Specification testing in nonparametric AR-ARCH models

Marie Hušková\textsuperscript{1} | Natalie Neumeyer\textsuperscript{2} | Tobias Niebuhr\textsuperscript{2} | Leonie Selk\textsuperscript{2}

\textsuperscript{1}Department of Statistics, Charles University, Prague, Czech Republic
\textsuperscript{2}Department of Mathematics, University of Hamburg, Hamburg, Germany

**Correspondence**
Natalie Neumeyer, Department of Mathematics, University of Hamburg, 20146 Hamburg, Germany.
Email: Natalie.Neumeyer@math.uni-hamburg.de

**Funding information**
Grantová agentura České republiky, Deutsche Forschungsgemeinschaft

**Abstract**
In this paper, an autoregressive time series model with conditional heteroscedasticity is considered, where both conditional mean and conditional variance function are modeled nonparametrically. Tests for the model assumption of independence of innovations from past time series values are suggested. Tests based on weighted $L^2$-distances of empirical characteristic functions are considered as well as a Cramér–von Mises-type test. The asymptotic distributions under the null hypothesis of independence are derived, and the consistency against fixed alternatives is shown. A smooth autoregressive residual bootstrap procedure is suggested, and its performance is shown in a simulation study.

**KEYWORDS**
autoregression, conditional heteroscedasticity, empirical characteristic function, kernel estimation, nonparametric CHARN model, testing independence

**1 | INTRODUCTION**

Assume that we have observations from a one-dimensional stationary weakly dependent time series $X_j$, $j \in \mathbb{Z}$. Nonparametric modeling avoids misspecification problems, and thus, such models have gained much attention over the last years, see Fan and Yao (2003) and Gao (2007) for extensive overviews. One popular possibility is to analyze data by fitting a nonparametric AR(1)-ARCH(1)-model (also called CHARN-model), that is,

$$X_j = m(X_{j-1}) + \sigma(X_{j-1}) \epsilon_j, \quad j \in \mathbb{Z},$$

(1)
with autoregression function $m(x) = \mathbb{E}[X_j \mid X_{j-1} = x]$, conditional variance function $\sigma^2(x) = \text{Var}(X_j \mid X_{j-1} = x)$, and innovations $\varepsilon_j$, independent from past time series values $X_{j-1}, X_{j-2}, \ldots$. Before applying any procedure developed for a time series model like the one defined, model assumptions need to be tested. Thus, we are interested in testing the hypothesis

$$H_0: \varepsilon_j \text{ and } (X_{j-1}, X_{j-2}, \ldots) \text{ are stochastically independent.}$$

Although testing for this model assumption is essential for applications to obtain correct forecasts, it seems that the problem has not been considered before in the literature for the nonparametric case. The reason is presumably that tests for hypotheses involving the innovation distribution would typically be based on the empirical distribution function of nonparametrically estimated innovations (residuals). Only recently, asymptotic results for such processes in nonparametric autoregressive models are available. Müller, Schick, and Wefelmeyer (2009) consider the above model in the homoscedastic case with constant $\sigma$. They prove an asymptotic expansion of the empirical process of residuals obtained from local-polynomial estimation of the autoregression function $m$. Furthermore, Dette, Pardo–Fernández, and Van Keilegom (2009) base a test for the multiplicativity hypothesis $m = c \sigma$ on the estimated innovation distribution. Selk and Neumeyer (2013) consider a sequential empirical process of residuals and apply it to test for a change point in the innovation distribution. In order to test an implication of the null hypothesis $H_0$, one can consider, for some fixed and prespecified $k \in \mathbb{N}$, test statistics based on an estimated difference of the joint empirical distribution function of $\varepsilon_j$ and $(X_{j-1}, \ldots, X_{j-k})$ and the product of the marginal distributions. We follow this approach and consider a Cramér–von Mises-type test statistic and prove its asymptotic distribution under the null hypothesis. Note, however, that the assumptions for deriving asymptotic distributions of residual-based processes here and in the aforementioned literature are very restrictive. To avoid unnecessarily strong assumptions, we additionally follow a different path and base tests on an estimated weighted $L^2$-distance between the joint and the marginal characteristic functions of $\varepsilon_j$ and $(X_{j-1}, \ldots, X_{j-k})$. Similarly, in a time series context but for a parametric model Hlávka, Hušková, Kirch, and Meintanis (2012) test for a change in the innovation distribution of a linear autoregression model based on residual empirical characteristic functions. A survey of testing procedures based on empirical characteristic functions is given in Meintanis (2016).

From a certain point of view, the present paper can be viewed as a dynamic extension of specification tests on nonparametric location/scale models considered in the papers by Akritas and Van Keilegom (2001), Einmahl and Van Keilegom (2008), and Hlávka, Hušková, and Meintanis (2011) in iid context. The two former ones elaborate procedures on the basis of empirical distribution functions, whereas the last one employs empirical characteristic functions.

Next, we mention some important papers connected with AR-ARCH models. Particularly, asymptotic theory for vector ARMA-GARCH models in parametric setup is investigated in Ling and McAleer (2003). Lange, Rahbek, and Jensen (2011) studied asymptotic properties of quasi-maximum likelihood estimators and their modifications also in parametric setup. Carrasco, Chernov, Florens, and Ghysels (2007) studied estimation in parametric AR-ARCH model-based characteristic functions from the point of view of certain optimality. Escanciano, Lobato, and Zhu (2013) introduces an automatic test for the correct specification of parametric vector autoregression (VAR) models. The proposed test statistic is a Portmanteau statistic with an automatic selection of the order of the residual serial correlation tested. Du and Escanciano (2015) develop a nonparametric distribution-free test for serial independence of errors in parametric location/scale model, where the location and scale functions are known up to finite-dimensional parameters.
The focus is on regression setup with possible modification for a time series setup. Ghoudi, Kulperger, and Rémillard (2001) present nonparametric tests of independence that can be used to test the independence of $p$ random variables or serial independence for time series or residuals.

The remainder of the paper is organized as follows. In Section 2, we define the estimators and the test statistic on the basis of empirical characteristic functions. We state model assumptions, give the asymptotic distribution of the test statistic under the null hypothesis, and discuss consistency under fixed alternatives. In Section 3, we consider a Cramér–von Mises-type test based on the empirical distribution functions. A bootstrap procedure is suggested in Section 4, where also the finite-sample performance is investigated in a simulation study and by means of a real data example. Section 5 concludes the paper. Proofs of the main results are in Appendix A, while Appendix B contains auxiliary results.

2 | A TEST BASED ON EMPirical CHARACTERISTIC FUNCTIONS

2.1 | The test statistic

Assume that we have observations $X_{-k+1}, \ldots, X_n$ from the time series $X_j, j \in \mathbb{Z}$, considered in Equation (1). As test statistic for independence of innovations and past time series values, we consider the weighted $L^2$-distance

$$T_n = n \int \left| \hat{\varphi}_{\varepsilon, \bar{X}_k}(t_0, t_1, \ldots, t_k) - \hat{\varphi}_{\varepsilon}(t_0)\hat{\varphi}_{\bar{X}_k}(t_1, \ldots, t_k) \right|^2 W(t_0, \ldots, t_k) d(t_0, \ldots, t_k).$$

Here, $W$ denotes some weight function fulfilling assumption (A8) in Section 2.2. Furthermore,

$$\hat{\varphi}_{\varepsilon, \bar{X}_k}(t_0, t_1, \ldots, t_k) = \sum_{j=1}^n \hat{w}_j \exp \left( it_0 \hat{\epsilon}_j + \sum_{\nu=1}^k t_\nu X_{j-\nu} \right)$$

estimates the joint characteristic function of $\varepsilon_j$ and $\bar{X}_{k,j} = (X_{j-1}, \ldots, X_{j-k})$, whereas

$$\hat{\varphi}_{\varepsilon}(t) = \sum_{j=1}^n \hat{w}_j \exp \left( it\hat{\epsilon}_j \right),$$

$$\hat{\varphi}_{\bar{X}_k}(t_1, \ldots, t_k) = \sum_{j=1}^n \hat{w}_j \exp \left( i \sum_{\nu=1}^k t_\nu X_{j-\nu} \right)$$

estimate the marginal characteristic functions of $\varepsilon_j$ and $\bar{X}_{k,j}$, respectively. Here, the weights are defined as $\hat{w}_j = w_n(X_{j-1})/(\sum_{l=1}^n w_n(X_{l-1}))$, where we choose a weight function $w_n(x) = I_{[-a_n, a_n]}(x)$ for some sequence $a_n \to \infty$. Here and throughout, $I_A$ denotes the indicator function of set $A$. Other weight functions $w_n : \mathbb{R} \to [0, 1]$, which vanish outside $[\pm a_n, a_n]$, are possible as well but require slightly adapted assumptions. The weights are included in the definition of the empirical characteristic functions to avoid problems of kernel estimation in areas where only few data are available. Furthermore, the residuals are defined as

$$\hat{\epsilon}_j = \frac{X_j - \hat{m}(X_{j-1})}{\hat{\sigma}(X_{j-1})},$$

(2)
and we use Nadaraya–Watson-type estimators for the conditional mean and variance functions,

\[
\hat{m}(x) = \frac{1}{nc_n} \sum_{j=1}^{n} K \left( \frac{x - X_{j-1}}{c_n} \right) X_j
\]

\[
\hat{\sigma}^2(x) = \frac{1}{nc_n} \sum_{j=1}^{n} K \left( \frac{x - X_{j-1}}{c_n} \right) (X_j - \hat{m}(x))^2
\]

with kernel function \( K \) and sequence of bandwidths \( c_n, \ n \in \mathbb{N} \). Here,

\[
\hat{f}_X(x) = \frac{1}{nc_n} \sum_{j=1}^{n} K \left( \frac{x - X_{j-1}}{c_n} \right)
\]

denotes a kernel estimator for the marginal density \( f_X \) of \( X_j \). See, for example, Robinson (1983), Masry and Tjøstheim (1995), Härdle and Tsybakov (1997), and Hansen (2008) for properties of these estimators in the time series context.

### 2.2 Assumptions and asymptotic results under the null hypothesis

Under the null hypothesis, we state the following assumptions. Please note that, throughout, we write \( t = (t_0, t_1, \ldots, t_k) \) and use the notation \( g(t) \) for simplicity also for functions \( g \) that only depend on \( (t_1, \ldots, t_k) \) or \( (t_2, \ldots, t_k) \) (see, e. g., \( \psi(t,x) \) from assumption (A4)).

(A1) The process \( (X_t)_{t \in \mathbb{Z}} \) is strictly stationary and \( \alpha \)-mixing with mixing coefficient \( \alpha \) that satisfies \( \alpha(i) \leq Ai^{-\beta} \) for some \( A < \infty \) and \( \beta > (1 + (s - 1)(2 + 1/q))/(s - 2) \) for some \( q > 0 \), where \( s > 2 \) and \( E|X_0|^s < \infty \). \( X_1 \) has bounded marginal density \( f_X \) such that for some constant \( B_1 \),

\[
\sup_x E(|X_1|^s \mid X_0 = x) f_X(x) \leq B_1.
\]

Furthermore \((X_0, X_j)\) has bounded joint density \( f_j \) and there exists a constant \( B_2 \), such that for some \( j^* \),

\[
\sup_{x_0, x_j} E(|X_1 X_{j+1}| \mid X_0 = 0, X_j = x_j) f_j(x_0, x_j) \leq B_2,
\]

for all \( j \geq j^* \).

(A2) Let \( m, \sigma^2 \), and \( f_X \) be differentiable. Let the functions \( m, m', \sigma^2, (\sigma^2)' \), \( \frac{1}{m'} \), \( \frac{1}{f_X} \), and \( f_X' \) be of order \( O((\log n)^r) \), for some \( r > 0 \), uniformly on the interval \( I_n = [-a_n - Cc_n, a_n + Cc_n] \). (Here, the sequence \( a_n \) corresponding to the weight function is further specified in (A6), the sequence of bandwidths \( c_n \) fulfills assumption (A7) and \( C \) is from assumption (A5)). Furthermore, we assume Lipschitz continuity of the derivatives \( f_X', m', \) and \( (\sigma^2)' \) in the following sense:

\[
\sup_{x, y \in I_n \atop |x - y| \leq a_n} |g(x) - g(y)| = O \left( c_n (\log n)^r \right) \text{ for } g \in \{ f_X', m', (\sigma^2)' \}.
\]

(A3) The innovations \((\epsilon_t)_{t \in \mathbb{Z}}\) are independent, centered, and identically distributed. For each \( t \in \mathbb{Z} \), \( \epsilon_t \) is independent from the past \( X_{t-1}, X_{t-2}, \ldots \). For some \( \delta > 2/(\beta - 2) \), let \( E|\epsilon_1|^{2(2+2\delta)/\beta} < \infty \) and \( \sup_{x \in I_n} E \left[ |\epsilon_j|^{2(1+\delta)} | X_0 = x \right] = O((\log n)^r) \) uniformly in \( j \) with \( r \) and \( I_n \) from assumption (A2).
Define $\psi(t, x) = E[Y_1(t) \mid X_0 = x] - E[Y_1(t)]$ with $Y_1(t) = \cos(\sum_{k=1}^{\infty} t_k X_{1-k})$ and assume that

$$\sup_{x \in A_2} \sup_{|z| \leq C_n} |\psi(t, x) - \psi(t, z)| = O\left( (\log n)^d c_n^d \right)$$

for some $d > 0$ with $r$ from assumption (A2) and $C$ from assumption (A5). Assume the same condition holds for $\tilde{\psi}(t, x) = E[Z_1(t) \mid X_0 = x] - E[Z_1(t)]$ with $Z_1(t) = \sin(\sum_{k=1}^{\infty} t_k X_{1-k})$.

The kernel $K$ is a symmetric and Lipschitz continuous density with compact support $[-C, C]$ and $\int K(u)du = 0$.

For $q, s$, and $\beta$ from (A1), we have $a_n = O(n^{1/(2q)}) \log n$, and for

$$\theta = \frac{\beta - 2 - \frac{1}{q} - \frac{1+\beta}{s-1}}{\beta + 2 - \frac{1+\beta}{s-1}},$$

it holds that $\log n = o(n^{\theta} c_n)$. Let

$$a_n^* = \left( \frac{\log n}{n c_n} \right)^{1/2} + c_n^2,$$

then $a_n^* = O(\Delta n n^{-1/4})$ with $\Delta_n = \inf_{|x| \leq a_n} f_X(x)$.

Let the sequence of bandwidths fulfill $n c_n^2 (\log n)^{-D} \to \infty$, $n c_n^D (\log n)^D \to 0$ for all $D > 0$.

