992, 2003) that can capture the dependence structure of the process to perform their proposed test. Aït-Sahalia and Park (2012) analyze the asymptotic behavior of the specification test of Aït-Sahalia (1996) for the stationary density of a diffusion process, but when the diffusion is not stationary. They consider integrated and explosive processes, as well as nearly integrated ones in the spirit of the local to unity analysis in classical unit root theory. The paper finds that the behavior of the test predicted by the asymptotic distribution under an integrated process provides a better approximation to the finite sample distribution of the test than that predicted by the asymptotic distribution under strict stationarity.

Despite the innovative nature of their ideas, the tests in Aït-Sahalia (1996) and Corradi and Swanson (2005) are *not applicable* if the closed-form density and closed-form d.f. of the process are not available. This significantly reduces the applicability of these tests because the closed-form density and d.f. are typically *unavailable* for many prominent linear and non-linear time series models. To address this issue, Kim, Zhang and Wu (2015) introduce a convolution-type *density* estimate that is used to test for the framework (1.1), regardless of whether or not the closed-form density of X_i is available. They propose a test based on the maximal deviation of this convolution density estimate from the traditional kernel density estimate. However, because of the well-known slow consistency rate of the kernel density estimates, Kim et al. (2015) test leads to size distortion and a low power of the test for moderate sample sizes.

One way to address the issue with the Kim et al. (2015) test is to construct tests based on d.f.'s. More precisely, for any r.v. ξ, let F_ξ, f_ξ denote its d.f. and density, respectively. Let X, ε denote copies of X_0, ε_1, respectively and let $Z := m(X)$. Because of the independence between X and ε in (1.1), we propose a test based on a statistic that compares an estimate for the d.f. F_X based on the convolution between F_Z and F_ε to the traditional empirical d.f. of $X_i, 1 \leq n$. Since both the convolution and empirical d.f. estimates enjoy the $n^{1/2}$-consistency rate, which is faster than that of the kernel density estimate, our test based on these d.f. estimates is expected to perform better than that of the Kim et al. (2015) test in terms of preserving the level of significance in finite sample applications.

The organization of the paper is the following: Section 2 introduces the convolution d.f. estimate of F_X and the test statistic based on the integrated square difference between this estimate and the empirical d.f. estimate of F_X. Section 3 provides the assumptions required for the asymptotics, technical lemmas, and the main theoretic result of the paper. Throughout the paper, we assume that the innovation density in *unknown*. Section 4 derives the asymptotic null distribution of the test statistic of Section 2. The proposed test is compared to competing benchmark tests in the literature. Section 5 provides a simulation study that investigates the finite-sample properties of the proposed statistic based on the block bootstrap methodology. Section 6 concludes the paper.

2 Test statistics

In this section, we combine the ideas of Corradi and Swanson (2005) and Kim et al. (2015) to propose a test statistic based on the empirical d.f. estimate and the convolution-type d.f. estimate of F_X. Throughout the rest of the paper, $Z = m_\theta(X), Z_i = m_\theta(X_i)$, where θ is the true parameter value for which H_0 holds.

Let K be a density kernel, $G(y) = \int_{-\infty}^{y} K(x)dx$, $y \in \mathbb{R}$ and $b \equiv b_n$ be a bandwidth sequence. For $x \in \mathbb{R}$, $\vartheta \in \Theta$, define

$$F_n(x,\vartheta) := \frac{1}{n}\sum_{i=1}^n G\Big(\frac{x - X_i + m_\vartheta(X_{i-1})}{b}\Big), \qquad f_n(x,\vartheta) := \frac{1}{nb}\sum_{i=1}^n K\Big(\frac{x - m_\vartheta(X_{i-1})}{b}\Big).$$

When the null hypothesis H_0 is true, then specifying the true parameter θ among all possible ϑ's will lead to $X_i - m_\vartheta(X_{i-1}) = X_i - m_\theta(X_{i-1}) = \varepsilon_i$. Consequently, we have

$$F_n(x,\theta) = \frac{1}{n}\sum_{i=1}^n G\Big(\frac{x-\varepsilon_i}{b}\Big), \qquad f_n(x,\theta) = \frac{1}{nb}\sum_{i=1}^n K\Big(\frac{x - m_\theta\big(X_{i-1}\big)}{b}\Big).$$

A natural estimate of the d.f. F_X is the empirical d.f. defined by

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n I\Big(X_i \le x\Big), \qquad x \in \mathbb{R}. \tag{2.1}$$

Alternately, we can estimate F_X in the following way. Let $\hat\theta$ be a $n^{1/2}$-consistent estimator of θ, under H_0. Let $\hat\varepsilon_i = X_i - m_{\hat\theta}(X_{i-1})$ and

$$\hat F_\varepsilon(x) := F_n(x,\hat\theta) \equiv \frac{1}{n}\sum_{i=1}^n G\Big(\frac{x-\hat\varepsilon_i}{b}\Big).$$

Next, suppose $Z := m_\theta(X)$ has Lebesgue density denoted by f_Z. Then, under H_0, an estimate of f_Z is given by

$$\hat f_Z(x) := f_n(x,\hat\theta) \equiv \frac{1}{nb}\sum_{i=1}^n K\Big(\frac{x - m_{\hat\theta}\big(X_{i-1}\big)}{b}\Big).$$

Given the independence between X_{i-1} and ε_i, the d.f. $F_X(x) = \int F_\varepsilon(x-z) f_Z(z)dz$ can also be estimated by the convolution d.f.

$$\hat F_c(x) = \int \hat F_\varepsilon(x-z)\hat f_Z(z)dz. \tag{2.2}$$

Because $X_i, i \in \mathbb{Z}$ is stationary and ergodic, by the Ergodic Theorem, the estimator $\widehat{F}_n$ is uniformly consistent for the true d.f. F_X, regardless of whether H_0 is true or not, while, typically under H_0, $\hat F_c(x)$ will be uniformly consistent for F_X. Hence a suitably centered and scaled difference between the two d.f.'s $\widehat{F}_n$ and $\hat F_c$ can be used as a test statistic for testing H_0. Let ψ be a density on $\mathbb{R}$. In this paper, we propose a class of tests, one corresponding to each ψ, based on the statistics

$$\mathcal{V}_n := \int V_n^2(y)\psi(y)dy, \qquad V_n(y) := n^{1/2}\big(\widehat{F}_n(y) - \hat F_c(y)\big), \quad y \in \mathbb{R}. \tag{2.3}$$

We first derive the asymptotic null distribution of $\mathcal{V}_n$ and then illustrate how the bootstrap method can be used to implement the test for moderate sample sizes.

3 Assumptions and some preliminaries

In this section, we shall describe the assumptions and some preliminary results needed for deriving the asymptotic null distribution of V_n. In the sequel, all limits are taken as $n \to \infty$, unless mentioned otherwise. For any Euclidean vector x and any smooth function ℓ from $\mathbb{R}$ to $\mathbb{R}$, let $\|x\|$ denote the Euclidean norm of x, $\|\ell\|_\infty := \sup_{x\in\mathbb{R}} |\ell(x)|$, and $\dot{\ell}$ and $\ddot{\ell}$ denote the first and second derivatives of ℓ, respectively. We are now ready to state the needed assumptions.

Assumption K. The kernel K is a Lebesgue density on $[-1, 1]$, vanishes off $(-1, 1)$, $\int uK(u)du = 0$ and K is differentiable, having a bounded derivative $\dot{K}$.

