Joint parametric specification checking of conditional mean and volatility in time series models with martingale difference innovations

Kilani Ghoudi a, Naâmane Laïb b and Mohamed Chaouch c

aCollege of Business and Economics, United Arab Emirates University, Al Ain, United Arab Emirates; bLaboratoire AGM, UMR 8088-CNRS, CY Cergy Paris Université, Cergy, France; cProgram of Statistics, Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, Doha, Qatar

ABSTRACT
Using cumulative residual processes, we introduce powerful joint specification tests for conditional mean and variance functions in the context of nonlinear time series with martingale difference innovations. The main challenge comes from the fact that, the cumulative residual process no longer admits a distribution-free limit. To obtain a practical solution one either transforms the process to achieve a distribution-free limit or approximates the non-distribution free limit using numerical or re-sampling techniques. In this paper, the three solutions are considered and compared. The proposed tests have nontrivial power against a class of root-n local alternatives and are suitable when the conditioning set is infinite-dimensional, which allows including more general models such as ARMAX-GARCH with dependent innovations. Numerical results based on simulated and real data show that the powers of tests based on re-sampling or numerical approximation are in general slightly better than those based on martingale transformation.

ARTICLE HISTORY
Received 29 April 2021
Accepted 24 October 2022

KEYWORDS
conditional mean; conditional variance; cumulative residual process; martingale transform; specification tests

1. Introduction
Diagnostic tests are integral part of any modelling exercise. In particular, when fitting a parametric model to a data set, it is essential to check the goodness-of-fit of the postulated model. Several time series models are given by specifying conditional mean and conditional variance functions. Testing the correct specification of these quantities is of a major importance in model validation. A great deal of tests proposed in the literature focuses on testing either the mean function or the volatility function for time series, but usually not both. Note however, testing conditional mean and conditional variance functions separately requires repeated use of significance level of these tests; each one has its own probability of leading to a wrong conclusion. On the other hand, the simultaneous approach takes into account the correlation between these parts. Simultaneous correct specification testing (SCST) approaches are usually used in finance and econometric area.
and constitute vital tools for inference procedures, such as the consistency of the quasi-
maximum likelihood estimator, which depends crucially on the correct joint specification
of the conditional mean and variance functions.

Despite their importance, literature on simultaneous approaches is quite limited. One
cites, however, Escanciano (2008, 2010), who discussed joint tests for parametric form of
the mean and volatility functions. He also argued that, if the mean is misspecified, tests
of volatility functions are usually misleading. Chen and Gao (2011) used a nonparamet-
ric simultaneous approach test for parametric specification of the conditional mean and
variance functions. They argued that using multiple hypotheses leads to an increase of
false rejection of the null hypothesis. This phenomenon of increased false rejection due
to multiple hypotheses is known as the multiplicity effect (see Simes 1986; Benjamini
and Hochberg 1995). A solution for this problem is to construct simultaneous testing proce-
dure, since it takes into account the multiplicity effect while attaining the exact (at least
asymptotically) level of significance. Moreover, a simultaneous test is useful for situations
where we do not have any prior knowledge whether the mean function, the variance
function or both have been correctly specified. If the simultaneous test rejects the null
hypothesis, then one can further investigate the causes of this rejection by looking into
mean and variance specification tests separately.

The literature on SCST for semiparametric/nonparametric regression is mostly con-
cerned with nonlinear models of the form \( Y_i = \mu(X_i) + \sigma(X_i)\epsilon_i \) with independent and
identically distributed (i.i.d.) errors \((\epsilon_i)_{1 \leq i \leq n}\). This assumption turns out to be restrictive in some cases. For instance, Drost and Nijman (1993) pointed out that, the common assumption imposed to GARCH models that rescaled innovations are independent, is disputable since it depends upon the available data frequency. Alternatively, one can suppose that the \( \epsilon_i \)'s form a martingale difference sequence. Such hypothesis is very important in economics theory, because dynamic equilibrium approaches to macroeconomics have imposed martingale restrictions on numerous time series of interest. Further discussion on the martingale hypothesis arising in other contexts of economic theory can be found in, e.g. Durlauf (1991).