The weight function $W$ is nonnegative and symmetric such that $W(\pm t_0, \pm t_1, \ldots, \pm t_k) = W(t_0, \ldots, t_k)$. Furthermore, $\int t_0^k W(t_0, \ldots, t_k)dt_0 \ldots dt_k < \infty$.

Remark 1. Apart from the typical assumptions on the kernel, bandwidths, and weight functions, we need smoothness assumptions on the unknown functions as well as moment assumptions and the mixing property; for example, in order to obtain uniform rates of convergence for the kernel estimators, similar to Hansen (2008). Note that for (A6) and (A7) to both be satisfied, one needs $\theta > 1/4$.

We have the following asymptotic distribution of the test statistic under the null.

**Theorem 1.** Under model (1) with assumptions (A1)–(A8), the test statistic $T_n$ converges in distribution to $T_0 = \int_{\mathbb{R}^{k+1}} S^2(t) W(t) dt$, where $S(t)$, $t \in \mathbb{R}^{k+1}$, denotes a centered Gaussian process with the same covariance structure as

$$\tilde{S}(t_0, \ldots, t_k) = (\cos(t_0e_1) - E[\cos(t_0e_1)]) (Y_1(t) + Z_1(t) - E[Y_1(t) + Z_1(t)])$$

$$+ (\sin(t_0e_1) - E[\sin(t_0e_1)]) (Y_1(t) - Z_1(t) - E[Y_1(t) - Z_1(t)])$$

$$+ t_0 \left( \epsilon_1 E[\sin(t_0e_1)] + \frac{1}{2} (\epsilon_1^2 - 1) E[\sin(t_0e_1)e_1] \right) (E[Y_1(t) + Z_1(t) \mid X_0] - E[Y_1(t) + Z_1(t)])$$

$$- t_0 \left( \epsilon_1 E[\cos(t_0e_1)] + \frac{1}{2} (\epsilon_1^2 - 1) E[\cos(t_0e_1)e_1] \right) (E[Y_1(t) - Z_1(t) \mid X_0] - E[Y_1(t) - Z_1(t)]),$$

where $Y_1(t)$ and $Z_1(t)$ have been defined in assumption (A4).

The proof is given in the appendix A. An asymptotic level-$\alpha$ test is obtained by rejecting $H_0$ whenever $T_n > c_{1-\alpha}$, where $P(T > c_{1-\alpha}) = \alpha$. Because of the complicated distribution of $T$, we suggest a bootstrap procedure to estimate the critical value $c_{1-\alpha}$ in Section 4.
Remark 2.

(a) The replacement of true but unknown innovations $\varepsilon_j$ by the estimated residuals $\hat{\varepsilon}_j$ changes the asymptotic distribution drastically. Were the true innovations known and used in the test statistic instead of residuals, the statistic $\tilde{S}$ in Theorem 1 would simplify to

$$
\tilde{S}(t_0, \ldots, t_k) = (\cos(t_0\varepsilon_1) - E[\cos(t_0\varepsilon_1)]) (Y_1(t) + Z_1(t) - E[Y_1(t) + Z_1(t)])
+ (\sin(t_0\varepsilon_1) - E[\sin(t_0\varepsilon_1)]) (Y_1(t) - Z_1(t) - E[Y_1(t) - Z_1(t)]).
$$

The same phenomenon has been observed in other contexts, see, for example, Akritas and Van Keilegom (2001), Müller et al. (2009), Hlávka et al. (2012), and Selk and Neumeyer (2013).

(b) If the aim is to test for independence of innovations and past time series values in a (homoscedastic) AR(1) model

$$
X_j = m(X_{j-1}) + \varepsilon_j,
$$

one simply sets $\hat{\sigma} \equiv 1$ in the definition of the residuals. Then, the statistic $\tilde{S}$ in Theorem 1 changes to

$$
\tilde{S}(t_0, \ldots, t_k) = (\cos(t_0\varepsilon_1) - E[\cos(t_0\varepsilon_1)]) (Y_1(t) + Z_1(t) - E[Y_1(t) + Z_1(t)])
+ (\sin(t_0\varepsilon_1) - E[\sin(t_0\varepsilon_1)]) (Y_1(t) - Z_1(t) - E[Y_1(t) - Z_1(t)])
+ t_0\varepsilon_1 E[\sin(t_0\varepsilon_1)] E[Y_1(t) + Z_1(t) | X_0] - E[Y_1(t) + Z_1(t)]
- t_0\varepsilon_1 E[\cos(t_0\varepsilon_1)] E[Y_1(t) - Z_1(t) | X_0] - E[Y_1(t) - Z_1(t)].
$$

(c) Assumption (A8) is satisfied for a number of choices of weight functions $W$. The question of choice of $W$ that leads to asymptotically locally optimal tests is an open and challenging question, and we are not going to elaborate it here. Carrasco et al. (2007) and Carrasco and Florens (2014) derive optimality in a related context but consider the estimation of finite-dimensional parameters.

For practitioners, the computation of the test statistic $T_n$ might be challenging. For that reason, we suppose using another representation of $T_n$, which avoids for solving complicated integrals. The alternative representation is stated in the following lemma.

**Lemma 1.** Let $F[V](x)$ denote the Fourier transformation of $V$ at point $x$. Under Assumption (A8) and if $W$ yields $W(t_0, \ldots, t_k) = V_0(t_0) \prod_{i=1}^k V_i(t_i)$, it holds $F[V](x) = \int \cos(tx)V(t)dt$, and the test statistic $T_n$ can be represented by

$$
T_n = n \sum_{s_1,s_2=1}^{n} \tilde{w}_{s_1} \tilde{w}_{s_2} F[V_0] ((\hat{\varepsilon}_{s_1} - \hat{\varepsilon}_{s_2}) \sum_{s_3,s_4=1}^{n} \tilde{w}_{s_3} \tilde{w}_{s_4} \prod_{j=1}^{k} F[V_j] (X_{s_3-j} - X_{s_4-j})
+ n \sum_{s_1,s_2=1}^{n} \tilde{w}_{s_1} \tilde{w}_{s_2} F[V_0] ((\hat{\varepsilon}_{s_1} - \hat{\varepsilon}_{s_2}) \prod_{j=1}^{k} F[V_j] (X_{s_3-j} - X_{s_4-j})
- 2n \sum_{s_1,s_2,s_3=1}^{n} \tilde{w}_{s_1} \tilde{w}_{s_2} \tilde{w}_{s_3} F[V_0] ((\hat{\varepsilon}_{s_1} - \hat{\varepsilon}_{s_2}) \prod_{j=1}^{k} F[V_j] (X_{s_3-j} - X_{s_4-j})).
$$

Since the choice of the weighting function $W$ belongs to the user, the additional assumption on its multiplicative form is very weak. If one further chooses $W$ such that the Fourier transformations of the corresponding functions $V_i$, $i = 0, \ldots, k$, are known, the test statistic $T_n$ can straightforwardly
be computed. Even more important, the implementation then simplifies a lot since the computation of the \((k + 1)\)-fold integral is omitted.

**Example 1.** Some choices of \(W\) fulfilling the assumptions of the lemma are

1. \(W(t_0, \ldots, t_k) = \sum_{j=1}^{k} \prod_{j=1}^{k} e^{-\gamma_j |t_j|^2}, \) where the Fourier transformation of \(V_j(t_j) := e^{-\gamma_j |t_j|^2}, \) \(j = 0, \ldots, k,\) is given by \(\hat{F}[V_j](\omega) = 2\gamma_j / (\gamma_j^2 + 4\pi^2 x^2).\)

2. \(W(t_0, \ldots, t_k) = \sum_{j=1}^{k} \prod_{j=1}^{k} e^{-\gamma_j |t_j|^2}, \) where the Fourier transformation of \(V_j(t_j) := e^{-\gamma_j t_j^2}, \) \(j = 0, \ldots, k,\) is given by \(\hat{F}[V_j](\omega) = (\pi / \gamma_j)^{1/2} e^{-(\pi x^2) / \gamma_j} \).

### 2.3 Fixed alternatives

Note that by construction the test statistic \(T_n\) cannot detect alternatives where the innovation \(\epsilon_j\) is independent of \((X_{j-1}, \ldots, X_{j-k})\), but depends on some \(X_{j-l}\) for \(l > k\). However, the test is consistent against any fixed alternative

\[ H_1 : \epsilon_j \text{ and } X_{j-l} \text{ are stochastically dependent for some } l \in \{1, \ldots, k\} \]

under the following model. Assume that \((X_j)_{j \in \mathbb{Z}}\) is a strictly stationary and weakly dependent time series that fulfills assumption (A1). Further define \(m(x) = E[X_{j+1} | X_j = x]\) and \(\sigma^2(x) = \text{Var}(X_{j+1} | X_j = x)\). Let \(m, \sigma^2\) and the marginal density \(f_X\) fulfill assumption (A2). Let the kernel, weight function and sequence of bandwidths fulfill (A5)–(A8). Then, we have the following result.

**Theorem 2.** Under the assumptions listed in this section, \(T_n / n\) converges to

\[ \bar{T} = \int \left| \varphi_{\epsilon,X_k}(t_0, t_1, \ldots, t_k) - \varphi_\epsilon(t_0) \varphi_{X_k}(t_1, \ldots, t_k) \right|^2 W(t_0, \ldots, t_k) \, dt_0 \cdots dt_k \]

in probability, where \(\varphi_{\epsilon,X_k}\) is the joint characteristic function of \(\epsilon_j\) and \((X_{j-1}, \ldots, X_{j-k})\) and \(\varphi_\epsilon\) and \(\varphi_{X_k}\) are the corresponding marginal characteristic functions.

The proof is given in the appendix. Note that, under \(H_1\), one has \(\bar{T} > 0\), and hence, \(P(T_n > c_{1-\alpha}) \to 1\) for \(n \to \infty\).

From rejection of \(H_0\), one should conclude that the AR(1)-ARCH(1) model is not suitable to describe the data. Possible reasons are explained in the following example.

**Example 2.**

(a) Consider the conditional distribution of \(\epsilon_j\), given \(X_{j-1}\). The first two moments of this distribution do not depend on \(X_{j-1}\) by construction. Higher order moments could depend on \(X_{j-1}\), that is, \(E[\epsilon_j^\ell | X_{j-1}] = h_\ell(X_{j-1})\) for some \(\ell \geq 3\). In the simulation study, we will consider a skew normal innovation distribution with mean zero, variance one, and skewness dependent on \(X_{j-1}\).

(b) The conditional distribution of \(\epsilon_j\), given \(X_k = (X_{j-1}, \ldots, X_{j-k})\), may still depend on \(X_k\). If this distribution does still depend on the first component \(X_{j-1}\) but only on this component, modeling the autoregression and conditional variance function with lag 1 is appropriate, but one should not apply any procedures that assume the independence of innovations and past time series values.

(c) An AR(\(\ell\))-ARCH(\(\ell\)) model could be appropriate for the data for some \(\ell > 1\), that is,

\[ X_j = \tilde{m}(X_{j-1}, \ldots, X_{j-\ell}) + \tilde{\sigma}(X_{j-1}, \ldots, X_{j-\ell}) \eta_j \]

with innovations \(\eta_j\) independent from \(X_{j-1}, X_{j-2}, \ldots\).
3 | A CRAMÉR–VON MISÉS-TYPE TEST

Alternative testing procedures can be given by, for example, Kolmogorov–Smirnov or Cramér–von Misés-type statistics based on the \( \hat{F}_{\hat{X}_k} - \hat{F}_X \otimes \hat{F}_{\hat{X}_k} \), that is, the difference between the weighted empirical joint distribution function of residuals \( \hat{e}_j \) (defined in (2) and \( \hat{X}_{k,j} = (X_{j-1}, \ldots, X_{j-k}) \) \( j = 1, \ldots, n \)) and the product of the marginals. We concentrate here on the Cramér–von Misés-type statistic for two reasons. First, in our simulation study, the Cramér–von Misés-type statistics showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the

statistic how showed better power than the Kolmogorov–Smirnov test in all models under consideration. Second, whereas methods of proof for the Cramér–von Misés statistic are similar to the
is valid for all $j > j^* + 1$, for $k = 1, 2, n \to \infty$ with

$$J_n = \left[ -a_n - \left( C + c_n^{-\frac{1}{3}} n^{-\frac{1}{3}} \log n \right) c_n, a_n + \left( C + c_n^{-\frac{1}{3}} n^{-\frac{1}{3}} \log n \right) c_n \right].$$

(A14) The regression function $m$ and the scale function $\sigma$ are four times differentiable, and there exist some $r > 0$ and sequence $q_n = O((\log n)^r)$ with $\sup_{x \in [a_n-c_n-b_n+c_n]} |m^{(r)}(x)| = O(q_n)$; $\sup_{x \in [a_n-c_n-b_n+c_n]} |\sigma^{(4)}(x)| = O(q_n)$; $\mu = 0, 1, 2, 3, 4$, $(\inf_{x \in [a_n-c_n-b_n+c_n]} |\sigma(x)|)^{-1} = O(q_n)$; and $(q_n)^{-1} = O(1)$.

(A15) The innovations $\varepsilon_j$, $j \in \mathbb{Z}$, are absolutely continuous with distribution function $F_\varepsilon$. Their density $f_\varepsilon$ is continuously differentiable and $\sup_{t \in \mathbb{R}} |f_\varepsilon(t)| < \infty$, $\sup_{t \in \mathbb{R}} |f_\varepsilon(t)^2| < \infty$. Furthermore, $E \left[ \left| \varepsilon_1 \right|^{2b} \right] < \infty$ for $b$ from assumption (A12).

(A16) Define $\phi(t, x) = P(X_{-1} \leq t_2, \ldots, X_{1-k} \leq t_k \mid X_0 = x)$ and assume that

$$\sup_{x \in \mathbb{R}} \sup_{t \in \mathbb{R}} \left| \phi(t, x) - \phi(t, z) \right| = O\left( (\log n)^r c_n^d \right)$$

for some $d > 0$ with $r$ from assumption (A2) and $C$ from assumption (A9).

(A17) For $\phi$, from assumption (A16), it is valid that $|\phi(t_1, x) - \phi(t_2, x)| \leq \|t_1 - t_2\|_2^\rho h(x)$ for some $\rho \in (0, 1]$ and some function $h$ with $E [h^2(X_1)] < \infty$.

(A18) The weight function $\tilde{W}$ is nonnegative and integrable.

Note that some of the assumptions are stronger than needed for the empirical characteristic function approach, particularly the assumptions on the innovation distribution.