Assumption M. The process $X_i, i \in \mathbb{Z}$ is strictly stationary, ergodic and strongly mixing with the mixing coefficient sequence α_i satisfying

$$\sum_{i=1}^{\infty} i^{\delta}\alpha_i < \infty, \quad \exists\, \delta > 0. \tag{3.1}$$

Assumption F1. F_ε has Lebesgue density f_ε that is bounded and twice differentiable with the two bounded derivatives $\dot{f}_\varepsilon, \ddot{f}_\varepsilon$.

Assumption F2. The r.v. Z takes values in a bounded set $\mathcal{B}$ and F_Z has Lebesgue density f_Z that is bounded from below on $\mathcal{B}$ and twice differentiable with the two derivatives $\dot{f}_Z, \ddot{f}_Z$ satisfying $\int \big|\dot{f}_Z(z)\big|dz + \int \big|\ddot{f}_Z(z)\big|dz < \infty$, $\|\ddot{f}_Z\|_\infty < \infty$.

Assumption T1. The estimator $\hat{\theta}$ of θ satisfying the following condition.

$$n^{1/2}\|\hat{\theta} - \theta\| = O_p(1), \qquad \text{under } H_0. \tag{3.2}$$

Assumption T2. The estimator $\hat{\theta}$ is asymptotically linear, i.e.,

$$n^{1/2}\big(\hat{\theta} - \theta\big) = n^{-1/2}\sum_{i=1}^{n} \varphi(\varepsilon_i, X_{i-1}, \theta) + o_p(1), \quad \text{under } H_0, \tag{3.3}$$

where $\varphi_i := \varphi(\varepsilon_i, X_{i-1}, \theta), i \geq 1$ are p-vectors of stationary and ergodic r.v.'s with $E\varphi_1 = 0, E\|\varphi_1\|^2 < \infty$ and such that $\lim_n n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} E\big(\varphi_i\varphi_j'\big)$ exists.

Assumption M1. There exists a vector of q functions $\dot{m}_\theta(x), \theta \in \Theta, x \in \mathbb{R}$ such that for every $0 < k < \infty$,

$$\max_{1\leq i\leq n, n^{1/2}\|\vartheta-\theta\|\leq k} n^{1/2}\big|m_\vartheta(X_{i-1}) - m_\theta(X_{i-1}) - (\vartheta - \theta)'\dot{m}_\theta(X_{i-1})\big| = o_p(1), \tag{3.4}$$

$$E\|\dot{m}_\theta(X_0)\|^2 < \infty. \tag{3.5}$$

Assumption M2. $\sup_{x\neq x'} |m_\theta(x) - m_\theta(x')| / |x - x'| < 1$ and $E\left[|\epsilon_i|^p\right] < \infty$, for some $p > 0$.

Assumption b. $b \to 0, n^{1/2}b^3 \to 0$ and $n^{1/2}b^2 \to \infty$.

Assumption K implies that K is bounded. It is satisfied by many popular kernels such as Parzen, Epanechnikov and uniform kernels. Assumption M and (1.1) imply that the process $Z_i := m_\theta(X_i), i \in \mathbb{Z}$ is strongly mixing with the mixing sequence α_i, under H_0. Assumptions F1 and F2 are satisfied by numerous distributions. Assumptions T1 and T2 are important, intermediate assumptions which turn out to be useful for obtaining an approximation of the test statistic and its Gaussian process limit. Under some smoothness conditions on m_θ, it is satisfied by a class of M-estimators, see, e.g., Koul (2002, Ch. 8). Clearly, Assumption T2 implies Assumption T1. Assumption M1 is a minimal smoothness assumption on the null model under which the asymptotic results of this paper are obtained. It is obviously satisfied by $m_\theta(x) = \theta' h(x)$, provided $E\|h(X)\|^2 < \infty$, where $h(x) = (h_1(x), \cdots, h_p(x))'$ is a vector of p real valued functions. Assumption M2 represents a contraction condition. It ensures that the process $X_i, i \in \mathbb{Z}$ is stationary and ergodic, cf. Tong (1991). The Assumption b on the bandwidth sequence implies that $n^{1/2}b \to \infty$. It is in particular satisfied by any $b \propto n^{-1/5}$.

The following lemmas are useful in assessing the asymptotic behavior of $\widehat{F}_n$ and $\hat{F}_c$ and in deriving the asymptotic null distribution of $\mathcal{V}_n$. Their proofs are deferred to Section 7.

Lemma 3.1 *Suppose Assumptions* K, T1, M1 *and* b *hold. Then the followings hold for every θ for which H_0 holds.*

$$\sup_{x\in\mathbb{R}} \big|F_n(x,\hat{\theta}) - F_n(x,\theta)\big| = O_p\big((n^{1/2}b)^{-1}\big), \tag{3.6}$$

$$\sup_{x\in\mathbb{R}} \big|F_n(x,\theta) - EF_n(x,\theta)\big| = O_p\big(n^{-1/2}\big), \tag{3.7}$$

$$\sup_{x\in\mathbb{R}} \big|F_n(x,\hat{\theta}) - F_\varepsilon(x)\big| = O_p\big((n^{1/2}b)^{-1} + n^{-1/2}\big) + O\big(b^2\big) = O_p\big((n^{1/2}b)^{-1}\big). \tag{3.8}$$

Lemma 3.2 *Suppose Assumptions* K, M, F2 *and* b *hold. Then, the following results hold.*

$$\int \big|f_n(z,\theta) - Ef_n(z,\theta)\big| dz = O_p\big((n^{1/2}b)^{-1}\big), \tag{3.9}$$

$$\int \big|f_n(z,\hat{\theta}) - f_Z(z)\big| dz = O_p\big((n^{1/2}b)^{-1}\big) + O_p\big(n^{-1/2}\big) + O_p\big(b^2\big) = O_p\big((n^{1/2}b)^{-1}\big). \tag{3.10}$$

Lemma 3.3 *Assume f_ε, f_Z are twice differentiable with the second order derivatives $\ddot{f}_\varepsilon, \ddot{f}_Z$ bounded. Then the followings hold:*

$$\sup_x \left| \frac{1}{n}\sum_{i=1}^n \Big[\dot{f}_\varepsilon(x - m_{\hat{\theta}}(X_{i-1})) - \dot{f}_\varepsilon(x - m_{\theta_0}(X_{i-1}))\Big] \right| = O_p\Big(\frac{1}{\sqrt{n}}\Big), \tag{3.11}$$

$$\sup_x \left| \frac{1}{n}\sum_{i=1}^n \Big[\dot{f}_Z(x - \hat{\varepsilon}_i) - \dot{f}_Z(x - \varepsilon_i)\Big] \right| = O_p\Big(\frac{1}{\sqrt{n}}\Big). \tag{3.12}$$

Lemma 3.4 *Assume f_ε, f_Z are differentiable with $E\big|\dot f_\varepsilon(x-Z)\big| + E\big|\dot f_Z(x-\varepsilon)\big| < \infty$, for all $x \in \mathbb{R}$. Then*

$$E\big\{\dot f_\varepsilon(x-Z) + \dot f_Z(x-\varepsilon)\big\} = 0, \qquad \forall\, x \in \mathbb{R}. \tag{3.13}$$

Moreover,

$$\sup_x \Big|\frac{1}{n}\sum_{i=1}^n \big[\dot f_\varepsilon\big(x - m_{\hat\theta}(X_{i-1})\big) + \dot f_Z\big(x - \hat\varepsilon_i\big)\big]\Big| \;=\; O_P\big(n^{-1/2}\big). \tag{3.14}$$

Next, let

$$\begin{aligned} \hat S_{n1}(x) &:= n^{-1}\sum_{i=1}^n F_\varepsilon\big(x - m_{\hat\theta}(X_{i-1})\big), \quad & S_{n1}(x) &:= n^{-1}\sum_{i=1}^n F_\varepsilon\big(x - m_\theta(X_{i-1})\big), \\ \hat S_{n2}(x) &:= n^{-1}\sum_{i=1}^n F_Z\big(x - \hat\varepsilon_i\big), \quad & S_{n2}(x) &:= n^{-1}\sum_{i=1}^n F_Z\big(x - \varepsilon_i\big), \quad x \in \mathbb{R}. \end{aligned} \tag{3.15}$$

The following proposition describes the asymptotic uniform linearity of $\hat S_{nj}, j = 1, 2$ in the standardized estimate $n^{1/2}(\hat\theta - \theta)$. Its proof is given in Section 7.