In this paper, we aim to construct asymptotically correct simultaneous specification
testing for both the conditional mean \( \mu(\cdot) \) and the conditional variance \( \sigma^2(\cdot) \) functions
for strictly stationary and ergodic time series \( \{ (X_i, Z_i), \ i \in \mathbb{Z} \} \) with martingale difference
innovations \( X_i - \mathbb{E}(X_i|\mathcal{F}_{i-1}) \) without imposing any type of autoregressive model. Here, \( Z \)
represents exogenous variables and \( \mathcal{F}_{i-1} \) is the past sigma-field which will be formally
defined later. The assumption of martingale difference innovations is more general and
allows some dependence structure in the innovations. The framework we are consider-
ing here is then suitable for cases in which the conditioning set is infinite-dimensional.
It may be used for models that do not necessary satisfy Markov property, particularly,
semiparametric models, where the conditional mean and the conditional variance of \( X_i \) given \( \mathcal{F}_{i-1} \) have parametric forms. This includes most processes usually used for mod-
elling financial time series with dependent innovations, such as GARCH, ARMA-GARCH,
exponential and threshold autoregressive processes with GARCH errors. We propose here,
test statistics that are functions of marked cumulative empirical processes. The main chal-
lenge comes from the fact that, the marked cumulative residual processes do not admit a
distribution-free limit in general. To obtain a practical solution, one either transforms the
processes in order to achieve a distribution-free limit or approximates the non-distribution
free limit using numerical or re-sampling techniques. In this work, the three solutions will be discussed. The first approach consists in applying Khmaladze transform to the cumulative residual process. The second one uses multiplier bootstrap re-sampling procedure to approximates the limit distribution of the test statistics, and the third one approximates the limiting distribution of the test statistics using numerical integration. The finite sample performances of the three approaches are compared between each others and also compared to the results of Escanciano (2010)’s procedure.

Khmaladze transform of the cumulative residual process was considered in previous works (cf. Chen, Zheng, and Pan 2015; Stute, Thies, and Zhu 1998). Chen et al. (2015) developed two tests for parametric volatility function of a diffusion model, with i.i.d. innovations, based on Khmaladze’s martingale transformation. Their tests use the structural properties of the diffusion process and do not require the estimation or the specification of the drift function. The setting addressed in our context is more general and does not assume any explicit data-generating model or independent innovations. It may be applied for several nonlinear time series models with martingale innovations. The second added value of this paper is that it compares numerically the three classical procedures used for constructing joint tests based on the cumulative residual process. The Khmaladze transform was applied for each component separately, it will only provide distribution free limit for joint-specification tests in the special case of independent components. The numerical study revealed that it is slightly less performing than re-sampling and numerical approximation techniques.

In the literature, we notice that Escanciano (2008) proposed a class of joint and marginal spectral diagnostic tests for parametric conditional means and variances of time series models. The proposed tests are not distribution-free. To overcome this drawback, the author designed a bootstrap procedure. Based on weighted residual empirical process, Escanciano (2010) constructed asymptotic distribution-free joint specification tests, that can be applied in many financial and economic time series including GARCH and ARMA-GARCH models. The weights in these test statistics are chosen in a way to ensure that the weighted empirical process of residuals admits a distribution free limit. It was shown that the proposed tests generalise those of Wooldridge (1990). The performance of these tests is heavily dependent on the choice of the weights. Notice that our work has similar objectives to that of Escanciano (2010), the solution that we propose does not require the choice of weight functions and is shown to perform better in general than that of Escanciano (2010).

Related work on Goodness-of-fit tests for parametric and semiparametric hypotheses of the regression function has been considered widely in the literature with emphasis on i.i.d innovations. For instance, Stute (1997) presented nonparametric full-model checks based on the limiting law of the residual marked process. The reader is referred to González-Manteiga and Crujeiras (2013) for a survey on the topic. Escanciano, Pardo-Fernández, and Van Keilegom (2018) discussed a general methodology for constructing nonparametric/semiparametric asymptotically distribution-free tests about regression models for possibly dependent data. Similar study investigated the autoregressive function in time series models (see, Diebolt 1990; Koul and Stute 1999; Laïb 1999; McKean and Zhang 1994). In the context of time series with martingale difference innovations, Stute, Quindimil, Manteiga, and Kou (2006) provided nonparametric tests based on residual cusums for testing the autoregressive function in higher-order time series models. In the same framework, Escanciano and Mayoral (2010) proposed data-driven asymptotically
distribution-free tests for testing the martingale difference hypothesis of possibly nonlinear time series.