**Theorem 3.** Under model (I) with assumptions (A1)–(A3), (A6), (A7), and (A9)–(A18), the test statistic $\tilde{T}_n$ converges in distribution to $\tilde{T} = \int_{\mathbb{R}^{k+1}} S^2(t) \tilde{W}(t) \, dt$, where $S(t), t \in \mathbb{R}^{k+1}$, denotes a centered Gaussian process with the same covariance structure as

$$\tilde{S}(t_0, \ldots, t_k) = (I \{ \varepsilon_1 \leq t_0 \} - F_\varepsilon(t_0)) \left( \tilde{Y}_1(t) - E \left[ \tilde{Y}_1(t) \right] \right) + \left( f_\varepsilon(t_0) \varepsilon_1 + \frac{1}{2} t_0 f_\varepsilon(t_0) (\varepsilon^2_1 - 1) \right) \left( E \left[ \tilde{Y}_1(t) \mid X_0 \right] - E \left[ \tilde{Y}_1(t) \right] \right),$$

where $\tilde{Y}_1(t) = I \{ X_0 \leq t_1, \ldots, X_{1-k} \leq t_k \}, E \left[ \tilde{Y}_1(t) \right] = F_{X_0, \ldots, X_{1-k}}(t)$ and $E \left[ \tilde{Y}_1(t) \mid X_0 = x \right] = I \{ x \leq t_1 \} P(X_{-1} \leq t_2, \ldots, X_{1-k} \leq t_k \mid X_0 = x)$.

The proof is given in the appendix. As in Remark 2 (a), one sees that the asymptotic distribution is influenced by the estimation of $m$ and $\sigma^2$. For known errors $\varepsilon_j$ the second line in the definition of $\tilde{S}$ would vanish. Similarly as in Remark 2 (b) and Theorem 2, the homoscedastic models and fixed alternatives can be treated. Details are omitted for the sake of brevity.

## 4 Bootstrap and Finite-Sample Performance

In this section, we investigate the finite-sample performance of our tests. Because of the complicated limiting distribution of $T$ from Theorem 1 (and $\tilde{T}$ from Theorem 3), we suggest using a smooth autoregressive residual bootstrap instead.
4.1 | Bootstrap

Our bootstrap strategy is as follows. Firstly, based on the estimators as introduced in Section 2, generate bootstrap innovations $\varepsilon^*_j$ from a smooth estimate of the innovation distribution, that is, given the original data $X_{-k+1}, \ldots, X_n$, the distribution of $\varepsilon^*_j$ reads

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} L \left( \frac{x - \tilde{\varepsilon}_i}{h_n} \right),$$

where $h_n$ denotes a positive bandwidth, $L$ is some smooth distribution function, and $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ denote the standardized versions of the residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$ defined in (2). Secondly, compute the bootstrap process via

$$X^*_j = \hat{m} (X^*_{j-1}) + \hat{\sigma} (X^*_{j-1}) \varepsilon^*_j, \quad j = 1, \ldots, n,$$

with some starting value $X^*_0$ and a sufficiently large number of forerunnings to ensure that the process is in balance. Thirdly, calculate the bootstrap analogue of the test statistic $T_n$, say $T^*_n$. Frequent repetitions of these steps give the distribution of $T^*_n$, which approximates the distribution of $T_n$ under $H_0$. By using the $(1 - \alpha)$-percentile of the distribution of $T^*_n$, say $c^*_{1-\alpha}$, the hypothesis of independence then is rejected if $T_n > c^*_{1-\alpha}$ (Analogous considerations hold for $\tilde{T}_n$).

To prove validity of the bootstrap procedure, one would mimic the proof of Theorem 1 (and Theorem 3 for $\tilde{T}_n$) to show that the conditional distribution of $T^*_n$, given the original series $X_t, t \in \mathbb{Z}$, converges to the distribution of $T$ in Theorem 1 (under suitable regularity conditions). To this end, it is of importance that the bootstrap data fulfill the null hypothesis. This is assured by conditional independence of the bootstrap innovations $\varepsilon^*_j$ from $X^*_{j-1}, X^*_{j-2}, \ldots$, given the original data. It is also important that the bootstrap errors $\varepsilon^*_j$ are absolutely continuous in accordance with assumption (A16). Furthermore, note that a nonsmooth residual bootstrap, that is, drawing $\varepsilon^*_j$ with replacement from (standardized) residuals, might result in a time series that is not mixing. Note that Neumeyer (2009) and Birke, Neumeyer, and Volgushev (2017) proved the validity of smooth residual bootstrap procedures in the context of residual processes for regression models with independent data. Similar methods could be applied in our context, but a rigorous proof is beyond the scope of the paper.

4.2 | Simulations

We investigated the finite-sample performance of our test statistics in simulations for various models. All results shown are based on 400 simulation runs with $B = 200$ bootstrap replications for each data set. Sample sizes are $n = 100$ and $n = 200$ (and $n = 1000$ in particular cases). For the smooth residual bootstrap procedure, we chose $L$ as the standard normal distribution, and $h_n$ was set to $n^{-1/4}$ for reasons given in Neumeyer (2009). For the estimation of the mean and variance function, the bandwidth $c_n$ was chosen by Silverman’s rule by thumb, see Silverman (1986), and we used a Gaussian kernel. We set $w_n \equiv 1$ for simplicity. We used R (R Core Team, 2013) for the simulations.

We considered the test statistic $T_n$ on the basis of characteristic functions with weight function as in example 1 (1). Here, $\gamma_0$ and $\gamma_j (j = 1, \ldots, k)$ are $\gamma \in \{2, 4, 10, 15, 20\}$ multiplied with the standard deviation of $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$ and of $X_1, \ldots, X_n$, respectively. Those tests are denoted $TW1(\gamma)$ in the tables. Furthermore, we considered $T_n$ with weight function as in example 1 (2) and $\gamma_j$ as for $TW1(\gamma)$. Those tests are denoted $TW2(\gamma)$ in the tables. We further show results for the Cramér–von Mises test statistic, denoted CM in the tables. Results for the Kolmogorov–Smirnov test are not
shown as this test throughout showed less power than the Cramér–von Mises test. Note that the testing procedure is always based on estimating the general AR-ARCH model (1), even when we generate data from AR or ARCH models.

We first investigate the performance of the tests for the case \( k = 1 \) in the settings A, B, and A2 described below. The data are generated as \( X_t = m(X_{t-1}) + \sigma(X_{t-1})\epsilon_t \). Denote \( m_A(x) = 0.9x \) and \( s_A(x) = 1 + 0.25x^2 \). Then, setting A consists of the following models:

1. AR-model with \( m = m_A, \sigma^2 \equiv 1, \epsilon_t \) standard normal
2. ARCH-model with \( m \equiv 0, \sigma^2 = s_A, \epsilon_t \) standard normal
3. AR-ARCH-model with \( m = m_A, \sigma^2 \equiv s_A, \epsilon_t \) standard normal
4. AR-model with \( m = m_A, \sigma^2 \equiv 1, \epsilon_t \) skew normal with skewness parameter \( 5X_{t-1}^2 \)
5. AR-model with \( m = m_A, \sigma^2 \equiv 1, \epsilon_t \) skew normal with skewness parameter \( 10X_{t-1}^2 \)
6. AR-model with \( m = m_A, \sigma^2 \equiv 1, \epsilon_t \) skew normal with skewness parameter \( 15X_{t-1}^2 \)
7. ARCH-model with \( m \equiv 0, \sigma^2 = s_A, \epsilon_t \) skew normal with skewness parameter \( 5X_{t-1}^2 \)
8. ARCH-model with \( m \equiv 0, \sigma^2 = s_A, \epsilon_t \) skew normal with skewness parameter \( 10X_{t-1}^2 \)
9. ARCH-model with \( m \equiv 0, \sigma^2 = s_A, \epsilon_t \) skew normal with skewness parameter \( 15X_{t-1}^2 \)
10. AR-ARCH-model with \( m = m_A, \sigma^2 \equiv s_A, \epsilon_t \) skew normal with skewness parameter \( 5X_{t-1}^2 \)
11. AR-ARCH-model with \( m = m_A, \sigma^2 \equiv s_A, \epsilon_t \) skew normal with skewness parameter \( 10X_{t-1}^2 \)
12. AR-ARCH-model with \( m = m_A, \sigma^2 \equiv s_A, \epsilon_t \) skew normal with skewness parameter \( 15X_{t-1}^2 \).

Here, cases 1–3 correspond to the null hypothesis of independence, whereas cases 4–12 are fixed alternatives. For the skew normal distribution, we use the notation of Fernández and Steel (1998). Setting B is completely analogous to setting A but with \( m_A \) replaced by \( m_B(x) = \sin(1 + 0.5x) \) and \( s_A \) replaced by \( s_B(x) = \cos^2(1 + 0.5x) \).

In all the simulations, the nominal level is 0.05. Tables 1, 2, 3, 4, 5, and 6 show the rejection probabilities for settings A and B for sample sizes \( n = 100 \) and \( n = 200 \). As expected, the power increases with increasing factor of the skewness parameter and with increasing sample size. The level is well kept. The characteristic-function-based tests are very conservative in the AR-ARCH model in setting A (case A3). All tests can well detect the alternatives generated from AR and ARCH models (settings A 4–9, B 4–9), whereas the power is lower for the data generated from AR-ARCH models (settings A 10–12, B 10–12). Moreover, TW2 has higher power than TW1. For \( n = 200 \), in setting A, TW1 shows the best power for \( \gamma = 10 \) but is still outperformed by CM.

| Setting A | 1  | 2  | 3  | Setting B | 1  | 2  | 3  |
|-----------|----|----|----|-----------|----|----|----|
| CM        | 0.035 | 0.040 | 0.030 | CM        | 0.062 | 0.067 | 0.060 |
| TW1(2)    | 0.007 | 0.042 | 0.002 | TW1(2)    | 0.055 | 0.047 | 0.037 |
| TW1(4)    | 0.015 | 0.055 | 0.002 | TW1(4)    | 0.052 | 0.052 | 0.055 |
| TW1(10)   | 0.020 | 0.045 | 0.000 | TW1(10)   | 0.037 | 0.055 | 0.107 |
| TW1(15)   | 0.022 | 0.035 | 0.000 | TW1(15)   | 0.032 | 0.057 | 0.125 |
| TW1(20)   | 0.012 | 0.030 | 0.000 | TW1(20)   | 0.035 | 0.052 | 0.130 |
| TW2(2)    | 0.022 | 0.060 | 0.010 | TW2(2)    | 0.065 | 0.045 | 0.050 |
| TW2(4)    | 0.040 | 0.057 | 0.007 | TW2(4)    | 0.070 | 0.045 | 0.050 |
| TW2(10)   | 0.035 | 0.062 | 0.007 | TW2(10)   | 0.052 | 0.045 | 0.087 |
| TW2(15)   | 0.030 | 0.045 | 0.007 | TW2(15)   | 0.040 | 0.045 | 0.085 |
| TW2(20)   | 0.027 | 0.040 | 0.010 | TW2(20)   | 0.030 | 0.050 | 0.090 |
TABLE 2  Power simulations for \( n = 100 \) for setting A 4–12

| Setting A | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
|-----------|------|------|------|------|------|------|------|------|------|
| CM        | 0.120| 0.150| 0.160| 0.447| 0.475| 0.477| 0.162| 0.170| 0.150|
| TW1(2)    | 0.042| 0.030| 0.045| 0.390| 0.442| 0.517| 0.032| 0.042| 0.022|
| TW1(4)    | 0.110| 0.117| 0.167| 0.580| 0.605| 0.662| 0.080| 0.090| 0.072|
| TW1(10)   | 0.117| 0.112| 0.155| 0.675| 0.675| 0.722| 0.062| 0.082| 0.065|
| TW1(15)   | 0.075| 0.072| 0.100| 0.645| 0.642| 0.700| 0.042| 0.055| 0.037|
| TW1(20)   | 0.032| 0.020| 0.030| 0.565| 0.572| 0.602| 0.012| 0.025| 0.022|
| TW2(2)    | 0.085| 0.097| 0.147| 0.417| 0.517| 0.560| 0.080| 0.110| 0.082|
| TW2(4)    | 0.145| 0.175| 0.252| 0.550| 0.602| 0.652| 0.130| 0.152| 0.130|
| TW2(10)   | 0.232| 0.247| 0.322| 0.677| 0.705| 0.730| 0.167| 0.195| 0.180|
| TW2(15)   | 0.230| 0.257| 0.322| 0.690| 0.717| 0.745| 0.155| 0.177| 0.170|
| TW2(20)   | 0.212| 0.220| 0.290| 0.702| 0.710| 0.735| 0.140| 0.155| 0.145|

TABLE 3  Power simulations for \( n = 100 \) for setting B 4–12

| Setting B | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
|-----------|------|------|------|------|------|------|------|------|------|
| CM        | 0.140| 0.175| 0.155| 0.182| 0.367| 0.467| 0.125| 0.152| 0.150|
| TW1(2)    | 0.107| 0.137| 0.132| 0.167| 0.325| 0.407| 0.105| 0.120| 0.122|
| TW1(4)    | 0.140| 0.172| 0.142| 0.205| 0.430| 0.557| 0.115| 0.122| 0.105|
| TW1(10)   | 0.122| 0.172| 0.157| 0.260| 0.485| 0.632| 0.187| 0.192| 0.172|
| TW1(15)   | 0.120| 0.165| 0.137| 0.245| 0.470| 0.620| 0.245| 0.252| 0.242|
| TW1(20)   | 0.092| 0.132| 0.110| 0.225| 0.442| 0.567| 0.250| 0.287| 0.277|
| TW2(2)    | 0.117| 0.150| 0.135| 0.165| 0.372| 0.455| 0.067| 0.082| 0.067|
| TW2(4)    | 0.145| 0.180| 0.177| 0.195| 0.422| 0.547| 0.085| 0.107| 0.090|
| TW2(10)   | 0.197| 0.210| 0.227| 0.232| 0.457| 0.617| 0.130| 0.160| 0.142|
| TW2(15)   | 0.197| 0.230| 0.240| 0.257| 0.472| 0.602| 0.257| 0.200| 0.157|
| TW2(20)   | 0.192| 0.225| 0.235| 0.255| 0.460| 0.605| 0.165| 0.222| 0.200|