Proposition 3.1 *Under the Assumptions* T1, F1, F2 *and* M1, *the followings hold.*

$$\sup_{x\in\mathbb{R}}\Big|n^{1/2}\big(\hat S_{n1}(x) - S_{n1}(x)\big) - n^{1/2}\big(\hat\theta - \theta\big)' E\big[\dot m_\theta(X) f_\varepsilon(x - m_\theta(X)\big]\Big| = o_p(1). \tag{3.16}$$

$$\sup_{x\in\mathbb{R}}\Big|n^{1/2}\big(\hat S_{n2}(x) - S_{n2}(x)\big) - n^{1/2}\big(\hat\theta - \theta\big)' E\big[\dot m_\theta(X)\big] f_Z(x)\Big| = o_p(1). \tag{3.17}$$

Let

$$\begin{aligned} \mu(x) &:= E\big[\dot m_\theta(X)\big\{f_Z(x) + f_\varepsilon\big(x - Z\big)\big\}\big], \\ \widehat S_n(x) &:= \hat S_{n1}(x) + \hat S_{n2}(x), \quad S_n(x) := S_{n1}(x) + S_{n2}(x). \end{aligned}$$

Proposition 3.2 *Under the Assumptions* K, T1, F1, F2 *and* M2, *the following holds:*

$$\sup_{x\in\mathbb{R}}\Big|\int n^{1/2}\widehat S_n(x - ub)K(u)du - n^{1/2}S_n(x) - n^{1/2}\Delta_n'\mu(x)\Big| = o_p(1). \tag{3.18}$$

Proof. By Proposition 3.1,

$$\sup_{x\in\mathbb{R}}\Big|\int \big[n^{1/2}\widehat S_n(x - ub) - n^{1/2}S_n(x - ub) - n^{1/2}\Delta_n'\mu(x - ub)\big]K(u)du\Big| = o_p(1).$$

By the uniform continuity of f_ε, the continuity of f_Z, and the assumptions that $E\|\dot m_\theta(X)\| < \infty$ and $b \to 0$, we have $\sup_{x\in\mathbb{R}, |u|\le 1} |\mu(x - ub) - \mu(x)| \to 0$.

Observe that

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}}\Big|\int n^{1/2}S_n(x-ub)K(u)du-n^{1/2}S_n(x)\Big|\\
&=(b^2/2)\sup_{x\in\mathbb{R}}\Big|\int n^{-1/2}\sum_{i=1}^n\big\{\dot f_\varepsilon(x-Z_{i-1}-\eta ub)+\dot f_Z(x-\varepsilon_i-\eta ub)\big\}u^2K(u)du\Big|\\
&\le b^2\sup_{y\in\mathbb{R}}\Big|n^{-1/2}\sum_{i=1}^n\big\{\dot f_\varepsilon(y-Z_{i-1})+\dot f_Z(y-\varepsilon_i)\big\}\Big|=O_p(b^2)=o_p(1),
\end{aligned}
$$

where the first equality follows from the Taylor's expansion and $\nu_1=0$, and the inequality holds true due to $\int u^2K(u)du\le 1$ while the last two equalities follow from (3.14) of Lemma 3.4 and $b\to 0$ of Assumption b, respectively. □

The stationarity of $\{X_i, i\in\mathbb{Z}\}$, the independence of X_{i-1} and ε_i for each $i\in\mathbb{Z}$, integration by parts and the change of variable formula yield that for all $x\in\mathbb{R}$,

$$
ES_{n1}(x)=\int F_\varepsilon(x-y)f_Z(y)dy=\int F_Z(x-z)f_\varepsilon(z)dz=ES_{n2}(x)=F_X(x). \tag{3.19}
$$

Next, recall the definition of φ_i from the Assumption T2 and let

$$
\begin{aligned}
\widetilde{T}_n(x)&:=n^{1/2}\big(S_n(x)-2F_X(x)\big)+n^{1/2}\Delta_n'\mu(x)\\
Y_i(x)&:=F_\varepsilon\big(x-m_\theta(X_{i-1})\big)+F_Z(x-\varepsilon_i)-2F_X(x)+\mu(x)'\varphi_i,\quad i\ge 1,\\
T_n(x)&:=n^{-1}\sum_{i=1}^n Y_i(x),\quad x\in\mathbb{R}.
\end{aligned}
$$

The following proposition describes an approximation of $n^{1/2}\big(\hat F_c-F_X\big)$. In the sequel, for any two stochastic processes $U_{n1}(x), U_{n2}(x)$, the statement $U_{n1}(x)=U_{n2}(x)+u_p(1)$ means that $\sup_x|U_{n1}(x)-U_{n2}(x)|=o_p(1)$.

Proposition 3.3 *Under the Assumptions* K, F1, T2, F2, M2 *and* H_0*,*

$$
\sup_{x\in\mathbb{R}} n^{-1/2}\big|\tilde T_n(x)-T_n(x)\big|=o_p(1), \tag{3.20}
$$

$$
\sup_x\big|n^{1/2}\big(\hat F_c(x)-F_X(x)\big)-T_n(x)\big|=o_p(1). \tag{3.21}
$$

Proof. The proof of (3.20) follows immediately from Assumption T2 and (3.3).

Proof of (3.21). Recall $\hat f_Z(z)\equiv f_n(z,\hat\theta)$ and $\hat F_\varepsilon(x)\equiv F_n(x,\hat\theta)$. Let

$$
\begin{aligned}
\mathscr{Z}_n(x)&:=\int\big[\hat F_\varepsilon(x-z)-F_\varepsilon(x-z)\big]\big[\hat f_Z(z)-f_Z(z)\big]dz\\
&=\int\big[F_n\big(x-z,\hat\theta\big)-F_\varepsilon(x-z)\big]\big[f_n\big(z,\hat\theta\big)-f_Z(z)\big]dz.
\end{aligned}
$$

By (3.8), (3.10), and $n^{1/2}b^2 \to \infty$ (guaranteed by Assumption b), we have:

$$\sup_x n^{1/2}\big|\mathcal{Z}_n(x)\big| \le \sup_x n^{1/2}\big|F_n\big(x,\hat\theta\big) - F_\varepsilon(x)\big| \int \big|f_n\big(z,\hat\theta\big) - f_Z(z)\big|dz$$
$$= n^{1/2}O_p\big((n^{1/2}b)^{-1}\big)O_p\big((n^{1/2}b)^{-1}\big) = O_p\big((n^{1/2}b^2)^{-1}\big) = o_p(1). \tag{3.22}$$

We also have the decomposition

$$F_c(x) - F_X(x) = \int F_\varepsilon(x-z)\hat f_Z(z)dz + \int \hat F_\varepsilon(x-z)f_Z(z)dz - 2F_X(x) + \mathcal{Z}_n(x). \tag{3.23}$$