Hypotheses testing of the conditional variance function in regression models is investigated by many authors in the past. For instance, in nonparametric regression models, Wang and Zhou (2005) considered a nonparametric diagnostic test for checking the constancy of the conditional variance function, Dette, Neumeyer, and Keilegom (2007) proposed a test procedure for testing parametric forms of the conditional variance. Koul and Song (2010) discussed the problem of fitting a parametric model to the conditional variance function in a class of heteroscedastic regression models. Their test is based on the supremum of the Khmaladze type martingale transformation of certain partial sum process of calibrated squared residuals. The proposed statistical test is shown to be consistent against a large class of fixed alternatives and to have nontrivial asymptotic power against a class of nonparametric local alternatives. Recently, Pardo-Fernández, Jiménez-Gamero, and El Ghouch (2015) proposed several nonparametric statistical tests for checking whether the conditional variances are equal in $k$ location-scale regression models. Their procedure is based on the comparison of the error distributions under the null hypothesis of equality of variances functions. Polonik and Yao (2008) proposed two tests for testing multivariate volatility functions using minimum volume sets and inverse regression. Their tests are based on cumulative sums coupled with either minimum volume sets or inverse regression ideas.

Tests of conditional variance functions in time series context were also previously considered in the literature. In particular, Auestad and Tjøstheim (1990), focused on kernel estimate of the one step lagged conditional mean and variance functions for the purpose of identifying common linear models such as threshold and exponential autoregressive. Diebolt (1990) established the consistency of regressogram type estimators of the conditional mean and conditional variance functions. He deduced nonparametric Goodness-of-fit tests for a known form of these functions. Chen and An (1997) proposed a Kolmogorov–Smirnov type statistic to test the homoscedasticity hypothesis when the observations are assumed to be strongly mixing. Their test only uses a subsample which induces a loss of information and power. Ngatchou-Wandji (2002) presented a procedure, based on marked empirical process, for testing the Goodness-of-fit of the conditional variance function of a Markov parametric model of order one.

In the framework of time series with martingale difference innovations, Laïb and Chebana (2011) considered a class of nonlinear semiparametric models and established the local asymptotic normality for cumulative residual process. They also derived an efficient simultaneous test for testing the conditional mean and the conditional variance functions. Laïb and Louani (2002) provided a test of conditional homoscedasticity hypothesis of the one-step forecast error in the context of first-order AR-ARCH model. Their work was extended by Laïb (2003) for the context of time series with martingale difference innovations. The author proposed a test based on the cumulative residual process for testing homoscedasticity when the innovations are independent of the past information.

The rest of the paper is organised as follows: Section 2 defines the problem and states some preliminaries results. Section 3 defines and establishes the asymptotic behaviour of the martingale transformation applied to the cumulative residual process. Section 4 introduces the marginals as well as the joint parametric specification Cramér von-Mises-type tests statistics for the conditional mean and conditional variance functions. Section 4.1
presents test statistics based on the martingale transform. A numerical approximation procedure for the asymptotic distribution of the test statistics based on the original process is given in Section 4.2. A re-sampling algorithm for test statistics based on the original cumulative residual process is detailed in Section 4.3. A comparison between these statistical procedures, via simulations, is outlined in Section 5 and an application to real data is given in Section 6. A conclusion summarising our findings is given in Section 7. All proofs are provided in Appendix 1 and a discussion on the implementation of tests based on Khmalalze transform is detailed in Appendix 2.