TABLE 4  Size simulations for \( n = 200 \) for settings A 1–3 and B 1–3

| Setting A | 1    | 2    | 3    | Setting B | 1    | 2    | 3    |
|-----------|------|------|------|-----------|------|------|------|
| CM        | 0.025| 0.050| 0.040| CM        | 0.050| 0.082| 0.060|
| TW1(2)    | 0.017| 0.042| 0.002| TW1(2)    | 0.060| 0.060| 0.045|
| TW1(4)    | 0.025| 0.047| 0.002| TW1(4)    | 0.057| 0.057| 0.047|
| TW1(10)   | 0.030| 0.037| 0.000| TW1(10)   | 0.045| 0.082| 0.067|
| TW1(15)   | 0.030| 0.035| 0.000| TW1(15)   | 0.047| 0.095| 0.127|
| TW1(20)   | 0.022| 0.035| 0.000| TW1(20)   | 0.045| 0.095| 0.150|
| TW2(2)    | 0.035| 0.047| 0.012| TW2(2)    | 0.065| 0.047| 0.045|
| TW2(4)    | 0.032| 0.035| 0.010| TW2(4)    | 0.057| 0.056| 0.048|
| TW2(10)   | 0.027| 0.042| 0.005| TW2(10)   | 0.062| 0.065| 0.057|
| TW2(15)   | 0.027| 0.037| 0.005| TW2(15)   | 0.057| 0.080| 0.057|
| TW2(20)   | 0.032| 0.035| 0.000| TW2(20)   | 0.050| 0.075| 0.060|

and TW2(15). Overall, the Cramér–von Mises test performs very well. However, in setting B for \( n = 200 \), it is even outperformed by TW2(20). As the power is rather low for the AR-ARCH cases 10–12 in setting B, we present exemplarily in Table 10 the rejection probabilities of TW1(4) and TW2(4) for sample size \( n = 1000 \) to demonstrate consistency of the tests.
To investigate how sensitive the procedures are with respect to heavier tails, we consider $t$-distributed errors in setting A2 as follows:

1. AR-model with $m = m_A$, $\sigma^2 = 1$, $\epsilon_t \sim t$-distributed with 3 degrees of freedom
2. ARCH-model with $m = 0$, $\sigma^2 = s_A$, $\epsilon_t \sim t$-distributed with 3 degrees of freedom
3. AR-ARCH-model with $m = m_A$, $\sigma^2 = s_A$, $\epsilon_t \sim t$-distributed with 3 degrees of freedom

### TABLE 5  Power simulations for $n = 200$ for setting A 4–12

| Setting A | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
|-----------|------|------|------|------|------|------|------|------|------|
| CM        | 0.580| 0.625| 0.587| 0.887| 0.922| 0.917| 0.502| 0.572| 0.617|
| TW1(2)    | 0.317| 0.362| 0.372| 0.812| 0.875| 0.907| 0.145| 0.215| 0.230|
| TW1(4)    | 0.605| 0.645| 0.635| 0.935| 0.965| 0.967| 0.250| 0.330| 0.370|
| TW1(10)   | 0.687| 0.687| 0.662| 0.980| 0.985| 0.982| 0.222| 0.277| 0.310|
| TW1(15)   | 0.526| 0.580| 0.595| 0.975| 0.987| 0.982| 0.172| 0.222| 0.227|
| TW1(20)   | 0.440| 0.435| 0.405| 0.960| 0.977| 0.970| 0.112| 0.147| 0.130|

### TABLE 6  Power simulations for $n = 200$ for setting B 4–12

| Setting B | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
|-----------|------|------|------|------|------|------|------|------|------|
| CM        | 0.420| 0.522| 0.512| 0.465| 0.800| 0.880| 0.190| 0.202| 0.222|
| TW1(2)    | 0.292| 0.420| 0.442| 0.350| 0.682| 0.802| 0.135| 0.142| 0.125|
| TW1(4)    | 0.420| 0.545| 0.525| 0.490| 0.820| 0.912| 0.145| 0.147| 0.170|
| TW1(10)   | 0.477| 0.597| 0.550| 0.607| 0.912| 0.955| 0.250| 0.262| 0.267|
| TW1(15)   | 0.432| 0.542| 0.500| 0.617| 0.905| 0.945| 0.362| 0.352| 0.380|
| TW1(20)   | 0.372| 0.505| 0.430| 0.595| 0.897| 0.927| 0.437| 0.422| 0.445|
| TW2(2)    | 0.310| 0.422| 0.437| 0.327| 0.717| 0.812| 0.097| 0.115| 0.127|
| TW2(4)    | 0.422| 0.535| 0.515| 0.430| 0.810| 0.892| 0.145| 0.135| 0.147|
| TW2(10)   | 0.507| 0.642| 0.622| 0.562| 0.875| 0.935| 0.195| 0.212| 0.205|
| TW2(15)   | 0.545| 0.677| 0.645| 0.592| 0.902| 0.932| 0.235| 0.257| 0.235|
| TW2(20)   | 0.560| 0.677| 0.637| 0.602| 0.900| 0.930| 0.265| 0.287| 0.275|

### TABLE 7  Size simulations for $n = 200$ for setting A2 1–6

| Setting A2 | 1    | 2    | 3    | 4    | 5    | 6    |
|------------|------|------|------|------|------|------|
| CM         | 0.057| 0.085| 0.130| 0.060| 0.065| 0.057|
| TW1(2)     | 0.012| 0.052| 0.000| 0.015| 0.042| 0.000|
| TW1(4)     | 0.030| 0.060| 0.010| 0.027| 0.040| 0.000|
| TW1(10)    | 0.030| 0.020| 0.007| 0.027| 0.022| 0.000|
| TW1(15)    | 0.032| 0.010| 0.002| 0.035| 0.015| 0.000|
| TW1(20)    | 0.027| 0.010| 0.000| 0.032| 0.007| 0.000|
| TW2(2)     | 0.027| 0.117| 0.022| 0.032| 0.090| 0.000|
| TW2(4)     | 0.045| 0.120| 0.030| 0.045| 0.082| 0.000|
| TW2(10)    | 0.082| 0.077| 0.025| 0.075| 0.067| 0.002|
| TW2(15)    | 0.077| 0.050| 0.025| 0.080| 0.047| 0.002|
| TW2(20)    | 0.082| 0.045| 0.025| 0.070| 0.040| 0.002|
4 AR-model with $m = m_A, \sigma^2 = 1, \varepsilon_t$ $t$-distributed with 5 degrees of freedom
5 ARCH-model with $m = 0, \sigma^2 = s_A, \varepsilon_t$ $t$-distributed with 5 degrees of freedom
6 AR-ARCH-model with $m = m_A, \sigma^2 = s_A, \varepsilon_t$ $t$-distributed with 5 degrees of freedom.

All models represent the null hypothesis. The results depicted in Table 7 show that, only in few cases, the rejection probability is slightly too large.
In setting C, the data are generated as $X_t = m(X_{t-1}) + \sigma(X_{t-1})\varepsilon_t$, where, under the alternative, the conditional distribution of $\varepsilon_t$ depends on $X_{t-2}$. Cases 1–3 correspond to the null hypothesis and

### Table 8: Size simulations for $n = 200$ for setting C 1–3

| Setting C | 1    | 2    | 3    |
|-----------|------|------|------|
| CM        | 0.060| 0.065| 0.030|
| TW1(2)    | 0.002| 0.020| 0.000|
| TW1(4)    | 0.005| 0.040| 0.000|
| TW1(10)   | 0.012| 0.055| 0.000|
| TW1(15)   | 0.015| 0.060| 0.000|
| TW1(20)   | 0.012| 0.055| 0.000|
| TW2(2)    | 0.002| 0.040| 0.000|
| TW2(4)    | 0.005| 0.052| 0.002|
| TW2(10)   | 0.010| 0.052| 0.000|
| TW2(15)   | 0.020| 0.062| 0.000|
| TW2(20)   | 0.025| 0.055| 0.000|

### Table 9: Power simulations for $n = 200$ for setting C 4–12

| Setting C | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
|-----------|------|------|------|------|------|------|------|------|------|
| CM        | 0.557| 0.590| 0.592| 0.322| 0.337| 0.387| 0.402| 0.467| 0.445|
| TW1(2)    | 0.007| 0.007| 0.005| 0.125| 0.180| 0.152| 0.002| 0.005| 0.005|
| TW1(4)    | 0.130| 0.172| 0.147| 0.435| 0.540| 0.602| 0.052| 0.032| 0.052|
| TW1(10)   | 0.352| 0.362| 0.447| 0.442| 0.555| 0.595| 0.107| 0.112| 0.070|
| TW1(15)   | 0.232| 0.320| 0.280| 0.307| 0.422| 0.405| 0.080| 0.057| 0.070|
| TW1(20)   | 0.147| 0.182| 0.157| 0.222| 0.272| 0.282| 0.035| 0.035| 0.042|
| TW2(2)    | 0.037| 0.057| 0.057| 0.290| 0.387| 0.392| 0.012| 0.020| 0.047|
| TW2(4)    | 0.177| 0.252| 0.245| 0.447| 0.522| 0.605| 0.082| 0.087| 0.117|
| TW2(10)   | 0.417| 0.460| 0.522| 0.575| 0.685| 0.697| 0.170| 0.247| 0.192|
| TW2(15)   | 0.470| 0.567| 0.567| 0.545| 0.660| 0.665| 0.192| 0.250| 0.277|
| TW2(20)   | 0.465| 0.567| 0.575| 0.512| 0.620| 0.622| 0.192| 0.230| 0.262|

### Table 10: Size and power simulations for $n = 1000$ in settings B 3, 10–12 and C 3, 10–12

| Setting B | 3    | 10   | 11   | 12   |
|-----------|------|------|------|------|
| TW1(4)    | 0.057| 0.530| 0.605| 0.582|
| TW2(4)    | 0.087| 0.665| 0.705| 0.715|

| Setting C | 3    | 10   | 11   | 12   |
|-----------|------|------|------|------|
| TW1(10)   | 0.000| 0.670| 0.677| 0.660|
| TW2(10)   | 0.005| 0.912| 0.927| 0.947|
FIGURE 1  Swiss Bond Index (SBI), Swiss Performance Index (SPI), and Swiss Immofund Index (SII); Left panels: Log returns $X_t$, $t = 1, \ldots, n$; Right panels: Scatter plots of $(X_{t-1}, \hat{\varepsilon}_t)$, $t = 2, \ldots, n$
4.3 Data analysis

We consider three financial time series that are part of the LPP2005 (Swiss Pension Fund from Pictet), namely the Swiss Bond Index (SBI), Swiss Performance Index (SPI), and Swiss Immofund Index (SII) from November 1, 2005 until April 11, 2007 ($n = 377$). The time series (log returns) are depicted in the left panel of Figure 1. The data are available in the R-package timeSeries (Rmetrics Core Team, Wuertz, Setz, and Chalabi 2015) and are used there to demonstrate (parametric) GARCH modeling. We model the three time series separately as nonparametric AR(1)-ARCH(1) series. In the right panel of Figure 1, we show scatter plots of $(X_{t-1}, \hat{\epsilon}_t)$, $t = 2, \ldots, n$. Although not much pattern is apparent in those plots, testing the hypothesis of independence of the innovation $\epsilon_t$ from the past value $X_{t-1}$ reveals some differences in the time series. The $p$-values (based on $B = 1000$ bootstrap replications) are shown in Table 11.

The large $p$-values obtained for SBI and SPI demonstrate that the null hypothesis of independence is widely accepted for the considered finite sequences of those time series, whereas some of the $p$-values for SII are borderline. Thus, one should rather be careful when fitting an AR(1)-ARCH(1) model to the finite sequence of the SII time series under consideration. Different models might be more appropriate.

5 CONCLUDING REMARKS AND OUTLOOK

In this paper, we suggested a test for independence of innovations and past time series observations in an AR-ARCH model, where both the conditional mean and conditional volatility...
function are modeled nonparametrically. For simplicity of presentation, we considered the AR(1)-ARCH(1) case. However, generalizations to AR(p)-ARCH(q) models are straightforward, while then local polynomial estimators for the mean and variance function should be used. Facing the curse of dimensionality also, semiparametric models might be of interest, see, for example, Yang, Härdle, and Nielsen (1999) for a model with an additive autoregression function and multiplicative volatility function. Including covariates is possible as well. Then one considers a model of type $X_j = m(T_j) + \sigma(T_j)\epsilon_j$, where the vector $T_j$ may include past observations. Testing independence of $\epsilon_j$ from $T_j, T_{j-1}, \ldots$ would be of interest here and can be conducted in an analogous manner.

A question related to the one considered in the paper at hand is whether the innovations form an iid sequence. Corresponding tests for parametric times series models have been considered by Ghoudi et al. (2001), among others. Presumably, with the methods developed in the paper at hand, such hypotheses tests for nonparametric time series models can be derived. We leave the consideration for future research.

**ACKNOWLEDGEMENTS**

The authors thank two referees and an associate editor for their constructive comments on previous versions of the manuscript. Financial support of the Deutsche Forschungsgemeinschaft (Research Unit FOR 1735 Structural Inference in Statistics: Adaptation and Efficiency) and Grantová agentura České republiky (GAČR15-09663S) is gratefully acknowledged.

**ORCID**

Marie Hušková [http://orcid.org/0000-0002-6868-1362]

Natalie Neumeyer [http://orcid.org/0000-0002-6649-1135]

**REFERENCES**

Akritas, M. G., & Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics, 28*, 549–567.

Birke, M., Neumeyer, N., & Volgushev, S. (2017). The independence process in conditional quantile location-scale models and an application to testing for monotonicity. *Statistica Sinica, 27*, 1815–1839.

Carrasco, M., Chernov, M., Florens, J.-P., & Ghysels, E. (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics, 140*, 529–573.

Carrasco, M., & Florens, J.-P. (2014). On the asymptotic efficiency of the GMM. *Econometric Theory, 30*, 372–406.

Dette, H., Pardo-Fernández, J. C., & Van Keilegom, I. (2009). Goodness-of-fit tests for multiplicative models with dependent data. *Scandinavian Journal of Statistics, 36*, 782–799.

Du, Z., & Escanciano, J. (2015). A nonparametric distribution-free test for serial independence of errors. *Econometric Reviews, 34*, 1011–1034.

Einmahl, J. H. J., & Van Keilegom, I. (2008). Specification tests in nonparametric regression. *Journal of Econometrics, 143*, 88–102.

Escanciano, J., Lobato, I. N., & Zhu, L. (2013). Automatic diagnostic checking for vector autoregressions. *Journal of Business and Economics Statistics, 31*, 426–437.

Fan, J., & Yao, Q. (2003). *Nonlinear time series: Nonparametric and parametric methods*. New York, NY: Springer Series in Statistics.

Fernández, C., & Steel, F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association, 93*, 359–371.