By the change of variable formula,

$$\int F_\varepsilon(x-z)\hat f_Z(z)dz = \int \frac{1}{nb}\sum_{i=1}^n K\Big(\frac{z - m_{\hat\theta}\big(X_{i-1}\big)}{b}\Big)F_\varepsilon(x-z)dz,$$
$$= \int \frac{1}{n}\sum_{i=1}^n F_\varepsilon\Big(x - bu - m_{\hat\theta}\big(X_{i-1}\big)\Big)K(u)du = \int \hat S_{n1}(x-bu)K(u)du.$$

Similarly, integration by parts and the change of variable formula yield

$$\int \hat F_\varepsilon(x-z)f_Z(z)dz$$
$$= \frac{1}{n}\sum_{i=1}^n \int G\Big(\frac{x-z-\hat\varepsilon_i}{b}\Big)dF_Z(z) = \frac{1}{nb}\sum_{i=1}^n \int F_Z(z)K\Big(\frac{x-z-\hat\varepsilon_i}{b}\Big)dz$$
$$= \frac{1}{n}\sum_{i=1}^n \int F_Z\Big(x - ub - \hat\varepsilon_i\Big)K(u)du = \int \hat S_{n2}(x-bu)K(u)du.$$

Thus, by Proposition 3.2, (3.20), (3.22) and (3.23),

$$n^{1/2}\Big(F_c(x) - F_X(x)\Big) = \int n^{1/2}\Big(\widehat S_n(x-bu) - 2F_X(x)\Big)K(u)du + n^{1/2}\mathcal{Z}_n(x)$$
$$= n^{1/2}\Big(S_n(x) - 2F_X(x)\Big) + n^{1/2}\Delta_n'\mu(x) + u_p(1)$$
$$= \tilde T_n(x) + o_p(1) = T_n(x) + u_p(1), \tag{3.24}$$

thereby completing the proof of (3.21) and of the lemma. □.

4 Main Result

In this section, we shall describe the asymptotic null distribution of the test statistic $\mathcal{V}_n$ of (2.3).

Theorem 4.1 *Under the assumptions* (1.1)*,* K, F1, T2, F2, M2 *and* H_0*, for any density* ψ *on* $\mathbb{R}$*,*

$$\mathcal{V}_n = \int V_n^2(x)\psi(x)dx \;\Rightarrow\; \mathcal{V} := \int \mathcal{G}^2(x)\psi(x)dx, \tag{4.1}$$

where $\mathcal{G}(\cdot)$ *is a mean zero Gaussian process with the covariance function* $\widetilde{\mathcal{C}}(x,y)$ *given at* (4.3) *below.*

Proof. Rewrite

$$V_n(x) = n^{1/2}\big(\widehat{F}_n(x) - F_X(x)\big) - n^{1/2}\big(\hat{F}_c(x) - F_X(x)\big). \tag{4.2}$$

Recall the definition of φ_i from the Assumption T2 and let

$$\begin{aligned} \widetilde{Y}_i(x) &:= I\big(X_i \le x\big) - F_X(x) \\ &\qquad - \big\{F_\varepsilon\big(x - m_\theta(X_{i-1}\big) + F_Z(x-\varepsilon_i) - 2F_X(x) + \mu(x)'\varphi_i\big\}, \quad i \ge 1, \\ \widetilde{V}_n(x) &:= n^{-1/2}\sum_{i=1}^n \widetilde{Y}_i(x), \qquad \widetilde{\mathcal{C}}(x,y) := \lim_n \mathrm{Cov}\big(\widetilde{V}_n(x), \widetilde{V}_n(y)\big). \end{aligned} \tag{4.3}$$

Let $\mathcal{G}(x), x \in \mathbb{R}$ be a continuous Gaussian process with mean zero and the covariance function $\mathrm{Cov}\big(\mathcal{G}(x), \mathcal{G}(y)\big) \equiv \widetilde{\mathcal{C}}(x,y)$.

Note that $\widetilde{V}_n(x) \equiv n^{1/2}\big(\widehat{F}_n(x) - F_X(x)\big) - T_n(x)$. By (3.21) of Proposition 3.3,

$$\sup_x \big|V_n(x) - \widetilde{V}_n(x)\big| = \sup_x \big|n^{1/2}\big(\hat{F}_c(x) - F_X(x)\big) - T_n(x)\big| = o_p(1).$$

The r.v. $n^{-1/2}\widetilde{V}_n(x)$ is an average of stationary and ergodic r.v.'s $\widetilde{Y}_i(x)$ where, in view of (3.19) and Assumption T2, $E\widetilde{Y}_0(x) \equiv 0$. Hence, by Theorem 2 in Wu and Shao (2004), $\widetilde{\mathcal{C}}(x,y)$ exists and is finite for all $x, y \in \mathbb{R}$ and $\widetilde{\mathcal{C}}(x,x) > 0$ for all $x \in \mathbb{R}$. Moreover, by Theorem 3 of the same paper, $\widetilde{V}_n(x) \to_D N\big(0, \mathcal{C}(x,x)\big)$ r.v., for every $x \in \mathbb{R}$. By the Cramér-Wold device, all finite dimensional distributions of $\widetilde{V}_n$ converge weakly to those of $\mathcal{G}$.

To complete the proof of (4.1), it remains to prove the tightness of the process $\widetilde{V}_n(x), x \in \mathbb{R}$ in the uniform metric. For this purpose, recall the definition of $S_{nj}, j = 1, 2$ from (3.15) and with $U_n := n^{-1/2}\sum_{i=1}^n \varphi_i$, rewrite

$$\begin{aligned} \widetilde{V}_n(x) &= n^{1/2}\big(\widehat{F}_n(x) - F_X(x)\big) - n^{1/2}\big(S_{n1}(x) - ES_{n1}(x)\big) \\ &\qquad - n^{1/2}\big(S_{n2}(x) - ES_{n2}(x)\big) - \mu(x)'U_n \\ &= T_{n1}(x) - T_{n2}(x) - T_{n3}(x) - \mu(x)'U_n, \quad \text{say.} \end{aligned}$$

The processes $T_{n1}(x), T_{n2}, x \in \mathbb{R}$ are empirical processes of bounded functions of a Markov chain. Their tightness in the uniform metric follows from Levental (1989). Similarly, $T_{n3}(x)$,

$x \in \mathbb{R}$ is an empirical process of bounded functions i.i.d. r.v.'s and its tightness follows from van der Vaart and Wellner (1996). These facts together with the uniform continuity of $\mu(x)$ and $\|U_n\| = O_p(1)$, guaranteed by Assumption T2, imply the tightness of the process $\widetilde{V}_n(x), x \in \mathbb{R}$, thereby completing the proof of (4.1) and of the theorem. $\square$

Remark 4.1 Similarly, we can formulate a test statistic based on density estimation

$$v_{n,b}(x) := \sqrt{nb}\Big(\hat{f}_k(x) - \hat{f}_c(x)\Big) \tag{4.4}$$

where $\hat{f}_k(x)$ is a kernel density estimate for the marginal density $f_X(x)$ of X, while $\hat{f}_c(x)$ is the convolution density estimate of $f_X(x)$. Kim and Wu (2007) proved that under H_0, $v_{n,b}(\cdot) \;\Rightarrow\; N\Big(0, \;\; f_X(\cdot)\int K^2(u)du\Big)$. Hence,

$$\mathcal{K}_n := \int v_{n,b}^2(u)\psi(u)du \;\Rightarrow\; \int \mathcal{Z}^2(u)\psi(u)du, \quad \text{for all densities } \psi \text{ on } \mathbb{R}, \tag{4.5}$$

where $\mathcal{Z}$ is a continuous Gaussian process. Note, that the consistency rate in (4.2) is faster than that in (4.4), given the order of the bandwidth b. Thus, the test based on (4.2) is expected to perform better than the test based on (4.4) in finite sample applications for moderate sample sizes. This is evidenced in the following simulation study.