2. Assumptions and main results

Consider a time series with conditional mean \( \mu(\cdot) \) and conditional variance \( \sigma^2(\cdot) \), the goal is to test whether \( \mu \) and \( \sigma^2 \) admit specific parametric forms. More formally, let \( \{(X_i, Z_i), i \in \mathbb{Z}\} \) be strictly stationary ergodic defined on the probability space \((\Omega, \mathcal{A}, \mathbb{P})\), where \( Z \) represents exogenous random variables. The random variables \( X_i \)'s are real-valued with common continuous distribution function \( F \). For each \( i \geq 1 \), let \( I_{i-1} = (X_{i-1}, X_{i-2}, \ldots, Z_i, Z_{i-1}, \ldots) \) denote the past information at time \( i \). We let \( \mathcal{F}_i = \sigma(I_0, I_1, \ldots, I_i) \) denotes \( \sigma \)-field generated by \( I_0, \ldots, I_i \). The purpose is to verify if the conditional mean \( \mu(I_{i-1}) := \mathbb{E}(X_i | \mathcal{F}_{i-1}) \) and the conditional variance \( \sigma^2(I_{i-1}) := \text{Var}(X_i | \mathcal{F}_{i-1}) \) of \( X_i \) given \( \mathcal{F}_{i-1} \) satisfy, respectively, the following relations \( \mu(I_{i-1}) = m(\theta_0, I_{i-1}) \) almost surely (a.s.) and \( \sigma^2(I_{i-1}) = \sigma^2(\theta_0, I_{i-1}) \) a.s., where \( m(\theta, \cdot) \) and \( \sigma^2(\theta, \cdot) \) are \( \mathcal{F}_{i-1} \)-measurable known parametric functions that depend on a finite dimensional vector of parameters \( \theta \in \mathbb{R}^d \) for \( d \geq 1 \), and assumed to be finite with probability one. More precisely, we are interested in examining composite hypotheses stating that

\[
\begin{align*}
(H_0) : \mu(I_{i-1}) = m(\theta_0, I_{i-1}) & \quad \text{a.s. and } \sigma^2(I_{i-1}) = \sigma^2(\theta_0, I_{i-1}) \quad \text{a.s. versus} \\
(H_1) : \mu(I_{i-1}) \neq m(\theta, I_{i-1}) & \quad \text{or } \sigma^2(I_{i-1}) \\
& \neq \sigma^2(\theta, I_{i-1}) \text{ with positive probability } \forall \theta \in \mathbb{R}^d.
\end{align*}
\]

We assume throughout this manuscript that the true value of \( \theta \), denoted \( \theta_0 \), belongs to the interior of some compact subset \( \Xi \subset \mathbb{R}^d \). We also assume that the sequence of innovations \( \{X_i - m(\theta_0, I_{i-1}) : i \geq 0\} \) is a sequence of martingale differences with respect to \( \mathcal{F}_i \), that is \( X_i - m(\theta_0, I_{i-1}) \) is \( \mathcal{F}_i \)-measurable and \( \mathbb{E}[(X_i - m(\theta_0, I_{i-1})) | \mathcal{F}_{i-1}] = 0 \) a.s. This condition combined with the fact that the set of information \( I_i \) could be infinite dimensional allow us to consider non markovian processes such as the ARMA-GARCH process. In practice the set \( I_i \) is not observable and may be estimated, see Remark 2.3 and Escanciano (2010) for more discussion.

To test the specification of the conditional mean function we will use \( W^1_\theta(X, I) := X - m(\theta, I) \) and to test the conditional variance function we will use \( W^2_\theta(X, I) := (X - m(\theta, I))^2 - \sigma^2(\theta, I) \). Following (Escanciano 2007a, 2008; Koul and Stute 1999; Laïb 2003; Ngatchou-Wandji 2002), we introduce the following cumulative empirical residuals processes

\[
\mathbb{D}^k_n(x) = n^{-1/2} \sum_{i=1}^n W^k_{\theta_0}(X_i, I_{i-1}) \mathbb{I}\{X_{i-1} \leq x\}, \quad x \in \mathbb{R} \quad \text{and} \quad k = 1, 2,
\]
\[
\hat{D}_n^k(x) = n^{-1/2} \sum_{i=1}^{n} W_{\theta_n}^k(X_i, I_{i-1}) \mathbb{I}\{X_{i-1} \leq x\}, \quad x \in \mathbb{R}, \quad \text{and} \quad k = 1, 2,
\]