Gao, J. (2007). *Nonlinear time series: Semiparametric and nonparametric methods*. Boca Raton, FL: Chapman & Hall/CRC.
Ghoudi, K., Kulperger, R. J., & Rémillard, B. (2001). A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis*, 79, 191–218.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24, 726–748.

Härdle, W., & Tsymbalov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, 81, 223–242.

Hlávka, Z., Hušková, M., & Meintanis, S. G. (2011). Tests for independence in non-parametric heteroscedastic regression models. *Journal of Multivariate Analysis*, 102, 816–827.

Hlávka, Z., Hušková, M., Kirch, C., & Meintanis, S. G. (2012). Monitoring changes in the error distribution of autoregressive models based on Fourier methods. *TEST*, 21, 605–634.

Hlávka, Z., Hušková, M., Kirch, C., & Meintanis, S. G. (2017). Fourier-type tests involving martingale difference processes. *Econometric Reviews*, 36, 468–492.

Ibragimov, I. A., & Has'minskii, R. Z. (1981). *Statistical estimation: Asymptotic theory*. New York, NY: Springer.

Lange, T., Rahbek, A., & Jensen, S. T. (2011). Estimation and asymptotic inference in the AR-ARCH model. *Econometric Reviews*, 30, 129–153.

Ling, S., & McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory*, 19, 280–310.

Masry, E., & Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory*, 11, 258–289.

Meintanis, S. G. (2016). A review of testing procedures based on the empirical characteristic function. *South African Statistical Journal*, 50, 1–14.

Müller, U. U., Schick, A., & Wefelmeyer, W. (2009). Estimating the innovation distribution in nonparametric autoregression. *Probability Theory and Related Fields*, 144, 53–77.

Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scandinavian Journal of Statistics*, 36, 204–228.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rmetrics Core Team, Wuertz, D., Setz, T., & Chalabi, Y. (2015). timeSeries: Rmetrics - Financial Time Series Objects. R package version 3012.99. http://CRAN.R-project.org/package=timeSeries

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4, 185–207.

Selk, L., & Neumeyer, N. (2013). Testing for a change of the innovation distribution in nonparametric autoregression: The sequential empirical process approach. *Scandinavian Journal of Statistics*, 40, 770–788.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York, NY: Chapman & Hall/CRC.

Su, L., & Xiao, Z. (2008). Testing structural change in time-series nonparametric regression models. *Statistics and Its Interface*, 1, 347–366.

Sun, S., & Chiang, C.-Y. (1997). Limiting behavior of the perturbed empirical distribution functions evaluated at U-statistics for strongly mixing sequences of random variables. *Journal of Applied Mathematics and Stochastic Analysis*, 10, 3–20.

Yang, L., Härdle, W., & Nielsen, J. P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis*, 20, 579–604.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 52, 45–57.

---

**How to cite this article:** Hušková M, Neumeyer N, Niebuhr T, Selk L. Specification testing in nonparametric AR-ARCH models. *Scand J Statist*. 2019;46:26–58. [https://doi.org/10.1111/sjos.12337](https://doi.org/10.1111/sjos.12337)
APPENDIX A

PROOFS: MAIN RESULTS

Throughout the proof, $D$ denotes some generic positive constant, independent of $t$, which may differ from line to line.

Proof. Proof of Theorem 1

Note that for the test statistic, we have

\[
T_n = n \int \left| \sum_{j=1}^{n} \tilde{w}_j \exp \left( i \left( t_0 \hat{\epsilon}_j + \sum_{v=1}^{k} t_v X_{j-v} \right) \right) - \sum_{j=1}^{n} \sum_{\ell=1}^{n} \tilde{w}_j \tilde{w}_\ell \exp \left( i \left( t_0 \hat{\epsilon}_j + \sum_{v=1}^{k} t_v X_{\ell-v} \right) \right) \right|^2 \times W(t_0, \ldots, t_k) d(t_0, \ldots, t_k)
\]

\[
= n \int \left[ \left( \sum_{j=1}^{n} \tilde{w}_j \cos \left( t_0 \hat{\epsilon}_j + \sum_{v=1}^{k} t_v X_{j-v} \right) \right) - \sum_{j=1}^{n} \sum_{\ell=1}^{n} \tilde{w}_j \tilde{w}_\ell \cos \left( t_0 \hat{\epsilon}_j + \sum_{v=1}^{k} t_v X_{\ell-v} \right) \right]^2 \times W(t_0, \ldots, t_k) d(t_0, \ldots, t_k)
\]

and with the addition theorems for trigonometric functions, one obtains

\[
T_n = n \int \left\{ \left[ \sum_{j=1}^{n} \tilde{w}_j \cos(t_0 \hat{\epsilon}_j) \left( \cos \left( \sum_{v=1}^{k} t_v X_{j-v} \right) - \sum_{\ell=1}^{n} \tilde{w}_\ell \cos \left( \sum_{v=1}^{k} t_v X_{\ell-v} \right) \right) \right]
- \sum_{j=1}^{n} \tilde{w}_j \sin(t_0 \hat{\epsilon}_j) \left( \sin \left( \sum_{v=1}^{k} t_v X_{j-v} \right) - \sum_{\ell=1}^{n} \tilde{w}_\ell \sin \left( \sum_{v=1}^{k} t_v X_{\ell-v} \right) \right) \right]^2 \times W(t_0, \ldots, t_k) d(t_0, \ldots, t_k)
\]

\[
+ \left[ \sum_{j=1}^{n} \tilde{w}_j \sin(t_0 \hat{\epsilon}_j) \left( \cos \left( \sum_{v=1}^{k} t_v X_{j-v} \right) - \sum_{\ell=1}^{n} \tilde{w}_\ell \cos \left( \sum_{v=1}^{k} t_v X_{\ell-v} \right) \right) \right]^2 \times W(t_0, \ldots, t_k) d(t_0, \ldots, t_k)
\]

From assumption (A8), by the symmetry properties of cosine and sine, we obtain

\[
T_n = \int (S_n(t))^2 W(t) dt,
\]
where

\[
S_n(t) = \sqrt{n} \sum_{j=1}^{n} \tilde{w}_j \cos(t_0 \hat{\epsilon}_j) \left[ \cos \left( \sum_{v=1}^{k} t_v X_{j-v} \right) + \sin \left( \sum_{v=1}^{k} t_v X_{j-v} \right) \right]
- \sum_{\ell=1}^{n} \tilde{w}_\ell \left[ \cos \left( \sum_{v=1}^{k} t_v X_{j-v} \right) + \sin \left( \sum_{v=1}^{k} t_v X_{j-v} \right) \right]
+ \sqrt{n} \sum_{j=1}^{n} \tilde{w}_j \sin(t_0 \hat{\epsilon}_j) \left[ \cos \left( \sum_{v=1}^{k} t_v X_{j-v} \right) - \sin \left( \sum_{v=1}^{k} t_v X_{j-v} \right) \right]
- \sum_{\ell=1}^{n} \tilde{w}_\ell \left[ \cos \left( \sum_{v=1}^{k} t_v X_{\ell-v} \right) - \sin \left( \sum_{v=1}^{k} t_v X_{\ell-v} \right) \right].
\]

For simplicity, for the moment, we consider only

\[
S_n^{(1)}(t) = \sqrt{n} \sum_{j=1}^{n} \tilde{w}_j \cos(t_0 \hat{\epsilon}_j) \left[ \cos \left( \sum_{v=1}^{k} t_v X_{j-v} \right) - \sum_{\ell=1}^{n} \tilde{w}_\ell \cos \left( \sum_{v=1}^{k} t_v X_{\ell-v} \right) \right].
\]

By a second-order Taylor expansion for

\[
\cos \left( t_0 \hat{\epsilon}_j \right) = \cos \left( t_0 \left( \epsilon_j + \epsilon_j \frac{\sigma - \hat{\sigma}}{\hat{\sigma}} (X_{j-1}) + \frac{m - \hat{m}}{\hat{\sigma}} (X_{j-1}) \right) \right)
\]

and introducing the notations

\[
\hat{k}_n = \frac{1}{n} \sum_{i=1}^{n} w_n(X_{i-1}) \quad \text{(A1)}
\]

\[
Y_j(t) = \cos \left( \sum_{v=1}^{k} t_v X_{j-v} \right), \quad j = 1, \ldots, n,
\]

we obtain the expansion \( S_n^{(1)} = S_n^{(1,1)} + S_n^{(1,2)} - \frac{1}{2} S_n^{(1,3)} \), where

\[
S_n^{(1,1)}(t) = \frac{1}{k_n} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \cos(t_0 \hat{\epsilon}_j) \left( Y_j(t) - \frac{1}{k_n} \frac{1}{n} \sum_{\ell=1}^{n} w_n(X_{\ell-1}) Y_\ell(t) \right)
= \frac{1}{k_n} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \left( \cos(t_0 \hat{\epsilon}_j) - E[\cos(t_0 \hat{\epsilon}_j)] \right) \left( Y_j(t) - \frac{1}{k_n} \frac{1}{n} \sum_{\ell=1}^{n} w_n(X_{\ell-1}) Y_\ell(t) \right)
\]

\[
S_n^{(1,2)}(t) = \frac{1}{k_n} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \hat{\epsilon}_j) t_0 \left( \frac{\hat{m} - m}{\hat{\sigma}} (X_{j-1}) + \epsilon_j \frac{\hat{\sigma} - \sigma}{\hat{\sigma}} (X_{j-1}) \right)
\times \left( Y_j(t) - \frac{1}{k_n} \frac{1}{n} \sum_{\ell=1}^{n} w_n(X_{\ell-1}) Y_\ell(t) \right)
\]

\[
S_n^{(1,3)}(t) = \frac{1}{k_n} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \cos(t_0 \hat{\epsilon}_j) t_0^2 \left( \frac{\hat{m} - m}{\hat{\sigma}} (X_{j-1}) + \epsilon_j \frac{\hat{\sigma} - \sigma}{\hat{\sigma}} (X_{j-1}) \right)^2
\times \left( Y_j(t) - \frac{1}{k_n} \frac{1}{n} \sum_{\ell=1}^{n} w_n(X_{\ell-1}) Y_\ell(t) \right)
\]
(with $\xi_j$ between $\varepsilon_j$ and $\hat{\varepsilon}_j$, $j = 1, \ldots, n$). The last term is negligible because

$$\int \left( S_n^{(1,2)}(t) \right)^2 W(t) dt \leq \int t_0^2 W(t) dt \left( \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 + 1 \right)^2 nO_p \left( \frac{a_n^4}{\Delta_n^4} \right) = o_P(1)$$

by assumptions (A3) and (A6) and Proposition 1 and Equation (B1) in Appendix B. Lemmata 2, 3, and 4 (in Appendix B) give further expansions of $S_n^{(1,1)}$ and $S_n^{(1,2)}$. With this, we obtain altogether that $\int (\hat{S}_n^{(1)}(t) - \tilde{S}_n^{(1)}(t))^2 W(t) dt = o_P(1)$, where

$$\tilde{S}_n^{(1)}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n w_n(X_{j-1}) \left( \cos(t_0 \varepsilon_j) - E \left[ \cos(t_0 \varepsilon_j) \right] \right) (Y_j(t) - E[Y_j(t)])$$

$$+ t_0 \left( \frac{1}{2} \varepsilon_j^2 - 1 \right) E \left[ \cos(t_0 \varepsilon_j) \right] (Y_j(t) - E[Y_j(t)])$$

$$\hat{S}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n w_n(X_{j-1}) \left[ \left( \cos(t_0 \varepsilon_j) - E \left[ \cos(t_0 \varepsilon_j) \right] \right) (Y_j(t) + Z_j(t) - E \left[ Y_j(t) + Z_j(t) \right]) \right]$$

$$+ \left( \sin(t_0 \varepsilon_j) - E \left[ \sin(t_0 \varepsilon_j) \right] \right) (Y_j(t) - Z_j(t) - E \left[ Y_j(t) - Z_j(t) \right])$$

$$+ t_0 \left( \varepsilon_j E \left[ \sin(t_0 \varepsilon_j) \right] + \frac{1}{2} \varepsilon_j^2 - 1 \right) E \left[ \cos(t_0 \varepsilon_j) \right] (E \left[ Y_j(t) + Z_j(t) \mid X_{j-1} \right] - E \left[ Y_j(t) + Z_j(t) \right])$$

$$- t_0 \left( \varepsilon_j E \left[ \cos(t_0 \varepsilon_j) \right] + \frac{1}{2} \varepsilon_j^2 - 1 \right) E \left[ \sin(t_0 \varepsilon_j) \right] (E \left[ Y_j(t) - Z_j(t) \mid X_{j-1} \right] - E \left[ Y_j(t) - Z_j(t) \right])$$

with

$$Z_j(t) = \sin \left( \frac{1}{k} \sum_{i=1}^k t_i X_{j-i} \right) \quad j = 1, \ldots, n.$$

To finish the proof of Theorem 1, we apply theorem 22 (pages 380, 381) in Ibragimov and Has’minskii (1981). In order to verify their assumptions, it suffices to show that

(i) $\tilde{S}_n(t)$ has asymptotically normal distribution with zero mean and finite variance;

(ii) for any compact set $F$ in $\mathbb{R}^{k+1}$,

$$\sup_n E \int_{F} \tilde{S}_n^2(t) W(t) dt < \infty;$$

(iii) $$E \left| \tilde{S}_n^2(t_1) - \tilde{S}_n^2(t_2) \right| \leq D \left\| t_1 - t_2 \right\|^{\gamma}, \forall t_1, t_2$$

for some $\gamma > 0$ and some $D > 0$;

(iv) for all $\eta > 0$, there exists some compact set $F_\eta$ in $\mathbb{R}^{k+1}$ with

$$E \int_{\mathbb{R}^{k+1} \setminus F_\eta} \tilde{S}_n^2(t) W(t) dt < \eta, \forall n, E \int_{\mathbb{R}^{k+1} \setminus F_\eta} \hat{S}_n^2(t) W(t) dt < \eta.$$

Since $\tilde{S}_n(t)$ is the sums of martingale differences for each $t$, the central limit theorem for martingale differences can be applied, which implies (i). Direct calculations gives (ii). Concerning (iii), we have

$$E \left| \tilde{S}_n^2(t_1) - \tilde{S}_n^2(t_2) \right| \leq E \left[ \left| \tilde{S}_n^2(t_1) \right| \left( \left| \tilde{S}_n^2(t_1) \right| + \left| \tilde{S}_n^2(t_2) \right| \right) \right]$$

$$\leq \left( E \left| \tilde{S}_n(t_1) - \tilde{S}_n(t_2) \right|^2 \times E \left[ \left| \tilde{S}_n(t_1) \right| + \left| \tilde{S}_n(t_2) \right| \right] \right)^{1/2},$$
and since $E(\tilde{S}_n(t)^2) \leq D_n$, it suffices to study $E(\tilde{S}_n(t_1) - \tilde{S}_n(t_2))^2$. We show here the needed inequality only for one of the terms in $\tilde{S}_n(t_1) - \tilde{S}_n(t_2)$; all others are treated in the same way. Particularly,

$$
E \left( \frac{1}{n} \sum_{j=1}^{n} W_n(X_{j-1}) \left( \left( \cos(t_0 \epsilon_j) - E \cos(t_0 \epsilon_j) \right) \left( Y_j(t_1) - EY_j(t_1) \right) \right) \right.
$$

$$
- \left( \left( \cos(t_0 \epsilon_j) - E \cos(t_0 \epsilon_j) \right) \left( Y_j(t_2) - EY_j(t_2) \right) \right) \right)^2
$$

$$
= E \left( W_n(X_{j-1}) \left( \left( \cos(t_0 \epsilon_j) - E \cos(t_0 \epsilon_j) \right) \left( Y_j(t_1) - EY_j(t_1) \right) \right) \right.
$$

$$
- \left( \left( \cos(t_0 \epsilon_j) - E \cos(t_0 \epsilon_j) \right) \left( Y_j(t_2) - EY_j(t_2) \right) \right) \right)^2,
$$

where we used smoothness of cosine and moment assumptions. Proceeding similarly with other terms and putting all together, we conclude that

$$
E \left( \tilde{S}_n(t_1) - \tilde{S}_n(t_2) \right)^2 \leq D |t_1 - t_2|.
$$

This implies the item (iii). Item (iv) follows straightforwardly by our moment assumptions and integrability of $W$.