5 Simulation studies

5.1 Setup

This section reports findings of a finite sample study that compares two tests for the comparison purpose: our proposed test and the benchmark test proposed by Kim et al. (2015). In this simulation study, test statistics considered are $\mathcal{V}_n$ in (4.1) and $\mathcal{K}_n$ in (4.5), respectively, with ψ equal to the uniform density of $(0, 10)$. These tests are similar in that both tests exploit the fact that the deviation from the null hypothesis will lead to the significant discrepancy between the convolution estimate and classical kernel density estimates. On the other hand, the quintessential difference between the two tests lies in the fact that the proposed test based on $\mathcal{V}_n$ employs the d.f.'s while the density functions are used in the benchmark test based on $\mathcal{K}_n$. In this section, we demonstrate that our proposed test outperforms the benchmark test when testing various hypotheses for the chosen sample sizes and alternatives.

Consider the following two sets of hypotheses

$$H_{01}: X_i = 0.5|X_{i-1}| + \varepsilon_i, \qquad H_{a1}: X_i = 0.5|X_{i-1}| + \varepsilon_i\sqrt{0.8(1 + X_{i-1}^2)},$$
$$H_{02}: X_i = 0.1 + 0.5X_{i-1} + \varepsilon_i, \qquad H_{a2}: X_i = 0.1 + 0.5X_{i-1} + 0.8e^{-|X_{i-1}|} + \varepsilon_i,$$

where $\varepsilon_i, 1 \leq i \leq n$ is a random sample from $N(0,1)$ distribution. The null model H_{01} is the threshold autoregressive (TAR) model (Tong, 1990) of order one, while the alternative H_{a1} is the TAR-ARCH(1,1) model (Zakoïan, 1994). The purpose of testing hypotheses H_{01} versus H_{a1} is to see whether our test can differentiate the conditional homoscedastic process (i.e. TAR) from the conditional heteroscedastic one (i.e. TAR-ARCH). The hypotheses H_{02} and H_{a2} simply test whether the autoregressive function is linear or non-linear function of the lag variable.

For $T = 200, 400, 600$, we generate $\{X_i : i = 1, 2, ..., T\}$ according to the given null hypothesis with the initial value $X_0 = 0.8$. When we compare the performance of the proposed test with that of the benchmark test, empirical level and power will be used to assess the performance. For the benchmark test, we emulate Monte Carlo simulation as described in Kim et al. (2015) while we employ block-wise bootstrap method proposed by Künsch (1989) and Liu and Singh (1992) for the proposed test. We first determine the size of the block l_B, so that the number of blocks, n_B, is T/l_B. Naik-Nimbalakar and Rajarshi (1994) showed that the weak convergence of block-wise bootstrapped empirical process depends on the order of the l_B. They obtained desired results when $l_B = O(n^{1/2} - \epsilon)$, with $0 < \epsilon < \frac{1}{2}$. Motivated by their work, we chose $l_B = \{8, 10, 16, 20\}$ in this finite sample study. We found that the proposed test displays the optimal result when $l_B = 10$, for all T. Once we determine the value of l_B, we construct a block: we draw any uniform random number between 1 and $T - l_B + 1$, say k, and choose l_B consecutive observations, $X_{k+1}, ..., X_{k+l_B}$. We repeat constructing a block n_B times, combine these n_B blocks all together, and obtain re-sampled observations, $X_1^*, ..., X_T^*$. For the calculation of the statistics, we use uniform kernel function $K(u) := 2^{-1} I(|u| \leq 1)$. Therefore,

$$G(u) = \int_{-\infty}^{u} K(x)dx = \begin{cases} 0, & u < -1; \\ \frac{u+1}{2}, & -1 \leq u < 1; \\ 1, & u \geq 1. \end{cases}$$

For $\hat{\theta}$, we use least squares estimator. Define $h(t) := (-t^2 + 2c_i t)/8b$ where $c_i := x + b - \hat{\varepsilon}_i$. Then $\hat{F}_c$ in (2.2) can be rewritten as

$$\hat{F}_c(x) = \frac{1}{n^2 b} \sum_{i=1}^{n} \sum_{j=1}^{n} \int G\Big(\frac{x - t - \hat{\varepsilon}_i}{b}\Big) K\Big(\frac{x - m_{\hat{\theta}}\big(X_{j-1}\big)}{b}\Big)\, dt = \frac{1}{n^2 b} \sum_{i=1}^{n} \sum_{j=1}^{n} IF_{ij}(x),$$

where, with $Z_{ij} := m_{\hat{\theta}}\big(X_{j-1}\big) + \hat{\varepsilon}_i$,

$$IF_{ij}(x) = \begin{cases} 0, & x < Z_{ij} - 2b, \\ h(x - \hat{\varepsilon}_i + b) - h(m_{\hat{\theta}}(X_{j-1}) - b), & Z_{ij} - 2b \leq x < Z_{ij}, \\ bG((x - Z_{ij} - b)/b) + h(m_{\hat{\theta}}(X_{j-1}) + b) & \\ \qquad\qquad -h(x - \hat{\varepsilon}_i - b), & Z_{ij} \leq x < Z_{ij} + 2b, \\ b, & \text{otherwise}. \end{cases}$$

Consequently, the great deal of simplification in the computation of $V_n(x)$ in (2.3) follows directly.

Define the bootstrap test statistic:

$$\mathcal{V}_n^* = \int_0^{10} \big(V_n^*(x) - V_n(x)\big)^2 dx \tag{5.1}$$

where $V_n^*(x)$ denotes the counterpart of $V_n(x)$, which is obtained from re-sampled observations. We repeat block-wise bootstrap B_{Iter} times, obtain $\mathcal{V}_n^*$'s, and calculate $100(1-\alpha)$ percentiles, $q_{1-\alpha}^*$, obtained from these values of $\mathcal{V}_n^*$'s. As various l_B's are tried, so are B_{Iter}'s. Our findings show that empirical levels approaches more closely to the suggested significance level α as B_{Iter} increases. See, e.g., Table 1. After $q_{1-\alpha}^*$ is obtained, we reject H_0 if $\mathcal{V}_n > q_{1-\alpha}^*$. As a final step, we repeat this procedure 1,000 times, count the number of rejections, and obtain empirical levels and powers by dividing it by 1,000.

5.2 Selection of B_{Iter}, b, and, l_B

In the simulation study, $b = \{0.05, 0.1, 0.15, 0.2\}$ are tried for the bandwidth. Since the choice of b does not affect the powers and levels much, we only report the result corresponding to $b = 0.1$. Tables 1 and 2 report empirical levels of the proposed test for H_{01} and H_{02} corresponding to various sizes of blocks and numbers of bootstrap iterations. As shown in the tables, we obtain the optimal results at $(l_B, B_{Iter}) = (10, 200)$ and $(l_B, B_{Iter}) = (20, 160)$ for H_{01} and H_{02}, respectively. Therefore, we use these optimal values in the following simulation studies. One point worth noting here is that the empirical levels of H_{01} show good convergence to corresponding α's while the same conclusion does not hold for H_{02}. For H_{01}, the empirical levels corresponding to $(l_B, B_{Iter}) = (10, 200)$ are 0.045 and 0.011 which are quite close to $\alpha = 0.05$ and 0.01, respectively. The counterparts of H_{02} corresponding to $(l_B, B_{Iter}) = (20, 160)$ are 0.094 and 0.037. In view of this, we conjecture that the proposed test for H_{02} might not be as efficient as it is for H_{01}: we will discuss more about this in the next section.