where \(\mathbb{I}(A)\) is the indicator function of the set \(A\), \(\theta_n\) is consistent estimator of \(\theta\). Note that, under \(H_0\), \(\mathbb{E}\{W_{\theta_0}^k(X_i, I_{i-1})|\mathcal{F}_{i-1}\} = 0\) a.s. for \(k = 1, 2\). To test specification of the conditional mean and variance jointly, we introduce the bivariate processes \(D_n\) and \(\hat{D}_n\) defined by \(D_n(x) := \left(\hat{D}_n^1(x), \hat{D}_n^2(x)\right)^\top\) and \(\hat{D}_n(x) := \left(\hat{D}_n^1(x), \hat{D}_n^2(x)\right)^\top\), where the script \(B^\top\) stands for the transpose of the matrix \(B\). Note that letting \(W_\theta^k(X, I) := \{W_\theta^1(X, I), W_\theta^2(X, I)\}^\top\) one sees that

\[
\mathbb{D}_n(x) = n^{-1/2} \sum_{i=1}^{n} W_{\theta_0}(X_i, I_{i-1}) \mathbb{I}\{X_{i-1} \leq x\} \quad \text{and} \quad \hat{D}_n(x) = n^{-1/2} \sum_{i=1}^{n} W_{\hat{\theta}_n}(X_i, I_{i-1}) \mathbb{I}\{X_{i-1} \leq x\}.
\]

**Remark 2.1:** Note that \(\mathbb{E}\{W_{\theta_0}^k(X_i, I_{i-1})|\mathcal{F}_{i-1}\} = 0\) if and only if the null hypothesis is true. However, in our construction of the test we only multiplied \(W_{\theta_n}(X_i, I_{i-1})\) by the indicator function \(\mathbb{I}\{X_{i-1} \leq x\}\). It follows from ergodicity and stationarity that \(n^{-1/2}D_n(x)\) converges a.s. to \(\mathbb{E}\{W_{\theta_0}(X_i, I_{i-1})\mathbb{I}\{X_{i-1} \leq x\}\}\), which is not equal to zero whenever \(\mathbb{E}\{W_{\theta_0}(X_i, I_{i-1})\mathbb{I}\{X_{i-1} \leq x\}\} \neq 0\) with positive probability. Note that there might be some alternatives that satisfy \(\mathbb{E}\{W_{\theta_0}(X_i, I_{i-1})|\mathcal{F}_{i-1}\} \neq 0\) while \(\mathbb{E}\{W_{\theta_0}(X_i, I_{i-1})|X_{i-1}\} = 0\). In particular, tests based on norm of \(D_n\) will have power going to one to detect any alternative satisfying \(\mathbb{E}\{W_{\theta_0}(X_i, I_{i-1})|X_{i-1}\} \neq 0\) with positive probability. The only alternatives that the above mentioned tests fail to detect are those satisfying \(\mathbb{E}\{W_{\theta_0}(X_i, I_{i-1})|X_{i-1}\} = 0\) a.s. Thus, tests based on the supremum or the integral of the square of the process \(D_n\) will be powerful in detecting any alternative except those satisfying \(\mathbb{E}\{W_{\theta_0}(X_i, I_{i-1})|X_{i-1}\} = 0\) a.s.

**Remark 2.2:** Comparing the empirical process \(\hat{D}_n^k\) with that used in Escanciano (2010), one notice that \(\hat{D}_n^k\) is a marked empirical process based on \(X_{i-1}\) with marks/weights given by \(W_{\hat{\theta}_n}^k\) while the process used in Escanciano (2010) is based on the residuals \(\hat{\epsilon}_i\), with weights carefully estimated from data.

**Remark 2.3:** As pointed in Escanciano (2010), the past information \(I_{i-1}\) is not completely observable and needs to be estimated by \(\hat{I}_{i-1}\). Such estimation, essentially, involves replacing the unknown initial state \(I_0\) (for \(t = 0\)) by some quantity \(\hat{I}_0\). This is commonly done in the context of parameter estimation. For all commonly used time series models it is established that parameter estimation is asymptotically not affected by the choice of initial states. In our context, one must ensure that replacing \(I_{i-1}\) by \(\hat{I}_{i-1}\) does not affect the asymptotic behaviour of the process \(\hat{D}_n\). One easily verifies that if for \(k = 1, 2,\)