Combining all the above arguments, we can infer that the assertion of Theorem 1 holds true; see lemma 7.1 and proof of theorem 4.1 (a) in Hlávka, Hušková, Kirch, and Meintanis (2017) for a similar argumentation.

**Proof of Lemma 1.** Using assumption (A8), it follows that $\int_{\mathbb{R}^{k+1}} |t_0^d W(t_0, \ldots, t_k)| \, dt(t_0, \ldots, t_k) < \infty$, and since $W(t_0, \ldots, t_k) = V_0(t_0) \prod_{i=1}^{k} V_i(t_i)$, by assumption, one obtains

$$
\int_{\mathbb{R}^{k+1}} |t_0^d W(t_0, \ldots, t_k)| \, dt(t_0, \ldots, t_k) = \int_{\mathbb{R}^{k+1}} t_0^d V_0(t_0) \prod_{i=1}^{k} V_i(t_i) \, dt(t_0, \ldots, t_k)
$$

$$
= \int_{\mathbb{R}^{k}} t_0^d V_0(t_0) \prod_{i=1}^{k} \int_{\mathbb{R}} V_i(t_i) \, dt_i,
$$

which gives that $V_i \in L^1(\mathbb{R})$ for any $i = 0, \ldots, k$. Hence, the Fourier transformation of any $V_i$, say $\mathcal{F}[V_i]$, exists. The representation of the test statistic is now straightforwardly computed by using the definition of the Fourier transformation and of the empirical characteristic functions besides the multiplicative structure of $W$. Since the computation is tedious but without further insights, this part of the proof is omitted here.

**Proof of Theorem 2.** We use the same decomposition of $T_n = \int (S_n(t))^2 W(t) \, dt$ as in the proof of Theorem 1. Please note that Proposition 1 remains true under the assumptions of Theorem 2. A careful inspection of the proof of Theorem 1 shows that applying this Proposition one obtains $S_n = \tilde{S}_n + R_n$, where $\int R_n^2(t) W(t) \, dt = o_P(1)$ and

$$
\frac{\tilde{S}_n(t)}{\sqrt{n}} = \frac{1}{n} \sum_{j=1}^{n} W_n(X_{j-1}) \left[ \left( \cos(t_0 \epsilon_j) - E \cos(t_0 \epsilon_j) \right) \left( Y_j(t) + Z_j(t) - E \left[ Y_j(t) + Z_j(t) \right] \right) \right.
$$

$$
+ \left( \sin(t_0 \epsilon_j) - E \sin(t_0 \epsilon_j) \right) \left( Y_j(t) - Z_j(t) - E \left[ Y_j(t) - Z_j(t) \right] \right) \right].
$$
The proof is finished as the end of the proof of Theorem 1 applying theorem 22 (pages 380, 381) in Ibragimov and Has’minskii (1981). To this end, condition (i) is replaced by convergence in probability of $\bar{S}_n(t)/n^{1/2}$ to

$$\tilde{S}(t) = E \left[ (\cos(t_0 \epsilon_j) - E[\cos(t_0 \epsilon_j)]) \left( Y_j(t) + Z_j(t) - E[Y_j(t) + Z_j(t)] \right) \right]$$

$$+ \left( \sin(t_0 \epsilon_j) - E[\sin(t_0 \epsilon_j)] \right) \left( Y_j(t) - Z_j(t) - E[Y_j(t) - Z_j(t)] \right) ,$$

for all $t$, whereas in conditions (ii)–(iv) $\bar{S}_n$ is replaced by $\tilde{S}_n/n^{1/2}$. Thus, we obtain the convergence of $T_n/n$ to $\int (\tilde{S}(t))^2 \tilde{W}(t) dt$ in probability. Note further that, by the addition theorems for trigonometric functions and symmetry properties of cosine and sine, it holds $\bar{T} = \int (\tilde{S}(t))^2 \tilde{W}(t) dt$. This completes the proof. □

**Proof of Theorem 3.** Note that $\bar{T}_n = \int (\tilde{S}_n(t))^2 \tilde{W}(t) dt$, where

$$\tilde{S}_n^{(1)}(t) = \sqrt{n} \sum_{j=1}^{n} w_j I[\epsilon_j \leq t_0] \left[ I\{X_{j-1} \leq t_1, \ldots, X_{j-k} \leq t_k\} - \sum_{\ell=1}^{n} w_\ell I\{X_{\ell-1} \leq t_1, \ldots, X_{\ell-k} \leq t_k\} \right]$$

has a very similar structure to $S_n^{(1)}(t)$ from the proof of Theorem 1. Analogous to lemma 3 in Selk and Neumeyer (2013) (setting $s = 1$), one can show the expansion

$$\tilde{S}_n^{(1)}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_j n^{-1} \left[ \tilde{Y}_j(t) - \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell (X_{\ell-1}) \tilde{\tilde{Y}}_\ell(t) \right]$$

$$+ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_j (X_{j-1}) \left( F_\epsilon \left( \frac{m - m_\sigma}{\sigma} (X_{j-1}) + t_0 \frac{\tilde{\sigma}}{\sigma} (X_{j-1}) \right) - F_\epsilon(t_0) \right)$$

$$\times \left[ \tilde{Y}_j(t) - \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell (X_{\ell-1}) \tilde{\tilde{Y}}_\ell(t) \right] + o_p(1)$$

uniformly with respect to $t$, where

$$\tilde{Y}_j(t) = I\{X_{j-1} \leq t_1, \ldots, X_{j-k} \leq t_k\}.$$ 

Now with a Taylor expansion of $F_\epsilon$, one obtains $\bar{T}_n = \int (\tilde{S}_n^{(1,1)}(t) + \tilde{S}_n^{(1,2)}(t))^2 \tilde{W}(t) dt + o_p(1)$ with

$$\tilde{S}_n^{(1,1)}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_j n^{-1} \left[ \tilde{Y}_j(t) - \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell (X_{\ell-1}) \tilde{\tilde{Y}}_\ell(t) \right]$$

$$\tilde{S}_n^{(1,2)}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_j (X_{j-1}) \left( \frac{m - m_\sigma}{\sigma} (X_{j-1}) + t_0 \frac{\tilde{\sigma}}{\sigma} (X_{j-1}) \right)$$

$$\times \left[ \tilde{Y}_j(t) - \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} w_\ell (X_{\ell-1}) \tilde{\tilde{Y}}_\ell(t) \right] .$$

Note that the second-order term arising from the Taylor expansion is negligible using (A15) and integrability of $\tilde{\tilde{W}}$ and an argument analogous to the one for $S_n^{(1,3)}$ in the proof of
Theorem 1. Now, completely analogous to the proof of Lemma 2, one shows that $\tilde{S}_n^{(1,1)}(t)$ can be replaced by

$$\tilde{S}_n^{(1,1)}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \left( I[\varepsilon_j \leq t_0] - F_\varepsilon(t_0) \right) \left[ \bar{Y}_j(t) - E \left[ \bar{Y}_j(t) \right] \right],$$

while as in the proof of Lemma 3, one can show that $\tilde{S}_n^{(1,2)}(t)$ can be replaced by

$$\tilde{S}_n^{(1,2)}(t) = \frac{f_\varepsilon(t_0)}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \left( \frac{\hat{m} - m}{\sigma}(X_{j-1}) + t_0 \frac{\hat{\sigma}}{\sigma}(X_{j-1}) \right) \left( \bar{Y}_j(t) - E \left[ \bar{Y}_j(t) \right] \right).$$

The further decomposition of $\tilde{S}_n^{(1,2)}$ is actually simpler than the one for $\tilde{S}_n^{(1,2)}$ in Lemma 4 because here we do not have the random denominator $\hat{\sigma}$. One obtains

$$\int \left( \tilde{S}_n^{(1,2)}(t) - \tilde{S}_n^{(1,2,1)}(t) - \tilde{S}_n^{(1,2,2)}(t) \right)^2 \tilde{W}(t) dt = o_p(1),$$

where

$$\tilde{S}_n^{(1,2,1)}(t) = \frac{f_\varepsilon(t_0)}{n^{3/2}} \sum_{j=1}^{n} \sum_{i=1}^{1} \frac{1}{c_n} K \left( X_{j-1} - X_{i-1} \right) \frac{w_n(X_{j-1}) \sigma(X_{i-1}) \varepsilon_i}{\sigma(X_{i-1}) f_X(X_{j-1})} \left( \bar{Y}_j(t) - E \left[ \bar{Y}_j(t) \right] \right) \times \left( \bar{Y}_j(t) - E \left[ \bar{Y}_j(t) \right] \right).$$

Now, mimicking the proof of Lemma 5 to show that

$$\int \left( \tilde{S}_n^{(1,2)}(t) - \tilde{S}_n^{(1,2,1)}(t) \right)^2 \tilde{W}(t) dt = o_p(1),$$

$$\int \left( \tilde{S}_n^{(1,2,2)}(t) - \tilde{S}_n^{(1,2,1)}(t) \right)^2 \tilde{W}(t) dt = o_p(1),$$

where

$$\tilde{S}_n^{(1,2,1)}(t) = \frac{f_\varepsilon(t_0)}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \varepsilon_j \left( E \left[ \bar{Y}_j(t) \mid X_{j-1} \right] - E \left[ \bar{Y}_j(t) \right] \right),$$

$$\tilde{S}_n^{(1,2,2)}(t) = \frac{t_0 f_\varepsilon(t_0)}{2 \sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \left( \varepsilon_j^2 - 1 \right) \left( E \left[ \bar{Y}_j(t) \right] \right) \left( E \left[ \bar{Y}_j(t) \mid X_{j-1} \right] - E \left[ \bar{Y}_j(t) \right] \right),$$

the only difference is in the term

$$\tilde{V}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma(X_{i-1}) \varepsilon_i \int \frac{1}{c_n} K \left( \frac{x - X_{i-1}}{\sigma} \right) \left( \frac{w_n(x) \varphi(t,x)}{\sigma} - \frac{w_n(X_{i-1}) \varphi(t,X_{i-1})}{\sigma(X_{i-1})} \right) dx$$

because, for

$$\varphi(t,x) = E \left[ \bar{Y}_j(t) \mid X_{j-1} = x \right] - E \left[ \bar{Y}_j(t) \right]$$

$$= I[x \leq t_1] P(X_{j-1} \leq t_2, \ldots, X_{i-k} \leq t_k) - F_{X_{j-1}}(t_1, \ldots, t_k),$$

no Lipschitz continuity as in assumption (A4) can be applied. Instead, to show that $E[\int \tilde{V}_n^2(t) \tilde{W}(t) dt] = o(1)$, one applies assumption (A16) and a direct calculation using that

$$\int \left( I[z + c_n u \leq t_0] - I[z \leq t_0] \right) w_n(z) dx(z) dz K(u) du = O(c_n).$$
Putting all results together, we have eventually shown that \( \tilde{T}_n = \int (\hat{S}_n(t))^2 \tilde{W}(t) \, dt + o_P(1) \) with

\[
\hat{S}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \omega_n(X_{j-1}) \left[ (I\{\varepsilon_j \leq t_0\} - F_{\varepsilon}(t_0)) \left( \tilde{Y}_j(t) - E[\tilde{Y}_j(t)] \right) + \left( f_{\varepsilon}(t_0) \varepsilon_j + \frac{1}{2} t_0 f_{\varepsilon}(t_0) \left( \varepsilon_j^2 - 1 \right) \right) \left( E[\tilde{Y}_j(t) \mid X_{j-1}] - E[\tilde{Y}_j(t)] \right) \right].
\]

The convergence in distribution follows as in the proof of Theorem 1 with an application of theorem 22 in Ibragimov and Has’minskii (1981). Here, in particular, to show that

\[
E \left| \frac{\hat{S}_n(t_1) - \hat{S}_n(t_2)}{\sqrt{\Delta_n}} \right|^2 \leq D \|t_1 - t_2\|\gamma, \forall t_1, t_2,
\]

for some \( \gamma > 0 \) and some \( D > 0 \), one has to make use of assumption (A17).

\[\Box\]

**APPENDIX B**

**AUXILIARY RESULTS**

First note that, for \( \hat{k}_n \) defined in (A1), one obtains directly that \( E[(\hat{k}_n - 1)^2] \) can be bounded by \( 1 - F_{X_{\hat{k}_n}}(\frac{a_n}{2}) - F_{X_{\hat{k}_n}}(-\frac{a_n}{2}) = o(1) \), and thus, we have

\[
\hat{k}_n = 1 + o_P(1).
\]