5.3 Tests $\mathcal{V}_n$ and $\mathcal{K}_n$ for H_{01} and H_{02}

This section compares the proposed test $\mathcal{V}_n$ and the benchmark test $\mathcal{K}_n$ for H_{01} and H_{02}. Tables 3 and 4 report their empirical levels and powers when $T = 200,\ 400,\ 600$ and $\alpha = 0.01,\ 0.05$. To begin with, consider the result for H_{01}. It is hard to tell which test is superior in terms of the level. However, there is no room for argument in terms of the power: the proposed test dominates the benchmark test. When $T = 200$, the differences in the power between two tests are approximately 0.3 and 0.17 for $\alpha = 0.05$ and 0.01, respectively. For other T's (400 and 600), the differences, however, decreases to approximately 0.19 and 0.15

	$l_B = 8$					$l_B = 10$				
α	$B_{Iter} = 40$	80	120	160	200	$B_{Iter} = 40$	80	120	160	200
0.05	0.079	0.072	0.063	0.059	0.055	0.074	0.064	0.053	0.047	0.045
0.01	0.050	0.025	0.016	0.014	0.016	0.029	0.013	0.012	0.014	0.011
	$l_B = 16$					$l_B = 20$				
α	$B_{Iter} = 40$	80	120	160	200	$B_{Iter} = 40$	80	120	160	200
0.05	0.064	0.057	0.055	0.054	0.052	0.097	0.076	0.075	0.067	0.065
0.01	0.025	0.015	0.015	0.012	0.013	0.058	0.029	0.020	0.019	0.016

Table 1: Levels of H_{01} when B_{Iter} and l_B vary with T being fixed at 400.

	$l_B = 8$					$l_B = 10$				
α	$B_{Iter} = 40$	80	120	160	200	$B_{Iter} = 40$	80	120	160	200
0.05	0.134	0.123	0.120	0.144	0.114	0.137	0.129	0.117	0.092	0.124
0.01	0.078	0.057	0.052	0.065	0.042	0.076	0.059	0.047	0.044	0.047
	$l_B = 16$					$l_B = 20$				
α	$B_{Iter} = 40$	80	120	160	200	$B_{Iter} = 40$	80	120	160	200
0.05	0.131	0.104	0.104	0.107	0.102	0.112	0.103	0.123	0.094	0.094
0.01	0.064	0.045	0.038	0.037	0.045	0.051	0.045	0.055	0.030	0.037

Table 2: Levels of H_{02} when B_{Iter} and l_B vary with T being fixed at 400.

for $\alpha = 0.05$ and 0.01, respectively. Note that the benchmark test does not obtain the power larger than 0.8 even when T reaches 600: 0.747 and 0.692 for α=0.05 and 0.01, respectively. On the contrary, the proposed test with $\alpha = 0.05$ accomplishes the power larger than 0.9 when $T = 400$.

	α	$T = 200$		$T = 400$		$T = 600$	
		$\mathcal{V}_n$	$\mathcal{K}_n$	$\mathcal{V}_n$	$\mathcal{K}_n$	$\mathcal{V}_n$	$\mathcal{K}_n$
Level	0.05	0.041	0.052	0.045	0.050	0.058	0.054
	0.01	0.013	0.007	0.013	0.010	0.012	0.007
Power	0.05	0.658	0.343	0.901	0.710	0.933	0.747
	0.01	0.387	0.215	0.714	0.561	0.854	0.692

Table 3: The proposed $\mathcal{V}_n$ and benchmark $\mathcal{K}_n$ tests for H_{01}

Next, we proceed to analyze the result for H_{02} which is reported in Table 4. A quick glance at the table reveals that the proposed and benchmark tests for H_{02} display poor performances (especially, power) for all α and T when compared with the result for H_{01}.

In the case of empirical level, both tests fail to show a sign of any convergence to α, as T increases. For example, the empirical level of the proposed test closest to $\alpha = .01$ is 0.031 when $T = 600$, which is still far away from 0.01.

While the empirical power of both tests for H_{02} is much smaller than worse that for H_{01}, there exists a stark difference between the two tests. The extent of decrease in the power displayed by the benchmark test $\mathcal{K}_n$ is extremely large, compared with that of the proposed test $\mathcal{V}_n$. For example, the empirical power of the $\mathcal{V}_n$ test corresponding to $T = 600$ and $\alpha = 0.05$ decreased from 0.933 to 0.702 while the power of the $\mathcal{K}_n$ test plummets from 0.747 to 0.233. Even though both tests show consistency, that is, display an increase in the power as T increases, the benchmark test always yields a power below 0.3 while the power of the proposed test starts with 0.424 and rises to 0.702 as T changes from 400 to 600. In summary, the proposed test for H_{02} does not retain the same efficiency as that for H_{01} anymore, but it still outperforms the benchmark test.

	α	$T = 200$		$T = 400$		$T = 600$	
		$\mathcal{V}_n$	$\mathcal{K}_n$	$\mathcal{V}_n$	$\mathcal{K}_n$	$\mathcal{V}_n$	$\mathcal{K}_n$
Level	0.05	0.134	0.023	0.096	0.013	0.091	0.013
	0.01	0.065	0.003	0.032	0.002	0.031	0.004
Power	0.05	0.424	0.045	0.617	0.120	0.702	0.233
	0.01	0.260	0.015	0.433	0.072	0.498	0.109

Table 4: The proposed ($\mathcal{V}_n$) and benchmark ($\mathcal{K}_n$) tests for H_{02}

This result accords closely with the fact that both tests assume the independence between the error term and the lagged X_i's. The presence of the dependence between them as in the alternative H_{1a} will lead to the instant rejection of the null H_{10} and beget a larger statistical power than would otherwise be the case. Unlike H_{1a}, the alternative H_{2a} does not violate the assumption of the independence, and hence, it is plausible for both tests to return smaller powers. Based on all findings in this section, it is undoubtedly the case that the proposed test is indeed superior to the benchmark test.

6 Conclusion

In this paper, we propose a test for a parametric specification of the autoregressive function. This test is based on the integrated square difference between the empirical d.f. and the convolution d.f. estimates of the stationary d.f. We prove the weak convergence of a suitably standardized convolution d.f. estimate process to a Gaussian process. This in turn is used to derive the asymptotic null distribution of the proposed test statistic. The consistency rate of this test statistic is $n^{1/2}$. The block bootstrap approach is employed to run the proposed test under a Monte Carlo setting. As expected, the simulation study shows the superiority of the proposed test over the benchmark test of Kim et al. (2015) based on density estimates in term of the empirical level and power.

The paper can be extended in the following directions. First, the framework considered in the current study can be extended to cover continuous-time processes. The continuous-time processes have played an important role in economic time series analysis, due to their usefulness in modeling financial data of high frequency. Some of the recent studies focus on the inference of potentially non-stationary continuous-time processes, see, e.g., Aït-Sahalia and Park (2012, 2016). The methodology suggested in the current study can be extended to this end. Secondly, a theoretical justification of the bootstrap methodology employed in the simulation study needs to be provided. The current study employs the bootstrap to handle the issue of non-pivotal asymptotic distribution for the proposed test statistic. A formal justification of the approach will make the simulation result more convincing.