**Proposition 1.** Let \( (X_j)_{j \in \mathbb{Z}} \) be a strictly stationary time series with marginal density \( f_X \). Define \( m(x) = E[X_{j+1} \mid X_j = x] \) and \( \sigma^2(x) = \text{Var}(X_{j+1} \mid X_j = x) \) and assume (A1), (A2), (A5), (A6). Let \( \Delta_n = \inf_{|x| \leq a_n} f_X(x) \), \( a_n^* = ((\log n)/(nc_n))^1/2 + (\log n)^p c_n^2 \) and \( b_n^* = (c_n/n)^{1/2} + c_n^2(\log n)^p \). Here, \( D > 0 \) is some multiple of \( r \) from assumption (A2) and may differ from line to line. We then have

\[\text{(i)}\]

\[
\sup_{|x| \leq a_n} \left| \widehat{f}_X(x) - f_X(x) \right| = O_P \left( a_n^* \right),
\]

\[
\sup_{|x| \leq a_n} \left| \hat{m}(x) - m(x) \right| = O_P \left( \frac{a_n^*}{\Delta_n} \right),
\]

\[
\sup_{|x| \leq a_n} \left| \hat{\sigma}(x) - \sigma(x) \right| = O_P \left( \frac{a_n^*}{\Delta_n} \right).
\]

\[\text{(ii)}\]

\[
\sup_{|x| \leq a_n} \left| \frac{1}{nc_n} \sum_{j=1}^{n} K \left( \frac{X_{j-1} - x}{c_n} \right) (m(X_{j-1}) - m(x)) \right| = O_P \left( b_n^* \right),
\]

\[
\sup_{|x| \leq a_n} \left| \frac{1}{nc_n} \sum_{j=1}^{n} K \left( \frac{X_{j-1} - x}{c_n} \right) (\sigma^2(X_{j-1}) - \sigma^2(x)) \right| = O_P \left( b_n^* \right),
\]

\[
\sup_{|x| \leq a_n} \left| \frac{1}{nc_n} \sum_{j=1}^{n} K \left( \frac{X_{j-1} - x}{c_n} \right) (m^2(X_{j-1}) - m^2(x)) \right| = O_P \left( b_n^* \right),
\]

\[\text{(iii)}\]

\[
\sup_{|x| \leq a_n} \left| \frac{1}{nc_n} \sum_{j=1}^{n} K \left( \frac{X_{j-1} - x}{c_n} \right) \sigma(X_{j-1}) \varepsilon_j (m(X_{j-1}) - m(x)) \right| = O_P \left( b_n^* \right).
\]
**Proof.** The first two results of (i) are stated in theorems 6 and 8 by Hansen (2008) without the 
(log \(n\))\(^3\) factor of the \(c_n^2\) term. In comparison with Hansen (2008), we use a different bounding 
for the expectation terms since we do not assume second derivatives. For example, we 
obtain, making use of the mean value theorem, the properties of the kernel function and our assumption (A2),

\[
\sup_{|x| \leq a_n} \left| E \left[ f_X(x) - f_X(x) \right] \right| = \sup_{|x| \leq a_n} \left| \int K(u) (f_X(x - c_n u) - f_X(x)) \, du \right| 
\leq \sup_{|x| \leq a_n} c_n \int K(u) |u| \sup_{\xi \text{ between } x \text{ and } x - c_n u} |f'(\xi) - f'(x)| \, du
\]

\[= O \left( c_n^2 (\log n)^r \right). \]

The result on \(\hat{\sigma}\) follows similarly to the derivations by Hansen (2008) by noting that \(\hat{\sigma}^2(x) = \hat{s}(x) - \hat{m}^2(x)\), where \(\hat{s}\) is the Nadaraya–Watson estimator for \(s(x) = E[X_j^2 | X_j = x]\) based on 
the observation pairs \((X_{j-1}, X_j^2), j = 1, \ldots, n\).

Toward the results in (ii), we treat only the first one since the others follow analogously. 
Note that, by the mean value theorem,

\[m(X_{i-1}) - m(x) = (X_{i-1} - x)m'(x) + (X_{i-1} - x) \left( m'(\xi_{X_{i-1}, x}) - m'(x) \right) \]

for some \(\xi_{X_{i-1}, x}\) between \([\min(X_{i-1}, x), \max(X_{i-1}, x)]\), where the absolute value of the second 
summand can be bounded by \((X_{i-1} - x)^2 (\log n)^r\) because of assumption (A2). It thus suffices 
to show that

\[
\sup_{|x| \leq a_n} \left| \frac{1}{nc_n} \sum_{i=1}^n K \left( \frac{X_{i-1} - x}{c_n} \right) (X_{i-1} - x)^2 \right| = O_P \left( c_n^2 \right)
\]

\[
\sup_{|x| \leq a_n} \left| \frac{1}{nc_n} \sum_{i=1}^n K \left( \frac{X_{i-1} - x}{c_n} \right) (X_{i-1} - x) m'(x) \right| = O_P \left( b_n^* \right).
\]

The first relation is straightforward by assumption (A2) and applying Theorem 2 in 
Hansen (2008) with \(Y_i = 1\) and the kernel \(u \mapsto K(u)u^2\). For the latter one, we receive, with 
the same theorem applied with \(Y_i = 1\) and kernel \(u \mapsto K(u)\),

\[
\sup_{|x| \leq a_n} \left| \frac{1}{nc_n} \sum_{i=1}^n K \left( \frac{x - X_{i-1}}{c_n} \right) (x - X_{i-1}) - E \left[ K \left( \frac{x - X_{i-1}}{c_n} \right) (x - X_{i-1}) \right] \right| = O_P \left( \left( \frac{\log n}{nc_n} \right)^{1/2} c_n \right).
\]

Furthermore, by direct calculation,

\[
\frac{1}{nc_n} \sum_{i=1}^n E \left[ K \left( \frac{x - X_{i-1}}{c_n} \right) (x - X_{i-1}) \right] = \frac{1}{c_n} \int K \left( \frac{x - y}{c_n} \right) (x - y) f_{X_{i-1}}(y) dy
\]

\[= c_n \int K(z) f_{X_{i-1}}(x - z c_n) dz = O \left( c_n^2 \int K(z) dz \sup_{x \in I_n} \left| f_X'(x) \right| \right) = O \left( c_n^2 (\log n)^r \right), \]

where we utilize assumptions (A2) and (A5).

The result (iii) can be proved in the same way as the results in (ii). Just set \(Y_i = |\varepsilon_i|\) for the first and \(Y_i = \sigma(X_{i-1}) \varepsilon_i\) for the second relation (when applying Theorem 2 by Hansen, 2008).
and note that
\[ E \left[ K \left( \frac{X - X_{i-1}}{c_n} \right) (x - X_{i-1})\sigma(X_{i-1})\varepsilon_i \right] = 0. \]

**Lemma 2.** Under the assumptions of Theorem 1, we have
\[ \int \left( S_n^{(1,1)}(t) - \tilde{S}_n^{(1,1)}(t) \right)^2 W(t) \, dt = o_P(1), \]
where
\[ \tilde{S}_n^{(1,1)}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \left( \cos(t_0\varepsilon_j) - E \left[ \cos(t_0\varepsilon_j) \right] \right) \left( Y_j(t) - E \left[ Y_j(t) \right] \right). \]

**Proof.** Because of (B1), we have
\[ S_n^{(1,1)}(t) = (1 + o_P(1)) \left( \tilde{S}_n^{(1,1)}(t) - J_n(t)J_n^{(1)}(t) - J_n(t)J_n^{(2)}(t) \right), \]
where
\[ J_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \left( \cos(t_0\varepsilon_j) - E \left[ \cos(t_0\varepsilon_j) \right] \right), \]
\[ J_n^{(1)}(t) = \frac{1}{n} \sum_{\ell=1}^{n} (Y_\ell(t) - E[Y_\ell(t)]) \]
\[ J_n^{(2)}(t) = \frac{1}{n} \sum_{\ell=1}^{n} Y_\ell(t) (w_n(X_{\ell-1}) - \hat{k}_n) \frac{1}{k_n}. \]

Note that, from assumption (A1), it follows that \( \beta > 2 \), and thus, \( \sum_{i=0}^{\infty} (i + 1)\alpha(i) < \infty \). From this, the centeredness of the summands (under the null) and the boundedness of cosine analogously to the proof of Theorem 2 by Yokoyama (1980), one obtains
\[ E \left[ (J_n(t))^4 \right] \leq D \tag{B2} \]
\[ E \left[ (J_n^{(1)}(t))^4 \right] \leq \frac{1}{n^2} D. \tag{B3} \]

The constant \( D \) can be chosen independent of \( t \) because of the boundedness of the cosine function. Thus, from the Cauchy Schwarz inequality, we obtain, directly,
\[ E \left[ \int \left( J_n(t)J_n^{(1)}(t) \right)^2 W(t) \, dt \right] = O \left( \frac{1}{n} \right). \]

Now, note that
\[ J_n^{(2)}(t) = (1 + o_P(1)) \left( \frac{1}{n} \sum_{\ell=1}^{n} Y_\ell(t) (w_n(X_{\ell-1}) - E[w_n(X_{\ell-1})]) \right. \]
\[ \left. - \frac{1}{n} \sum_{\ell=1}^{n} Y_\ell(t) \frac{1}{n} \sum_{j=1}^{n} (w_n(X_{j-1}) - E[w_n(X_{j-1})]) \right), \]
and thus, by the boundedness of \( Y_\ell \), we have, uniformly with respect to \( t \),
\[ |J_n^{(2)}(t)| = O_P(1) \frac{1}{n} \sum_{j=1}^{n} \left| w_n(X_{j-1}) - E[w_n(X_{j-1})] \right| = o_P(1) \tag{B4} \]
by a consideration of the expectation of the sum because of the properties of the weight function. We obtain
\[ \int \left( J_n(t) J_n^{(2)}(t) \right)^2 W(t) \, dt = o_P(1) \int \left( J_n(t) \right)^2 W(t) \, dt = o_P(1) \]
by an application of (B2).

Lemma 3. Under the assumptions of Theorem 1, we have
\[ \int \left( S_n^{(1,2)}(t) - \tilde{S}_n^{(1,2)}(t) \right)^2 W(t) \, dt = o_P(1), \]
where
\[ \tilde{S}_n^{(1,2)}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) t_0 \left( \frac{\hat{m} - m}{\hat{\sigma}} (X_{j-1}) + \epsilon_j \frac{\hat{\sigma} - \sigma}{\hat{\sigma}} (X_{j-1}) \right) (Y_j(t) - E[Y_j(t)]). \]

Proof. Because of (B1), we have
\[ S_n^{(1,2)}(t) = (1 + o_P(1)) \left( S_n^{(1,2)}(t) - t_0 I_n(t) J_n^{(1)}(t) - t_0 I_n(t) J_n^{(2)}(t) \right) \]
with \( J_n^{(1)} \) and \( J_n^{(2)} \) as in Lemma 2 and
\[ I_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) \left( \frac{\hat{m} - m}{\hat{\sigma}} (X_{j-1}) + \epsilon_j \frac{\hat{\sigma} - \sigma}{\hat{\sigma}} (X_{j-1}) \right). \]
Now,
\[ |I_n(t)| \leq \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (|\epsilon_j| + 1) O_P \left( \frac{a_n^*}{\Delta_n} \right) = o_P(\sqrt{n}) \]
uniformly with respect to \( t \) by assumption (A3) and Proposition 1 (i). Thus,
\[ \int \left( t_0 I_n(t) J_n^{(1)}(t) \right)^2 W(t) \, dt = o_P(n) \int t_0^2 \left( J_n^{(1)}(t) \right)^2 W(t) \, dt = o_P(1) \]
by (B3).
Furthermore, by (B4), we obtain
\[ \int \left( t_0 I_n(t) J_n^{(2)}(t) \right)^2 W(t) \, dt = o_P(1) \int t_0^2 (I_n(t))^2 W(t) \, dt = o_P(1), \]
where one yields the last equality as follows. Similar to the proof of Lemma 4, one can first replace the random denominators \( \hat{\sigma} \hat{f}_X \) in the definition of \( I_n \) by their true counterparts \( \sigma f_X \) applying Proposition 1. Let \( \tilde{I}_n \) denote the resulting term, then
\[ E[\int t_0^2 (\tilde{I}_n(t))^2 W(t) \, dt] = O(1) \]
is shown by straightforward calculations.

Lemma 4. Under the assumptions of Theorem 1, we have
\[ \int \left( \tilde{S}_n^{(1,2)}(t) - \tilde{S}_n^{(1,2,1)}(t) - \tilde{S}_n^{(1,2,2)}(t) \right)^2 W(t) \, dt = o_P(1), \]
where

\[
\tilde{S}_{n}^{(1,2)}(t) = \frac{t_0}{n^{3/2}} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{w_n(X_{j-1}) \sigma(X_{j-1}) \epsilon_j}{\sigma(X_{j-1}) f_X(X_{j-1})} \sin(t_0 \epsilon_j) \\
\times (Y_j(t) - E[Y_j(t)])
\]

\[
\tilde{S}_{n}^{(1,2,2)}(t) = \frac{t_0}{n^{3/2}} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{w_n(X_{j-1}) \sigma^2(X_{j-1}) (\epsilon_j^2 - 1)}{2 \sigma^2(X_{j-1}) f_X(X_{j-1})} \sin(t_0 \epsilon_j) \epsilon_j \\
\times (Y_j(t) - E[Y_j(t)]) .
\]

**Proof.** Recall the definition of \(\tilde{S}_{n}^{(1,2)}\) in Lemma 3 and note that

\[
\frac{t_0}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) \frac{m - \hat{m}}{\hat{\sigma}} (X_{j-1}) (Y_j(t) - E[Y_j(t)])
\]

\[
= \frac{t_0}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) \frac{\hat{m} - m}{\sigma}(X_{j-1}) \frac{\hat{\sigma} - \sigma}{\sigma} (X_{j-1}) (Y_j(t) - E[Y_j(t)]) + R_{n}^{(1)}(t) + R_{n}^{(2)}(t)
\]

\[
= \tilde{S}_{n}^{(1,2,1)}(t) + R_{n}^{(1)}(t) + R_{n}^{(2)}(t) + R_{n}^{(3)}(t),
\]

where

\[
R_{n}^{(1)}(t) = \frac{t_0}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) \frac{\hat{m} - m}{\sigma}(X_{j-1}) \frac{\hat{\sigma} - \sigma}{\sigma} (X_{j-1}) (Y_j(t) - E[Y_j(t)])
\]

\[
R_{n}^{(2)}(t) = \frac{t_0}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) \frac{\hat{m} - m}{\sigma}(X_{j-1}) \frac{f_X - \hat{f}_X}{f_X}(X_{j-1}) (Y_j(t) - E[Y_j(t)])
\]

\[
R_{n}^{(3)}(t) = \frac{t_0}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) \left\{ \frac{1}{nc_n} \sum_{i=1}^{n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{(m(X_{i-1}) - m(X_{j-1}))}{f_X(X_{j-1})} \right\} \\
\times (Y_j(t) - E[Y_j(t)]) .
\]

By Proposition 1 (i), one directly obtains that \(\int (R_n^{(j)}(t))^2 W(t) dt\) for \(j = 1, 2\) is of rate \(O_P(n(a_n^2/\Delta_n)^4) = o_P(1)\).