7 Proofs

This section contains the proofs of Lemmas 3.1–3.4 and Proposition 3.1. Before proceeding to the proof, we shall discuss some implications of the Assumption M1, which are often used in the proofs below. Let

$$\begin{aligned}(7.1)\qquad & d_i := m_{\hat\theta}(X_{i-1}) - m_\theta(X_{i-1}), \qquad \Delta_n := \hat\theta - \theta, \qquad D_n := \max_{1\le i\le n} |d_i|,\\ & \delta_i := d_i - \Delta_n' \dot m_\theta(X_{i-1}), \quad 1 \le i \le n.\end{aligned}$$

By (3.5), stationarity, ergodicity and the Ergodic Theorem imply that

$$\max_{1\le i\le n} n^{-1/2}\|\dot m_\theta(X_{i-1})\| = o_p(1), \qquad n^{-1}\sum_{i=1}^n \dot m_\theta(X_{i-1}) = E\dot m_\theta(X_0) + o_p(1).$$

Together with these facts, assumptions (3.2) and (3.4), in turn, yield the following results.

$$\begin{aligned}(7.2)\qquad & \max_{1\le i\le n} |\Delta_n'\dot m_\theta(X_{i-1})| \le n^{1/2}\|\Delta_n\| \max_{1\le i\le n} n^{-1/2}\|\dot m_\theta(X_{i-1})\| = o_p(1), \quad D_n = o_p(1)\\ & \max_{1\le i\le n} \big|\delta_i\big| = \max_{1\le i\le n} \big| m_{\hat\theta}(X_{i-1}) - m_\theta(X_{i-1}) - \Delta_n'\dot m_\theta(X_{i-1})\big| = o_p(n^{-1/2}),\\ & n^{-1/2}\sum_{i=1}^n d_i = n^{1/2}\Delta_n' n^{-1}\sum_{i=1}^n \dot m_\theta(X_{i-1}) + o_p(1) = n^{1/2}\Delta_n' E\dot m_\theta(X_0) + o_p(1) = O_p(1),\\ & n^{-1/2}\sum_{i=1}^n |d_i| \le n^{-1/2}\sum_{i=1}^n \big|d_i - \Delta_n'\dot m_\theta(X_{i-1})\big| + n^{1/2}\|\Delta_n\| n^{-1}\sum_{i=1}^n \big\|\dot m_\theta(X_{i-1})\|\\ & \qquad = O_p(1).\end{aligned}$$

In the proofs below, $\sup_x \equiv \sup_{x\in\mathbb{R}}$, unless mentioned otherwise. We also use the notation $\nu_j := \int u^j K(u)du$, $j = 1, 2$. Under Assumption K, $\nu_1 = 0$ and $\nu_2 \le 1$.

Proof of Lemma 3.1. Because $\kappa := \|K\|_\infty < \infty$,

$$\big|G(u) - G(v)\big| \;=\; \Big|\int_v^u K(y)dy\Big| \le |u-v|\kappa, \quad \forall\, u, v \in \mathbb{R}.$$

Hence

$$\begin{aligned}\sup_x \big|F_n(x,\hat\theta) - F_n(x,\theta)\big| &\le \sup_x \frac{1}{n}\sum_{i=1}^n \Big|G\Big(\frac{x-\hat\varepsilon_i}{b}\Big) - G\Big(\frac{x-\varepsilon_i}{b}\Big)\Big| \\ &\le \frac{\kappa}{n}\sum_{i=1}^n \Big|\frac{\varepsilon_i - \hat\varepsilon_i}{b}\Big| = \frac{1}{nb}\sum_{i=1}^n |d_i| = O_p\Big((n^{1/2}b)^{-1}\Big),\end{aligned}$$

by (7.2). This completes the proof of (3.6).

Proof of (3.7). The stochastic process $n^{1/2}\big(F_n(x,\theta) - EF_n(x,\theta)\big), x \in \mathbb{R}$ is a mean zero empirical process of bounded functions of i.i.d. r.v.s' ε_i, which are known to converge weakly to a Gaussian process in the uniform metric, see, e.g., van der Vaart and Wellner (1996). This fact in turn implies (3.7).

Proof of (3.8). The left hand side of (3.8) is bounded from the above by

$$\sup_x \big|\big(F_n(x,\hat\theta) - F_n(x,\theta)\big| + \sup_x \big|F_n(x,\theta) - EF_n(x,\theta)\big| + \sup_x \big|EF_n(x,\theta) - F_\varepsilon(x)\big|. \tag{7.3}$$

Consider the third term in this bound. The integration by parts, the change of variable, $\nu_1 = 0$, $\dot f_\varepsilon$ being bounded and the Taylor expansion yields that for some $0 < \eta < 1$,

$$\begin{aligned}&EF_n(x,\theta) - F_\varepsilon(x) = \int \big[F_\varepsilon(x-bu) - F_\varepsilon(x)\big]K(u)du = \frac{b^2}{2}\int \dot f_\varepsilon(x-\eta bu)K(u)du, \\ &\sup_x \big|EF_n(x,\theta) - F_\varepsilon(x)\big| \le \big\|\dot f_\varepsilon\|_\infty b^2.\end{aligned}$$

This bound together with the assumption $n^{1/2}b^3 \to 0$, guaranteed by Assumption b, (3.7), (3.8) and (7.3) completes the proof of (3.8), and that of the lemma. □

Proof of Lemma 3.2. Assumption M and (1.1) imply that, under H_0, the sequence $Z_i \equiv m_\theta(X_i), i \in \mathbb{Z}$ is stationary and ergodic α-mixing with the mixing sequence α_i satisfying (3.1). This fact and the compactness of the support of f_Z make Theorem 4.2.2(iii) of Györfi et al. (1989) applicable, which yields that the expected value of the left hand side of (3.9) is of the order $O\big((n^{1/2}b)^{-1}\big)$, which together with the Markov inequality implies (3.9). Note that this result does not need the smoothness of f_Z and K.

To prove (3.10), let

$$A_n := \int \big|f_n(z,\theta) - Ef_n(z,\theta)\big|dz, \qquad B_n := \int \big|Ef_n(z,\theta) - f_Z(z)\big|dz.$$

By the triangle inequality and (3.9),

$$\int \big|f_n(z,\hat\theta) - f_Z(z)\big|dz \le \int \big|f_n(z,\hat\theta) - f_n(z,\theta)\big|dz + A_n + B_n$$
$$= \int \big|f_n(z,\hat\theta) - f_n(z,\theta)\big|dz + O_p\big((n^{1/2}b)^{-1}\big) + B_n. \tag{7.4}$$

To obtain a bound on B_n, let $\gamma(z,\eta) := \int u^2 K(u)\ddot f_Z(z-\eta bu)du$, $z\in\mathbb{R}, 0<\eta<1$. With $Z_i = m_\theta(X_i)$, rewrite

$$f_n(z,\theta) = \frac{1}{nb}\sum_{i=1}^n K\Big(\frac{z - Z_{i-1}}{b}\Big).$$

Because $\nu_1 = 0$, by the Taylor expansion, for some $0<\eta<1$,

$$EK\Big(\frac{z-Z}{b}\Big) = b\int K(u) f_Z(z-ub)du$$
$$= b\int K(u)\left[f_Z(z) - ub\dot f_Z(z) + \frac{u^2b^2}{2}\ddot f_Z(z-\eta bu)\right]du = bf_Z(z) + \frac{b^3}{2}\gamma(z,\eta).$$

Hence, by stationarity,

$$Ef_n(z,\theta) = \frac{1}{b}EK\Big(\frac{z-Z}{b}\Big) = f_Z(z) + \frac{b^2}{2}\gamma(z,\eta).$$

Let $J_{2Z} := \int \big|\ddot f_Z(z)\big|dz$. By Assumption F2, $J_{2Z}<\infty$. By the Fubini Theorem,

$$\int |\gamma(z,\eta)|dz = \int\int \big|\ddot f_Z(z-\eta bu)\big|dz\, u^2K(u)du = \nu_2 J_{2Z}, \quad \forall\, n\ge 1.$$

Hence

$$B_n = \int \big|Ef_n(z,\theta) - f_Z(z)\big|dx = \frac{b^2}{2}\int |\gamma(z,\eta)|dx = \frac{b^2}{2}\nu_2 J_{2Z}, \quad \forall\, n\ge 1. \tag{7.5}$$

Next, K having bounded derivative implies that

$$\int \big|f_n(z,\hat\theta) - f_n\big(z,\theta\big)\big|dz$$
$$\le \frac{1}{nb}\sum_{i=1}^n \int \Big|K\Big(\frac{z - m_{\hat\theta}(X_{i-1})}{b}\Big) - K\Big(\frac{z - m_\theta(X_{i-1})}{b}\Big)\Big|dz$$
$$= \frac{1}{n}\sum_{i=1}^n \int \Big|K(u) - K\Big(u + \frac{d_i}{b}\Big)\Big|du \le \frac{\|\dot K\|_\infty}{n}\sum_{i=1}^n |d_i| = O_p\big(n^{-1/2}\big),$$

by (7.2). This fact together with (7.5) and (7.4) completes the proof of the lemma. □

Proof of Lemma 3.3. Because $\ddot f_\varepsilon$ is bounded, by the mean value theorem, the left hand side of (3.11) is bounded from the above by $\|\ddot f_\varepsilon\|_\infty n^{-1}\sum_{i=1}^n |d_i| = O_p(n^{-1/2})$, where the equality directly follows from (7.2). Similarly, the left hand side of (3.12) is bounded from the above by $\|\ddot{f}_Z\|_\infty n^{-1}\sum_{i=1}^n |\hat{\varepsilon}_i - \varepsilon_i| = n^{-1}\sum_{i=1}^n |d_i| = O_p(n^{-1/2})$. □

Proof of Lemma 3.4. The integration by parts and the change of variables yield that

$$\begin{aligned} E\big[\dot{f}_Z(x-\varepsilon)\big] &= \int \dot{f}_Z(x-v) f_\varepsilon(v)dv = -f_Z(x-v)f_\varepsilon(v)\Big|_{v=-\infty}^{v=\infty} + \int f_Z(x-v)\dot{f}_\varepsilon(v)dv \\ &= -\int f_Z(z)\dot{f}_\varepsilon(x-z)dz = -E\dot{f}_\varepsilon(x-Z), \quad \forall\, x\in\mathbb{R}. \end{aligned}$$

By Lemmas 3.3 and (3.13), the left hand side of (3.14) is bounded from the above by

$$\begin{aligned} &\frac{1}{n}\Big|\sum_{i=1}^n \big[\dot{f}_\varepsilon\big(x-m_\theta(X_{i-1})\big) - E\dot{f}_\varepsilon\big(x-m_\theta(X)\big) + \dot{f}_Z\big(x-\varepsilon_i\big) - E\dot{f}_Z\big(x-\varepsilon\big)\big]\Big| + O_p\big(n^{-1/2}\big) \\ &= O_p\big(n^{-1/2}\big), \end{aligned}$$

by the Ergodic Theorem. □

Proof of Proposition 3.1. Under Assumption F1, $\dot{f}_\varepsilon$ is bounded, which in turn implies that f_ε is uniformly continuous. Hence, by (7.2),

$$\begin{aligned} (7.6)\qquad &\sup_{x\in\mathbb{R}} n^{-1/2}\Big|\sum_{i=1}^n \big\{F_\varepsilon\big(x-m_{\hat\theta}(X_{i-1})\big) - F_\varepsilon\big(x-m_\theta(X_{i-1})\big) - d_i f_\varepsilon\big(x-m_\theta(X_{i-1})\big)\big\}\Big| \\ &= \sup_{x\in\mathbb{R}} n^{-1/2}\Big|\sum_{i=1}^n \int_0^{d_i} [f_\varepsilon\big(x-m_\theta(X_{i-1})-s\big) - f_\varepsilon(x-m_\theta(X_{i-1}))]ds\Big| \\ &\le n^{-1/2}\sum_{i=1}^n |d_i| \sup_{|y-z|\le D_n} \big|f_\varepsilon(y) - f_\varepsilon(z)\big| = o_p(1). \end{aligned}$$

Next, (7.2) and f_ε being bounded, guaranteed by Assumption F1, readily imply

$$(7.7)\qquad \sup_{x\in\mathbb{R}}\Big|n^{-1/2}\sum_{i=1}^n \delta_i\, f_\varepsilon\big(x-m_\theta(X_{i-1})\big)\Big| \;\le\; n^{1/2}\max_{1\le i\le n}\big|\delta_i\big|\sup_{y\in\mathbb{R}} f_\varepsilon(y) = o_p(1).$$

Thus, by (7.6) and (7.7),

$$\begin{aligned} n^{1/2}(\hat{S}_{n1}(x) - S_{n1}(x)) &= n^{1/2}\Delta_n' n^{-1}\sum_{i=1}^n \dot{m}_\theta(X_{i-1}) f_\varepsilon(x-m_\theta(X_{i-1})) + o_p(1) \\ &= n^{1/2}\Delta_n' E\big(\dot{m}_\theta(X) f_\varepsilon(x-m_\theta(X)\big) + o_p(1), \end{aligned}$$

by the Ergodic Theorem, thereby proving (3.16). The proof of (3.17) is exactly similar. □

References

[1] Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies* **9**, 385–426.

[2] Aït-Sahalia, Y. and Park, J.Y. (2012). Stationarity-based specification tests for diffusions when the process is nonstationary. *J. Econometrics* **169**, 279–292.

[3] Aït-Sahalia, Y. and Park, J.Y. (2016). Bandwidth selection and asymptotic properties of local nonparametric estimators in possibly nonstationary continuous-time models. *J. Econometrics* **192**, 119–138.

[4] Corradi, V. and Swanson, N.R. (2005). Bootstrap specification tests for diffusion processes. *J. Econometrics* **124**, 117–148.

[5] Györfi, L., Härdle, W., Sards, P. and Vieu, P. (1989). *Nonparametrie curve estimation from time series*. Springer-Verlag, Heidelberg, Germany.

[6] Kim, K.H. and Wu, W.B. (2007). Density estimation for nonlinear time series. manuscript.

[7] Kim, K.H., Zhang, T. and Wu, W.B. (2015). Parametric specification test for nonlinear autoregressive models. *Econometric Theory* **31**, 1078–1101.

[8] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, **17**, 1217-1241.

[9] Lahiri, S.N. (1992). Edgeworth correction by "moving block" bootstrap for stationary and nonstationary data. Exploring the limits of bootstrap, 183214, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., Wiley-Intersci. Publ., Wiley, New York.

[10] Lahiri, S. N. (3003). *Resampling methods for dependent data*. Springer Series in Statistics. Springer-Verlag, New York.

[11] Levental, S. (1989). A Uniform CLT for uniformly bounded families of martingale differences. *J. Theoretical Probability*, **2(3)**, 171–187.

[12] Liu, R.Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (R. Lepage and L. Billard, eds.) 225-248. Wiley, New York.

[13] Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies* **11**, 449–487.

[14] Stanton, R. (1997). A Nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* **52**, 1973–2002.

[15] Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press.

[16] van der Vaart, A.W. and Wellner, J. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer-Verlag, New York.

[17] Wu, W.B. and Shao, X.F. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability* **41**, 425–436.