Concerning \(\int (R_n^{(3)}(t))^2 W(t) dt\), notice that

\[
\left| R_{n}^{(3)}(t) \right| \leq D \frac{t_0}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sup_{|x| \leq a_n} \left| \frac{1}{nc_n^2} \sum_{i=1}^{n} K \left( \frac{x - X_{i-1}}{c_n} \right) (m(X_{i-1}) - m(x)) \right|
\]

uniformly in \(t\), which, together with assertion 1 (ii), implies the rate \(\int (R_n^{(3)}(t))^2 W(t) dt = O_P(n(\log n)^{4r}(b_n^*)^2) = o_P(1)\). The latter equality follows from assumption (A7).

Concerning the second term in the definition of \(\tilde{S}_{n}^{(1,2)}\) in Lemma 3, note that, because of \(\hat{\sigma} - \sigma = (\hat{\sigma}^2 - \sigma^2)/(\hat{\sigma} + \sigma)\), analogous to before, one shows that

\[
\frac{t_0}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) \frac{\hat{m} - m}{\sigma}(X_{j-1}) (Y_j(t) - E[Y_j(t)]) \\
= \frac{t_0}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) \frac{(\hat{\sigma}^2 - \sigma^2)(\hat{f}_X - f_X)}{2\sigma^2 f_X}(X_{j-1}) (Y_j(t) - E[Y_j(t)]) + R_{n}^{(4)}(t),
\]

where \(\int (R_n^{(4)}(t))^2 W(t) dt = o_P(1)\) now follows from Proposition 1 (i).
To treat the remaining term further, we first insert the definition of $\hat{\sigma}^2$ and then use the fact that $m^2 - \hat{m}^2 = 2m(m - \hat{m}) - (m - \hat{m})^2$ and insert the definition of $\hat{m}$. With this, one obtains

$$
\frac{t_0}{\sqrt{n}} \sum_{j=1}^{n} w_n(X_{j-1}) \sin(t_0 \epsilon_j) \epsilon_j \left( \frac{(\hat{\sigma}^2 - \sigma^2) f_X(X_{j-1})}{2\sigma^2 f_X(X_{j-1})} (Y_j(t) - E[Y_j(t)]) \right)
$$

where

\begin{align*}
R_n^{(5)}(t) &= t_0 \frac{1}{n^{3/2}} \sum_{j=1}^{n} \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{w_n(X_{j-1})}{2\sigma^2 f_X(X_{j-1})} \sin(t_0 \epsilon_j) \\
R_n^{(6)}(t) &= t_0 \frac{1}{n^{3/2}} \sum_{j=1}^{n} \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{w_n(X_{j-1})}{2\sigma^2 f_X(X_{j-1})} \sin(t_0 \epsilon_j) \\
R_n^{(7)}(t) &= t_0 \frac{1}{n^{3/2}} \sum_{j=1}^{n} \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{w_n(X_{j-1})}{\sigma^2 f_X(X_{j-1})} \sin(t_0 \epsilon_j) \\
R_n^{(8)}(t) &= t_0 \frac{1}{n^{3/2}} \sum_{j=1}^{n} \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{w_n(X_{j-1})}{\sigma^2 f_X(X_{j-1})} \sin(t_0 \epsilon_j) \\
R_n^{(9)}(t) &= t_0 \frac{1}{n^{3/2}} \sum_{j=1}^{n} \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{w_n(X_{j-1})}{2\sigma^2 f_X(X_{j-1})} \sin(t_0 \epsilon_j)
\end{align*}

and one can show $\int (R_n^{(j)}(t))^2 W(t) dt = o_p(1)$ completely analogous to the treatment of $R_n^{(3)}$ for $j = 5, 6, 7, 8$ and of $R_n^{(1)}$ for $j = 9$.

Lemma 5. Under the assumptions of Theorem 1, we have

$$
\int \left( \tilde{\mathcal{S}}_n^{(1,2,1)}(t) - \tilde{\mathcal{S}}_n^{(1,2,1)}(t) \right)^2 W(t) dt = o_p(1)
$$

$$
\int \left( \tilde{\mathcal{S}}_n^{(1,2,2)}(t) - \tilde{\mathcal{S}}_n^{(1,2,2)}(t) \right)^2 W(t) dt = o_p(1),
$$

where

\begin{align*}
\tilde{\mathcal{S}}_n^{(1,2,1)}(t) &= t_0 E [\sin(t_0 \epsilon_1)] \sum_{j=1}^{n} w_n(X_{j-1}) \epsilon_j (E[Y_j(t) | X_{j-1}] - E[Y_j(t)]) \\
\tilde{\mathcal{S}}_n^{(1,2,2)}(t) &= t_0 E [\sin(t_0 \epsilon_1)] \sum_{j=1}^{n} w_n(X_{j-1}) (\epsilon_j - 1)^2 (E[Y_j(t) | X_{j-1}] - E[Y_j(t)])
\end{align*}

Proof. We only prove the first assertion, and the second one can be shown completely analogous. We have the expansion

$$
\tilde{S}_n^{(1,2,1)}(t) - \tilde{S}_n^{(1,2,1)}(t) = U_n(t) + t_0 E[\sin(t_0 \epsilon_1)] V_n(t),
$$
where
\[ U_n(t) = \frac{1}{n^{3/2}} \sum_{j=1}^{n} \sum_{i=1}^{n} \varphi(t, \varepsilon_i, X_{i-1}, \zeta_j) \]
with \( \zeta_j = (X_{j-1}, \ldots, X_{j-k}) \),
\[
\varphi(t, \varepsilon_i, X_{i-1}, \zeta_j) = \sigma(X_{i-1}) \varepsilon_i \left( \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{w_n(X_{j-1})}{f_X(X_{j-1}) \sigma(X_{j-1})} \sin(t_0 \varepsilon_j) \right) (Y_j(t) - E[Y_j(t)]) \\
- \int \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \frac{w_n(x)}{\sigma(x)} E[\sin(t_0 \varepsilon_j)] \left( E[Y_j(t) | X_{j-1} = x] - E[Y_j(t)] \right) dx
\]
and
\[
V_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma(X_{i-1}) \varepsilon_i \int \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \left( \frac{w_n(x) \psi(t, x)}{\sigma(x)} - \frac{w_n(X_{i-1}) \psi(t, X_{i-1})}{\sigma(X_{i-1})} \right) dx
\]
with \( \psi(t, x) = E[Y_i(t) | X_{i-1} = x] - E[Y_i(t)] \). Straightforwardly, we obtain the negligibility of \( V_n \) by considering the expectation
\[
E \left[ \int_{t_0}^{t_2} V_n^2(t) W(t) dt \right]
= \int_{t_0}^{t_2} \int \sigma^2(z) \left( \int \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \left( \frac{w_n(x) \psi(t, x)}{\sigma(x)} - \frac{w_n(z) \psi(t, z)}{\sigma(z)} \right) dx \right)^2 f_X(z) dz W(t) dt.
\]
Note that the inner integral is zero for \( z \notin I_n \). We further separately consider the cases \( z \in K_n = [-a_n + c_n C, a_n - c_n C] \) and \( z \in I_n \setminus K_n \) to obtain
\[
E \left[ \int_{t_0}^{t_2} V_n^2(t) W(t) dt \right]
\leq \int_{t_0}^{t_2} \left[ \int \sigma^2(z) \left( \int \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \left( \frac{\psi(t, x) - \psi(t, z)}{\sigma(x)} + \frac{1}{\sigma(x)} - \frac{1}{\sigma(z)} \right) dx \right)^2 \times f_X(z) I[z \in K_n] dz \right.

\left. + \int \sigma^2(z) \left( \int \frac{1}{c_n} K \left( \frac{X_{j-1} - X_{i-1}}{c_n} \right) \left( \frac{1}{\sigma(x)} + \frac{1}{\sigma(z)} \right) dx \right)^2 f_X(z) I[z \in I_n \setminus K_n] dz \right] W(t) dt
= O((\log n)^{5r}(c_n + c_n^2)) = o(1)
\]
by assumptions (A2) and (A4).

We will now prove \( E[\int U_n^2(t) W(t) dt] = o(1) \). To this end, note that
\[
E \left[ \int U_n^2(t) W(t) dt \right] = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} E \left[ g(\xi_i, \xi_{i_1}, \xi_{i_2}, \xi_{i_3}, \xi_{i_4}) \right], \tag{B5}
\]
where \( \xi_i = (\varepsilon_i, \xi_j) \) and
\[
g(\xi_i, \xi_{i_1}, \xi_{i_2}, \xi_{i_3}, \xi_{i_4}) = \int \varphi(t, \varepsilon_i, X_{i-1}, \xi_j) \varphi(t, \varepsilon_{i_1}, X_{i_1-1}, \xi_{i_2}) \varphi(t, \varepsilon_{i_2}, X_{i_2-1}, \xi_{i_3}) \varphi(t, \varepsilon_{i_3}, X_{i_3-1}, \xi_{i_4}) W(t) dt.
\]

We first consider the case where all indices \( i_1, i_2, i_3, i_4 \) are different. Then, the expectation is zero if either \( i_1 \) or \( i_2 \) is the largest index because \( E[\varepsilon_i] = 0 \) and \( \varepsilon_i \) is independent of \( \varepsilon_j, X_{i-1}, X_{i-2}, \ldots \) (for \( j \neq i \)). All other cases are treated similarly, and thus, we only discuss the case \( i_1 < i_2 < i_3 < i_4 \) in detail. We will apply a version of lemma 2.1 by Sun and
Chiang (1997) for multivariate random variables (see Su and Xiao’s (2008) lemma D.1) in two separate subcases. First, let \( i_2 - i_1 \geq j_1 - i_2 \). Denote by the process \( \xi^*_i, i \in \mathbb{Z} \), an independent copy of \( \xi_i, i \in \mathbb{Z} \), that is, a process with the same distributional properties but independent of the original data. Then, \( E[\xi^*_1, \xi^*_2, \xi^*_j, \xi^*_j] = 0 \) and, for \( \delta > 0 \),

\[
E \left[ \left| g \left( \xi^*_1, \xi^*_2, \xi^*_j, \xi^*_j \right) \right|^{1+\delta} \right] 
\leq k_1 \sup_{x \in [-a_n-c_n, a_n+c_n]} \sigma^{2+2\delta}(x) \sup_{x \in [-a_n-c_n, a_n+c_n]} \sigma^{-2-2\delta}(x) E \left[ |\xi_1|^{1+\delta} \right] 
\times \sup_{x \in [-a_n-c_n, a_n+c_n]} \sigma^{2+2\delta}(x) \sup_{x \in [-a_n-c_n, a_n+c_n]} \sigma^{-2-2\delta}(x) 
\times c_n^{-\delta} E \left[ \left| \frac{1}{c_n} K \left( \frac{X_{j_1} - X_{j_1}}{c_n} \right) \frac{w_n(X_{j_1})}{f_X(X_{j_1})} \right|^{1+\delta} \right]
\]

for some constants \( k_1, k_2 \). This is of order \( O((\log n)^{\overline{r}}c_n^{-2\delta}) \) for \( \overline{r} = 5r(1 + \delta) \) by assumptions (A2)–(A5). An application of the aforementioned inequality gives

\[
\frac{1}{n^2} \sum_{i_1 < i_2 < j_1 < j_2 \atop i_2 - i_1 \geq j_1 - j_2} E \left[ g \left( \xi_{i_1}, \xi_{i_2}, \xi_{j_1}, \xi_{j_2} \right) \right] = O \left( (\log n)^{\overline{r}/(1+\delta)} \right) \frac{1}{n^3 c_n^{2\delta/(1+\delta)}} \sum_{i_1 < j_2 < i_2 < j_1 \atop i_2 - i_1 \geq j_1 - j_2} \left( \alpha(i_2 - i_1) \right)^{\delta/(1+\delta)}
\]

\[
= O \left( \frac{(\log n)^{5r}}{n c_n^{2\delta/(1+\delta)}} \right) \sum_{j=1}^{n} j (\alpha(j))^{\delta/(1+\delta)}
\]

\[
\leq o(1) \sum_{j=1}^{\infty} j^{-\frac{\delta}{1+\delta}} = o(1)
\]

by assumptions (A1), (A3), and (A7). Here, the mixing coefficient \( \alpha \) of \( \xi_i, i \in \mathbb{Z} \), is the same as the mixing coefficient of \( X_i, i \in \mathbb{Z} \), see Fan and Yao (2003). In the subcase \( i_2 - i_1 < j_1 - j_2 \), we apply the same inequality but by considering \( E[\xi^*_1, \xi^*_2, \xi^*_j, \xi^*_j] = 0 \) and obtain

\[
\frac{1}{n^2} \sum_{i_1 < i_2 < j_1 < j_2 \atop i_2 - i_1 \geq j_1 - j_2} E \left[ g \left( \xi_{i_1}, \xi_{i_2}, \xi_{j_1}, \xi_{j_2} \right) \right] = O \left( (\log n)^{\overline{r}/(1+\delta)} \right) \frac{1}{n^3 c_n^{2\delta/(1+\delta)}} \sum_{i_1 < j_2 < i_2 < j_1 \atop i_2 - i_1 \geq j_1 - j_2} \left( \alpha(j_2 - i_2) \right)^{\delta/(1+\delta)}
\]

\[
= O \left( \frac{(\log n)^{5r}}{n c_n^{2\delta/(1+\delta)}} \right) \sum_{j=1}^{n} j (\alpha(j))^{\delta/(1+\delta)} = o(1).
\]

For the case \( i_1 = i_2 \), we exemplarily consider the subcase \( i_2 < j_1 < j_2 \), other subcases are treated similarly. Note that \( E[\xi_{i_1}, \xi_{i_2}, \xi_{j_1}, \xi_{j_2}] = 0 \) by the definition of \( \varphi(\cdot) \), and
Finally, the cases where more than two indices in $i_1, i_2, j_1, j_2$ are equal always lead to negligible terms by direct calculation. For example, consider the term for $j_1 = i_1 \neq j_2 = i_2$ in the sum (B5). Applying assumption (A2), its absolute value can straightforwardly be bounded by $n^{-1}O((\log n)^{sr}(E[|\xi_1|])^2K^2(0)/c_n^2 = o(1)$ by assumption (A7). The remaining terms are treated analogously.