\[
n^{-1/2} \sum_{i=1}^{n} \mathbb{E}\left\{\sup_{\hat{\theta} \in \mathbb{R}} \left|W_{\hat{\theta}}^k(X_i, \hat{I}_{i-1}) - W_{\theta_0}^k(X_i, I_{i-1})\right|\right\} = o(1),
\]

then the asymptotic behaviour of \(\hat{D}_n\) is not altered when \(I_{i-1}\) is replaced by \(\hat{I}_{i-1}\). We will assume that such condition holds and will use \(I_{i-1}\) in the definition of \(\hat{D}_n\) throughout the
manuscript. The discussion in Escanciano (2010) shows that this condition holds true for ARMA-GARCH models in particular. Hafner and Kyriakopoulou (2021) showed that the effect of initial values on the parameters’ estimation vanishes for exponential-type GARCH models such as EGARCH or Log-GARCH. Using similar steps one verifies condition (1) holds for exponential-type GARCH models.

The limiting laws of $D_n^1$ and $\hat{D}_n^1$ are given in Koul and Stute (1999) and Escanciano (2007a). The limiting behaviour of $D_n^2$ is given in Laïb (2003), who also obtained the limit law of $D_n^2$ in the special case $\sigma^2(\cdot) = \zeta^2$ with $\zeta^2 \in (0, \infty)$. The next section studies the asymptotic properties of the bivariate process $\hat{D}_n$. 

### 2.1. Limiting law of the process $\hat{D}_n$ under the null hypothesis

The following notations are used in the rest of the paper. $\|v\|$ denotes the Euclidian norm of the vector $v$ and for any matrix $A$, $\|A\| = \sup_{\|v\|=1} \|Av\|$ is the associated matrix norm of the matrix $A$. For any bounded function $f$, let $\|f\| = \sup_{x} |f(x)|$ and for any $(x, y) \in \mathbb{R}^2$, set $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$.

The asymptotic results are stated under the following assumptions:

**Assumption A1:** For each $k = 1, 2$,

1. $\mathbb{E}(\|W_{\theta_0}^k (X_i, I_{i-1})\|^2) < \infty$.
2. $\lim_{n \to \infty} \mathbb{E}\{\sum_{i=1}^{n} |W_{\theta_0}^k (X_i, I_{i-1})| > \delta \sqrt{n}\} = 0$ for any real $\delta > 0$.
3. $K_k(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(\|W_{\theta_0}^k (X_i, I_{i-1})\| \|X_{i-1} \leq x\|)$ is non-decreasing continuous function of $x$.
4. $\mathbb{E}(\|W_{\theta_0}^k (X_i, I_{i-1})\| \|X_{i-1} \leq y\| |\mathcal{F}_{i-2}) = C_{k,i} |K_k(y) - K_k(x)|$ a.s. such that $\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}|C_{k,i}| = O(1)$.

**Assumption A2:** Under $H_0$, the estimator $\hat{\theta}_n$ of $\theta_0$ verifies

$$n^{1/2}(\hat{\theta}_n - \theta_0) = n^{-1/2} \sum_{i=1}^{n} \phi^*(X_i, I_{i-1}, \theta_0) + o_p(1),$$

where $\phi^*$ is an $\mathbb{R}^d$-valued measurable function satisfying $\mathbb{E}(\phi^*(X_i, I_{i-1}, \theta_0) | \mathcal{F}_{i-1}) = 0$ a.s. and

$$\Sigma_0 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left( \phi^*(X_i, I_{i-1}, \theta_0) \phi^*(X_i, I_{i-1}, \theta_0) ^\top \right)$$

exists and is positive definite.

**Assumption A3:** For $k = 1, 2$, let $\hat{W}_{\theta_0}^k (X_i, I_{i-1})$ denote the gradient of $W_{\theta_0}^k (X_i, I_{i-1})$ with respect to the components of $\theta$ evaluated at a fixed point $\theta^*$.

Assume that $\mathbb{E}\|\hat{W}_{\theta_0}^k (X_i, I_{i-1})\| \leq C < \infty$ and

$$\|W_{\theta_0}^k (X_i, I_{i-1}) - \hat{W}_{\theta_0}^k (X_i, I_{i-1}) - (\theta - \theta_0) ^\top \hat{W}_{\theta_0}^k (X_i, I_{i-1})\|$$
where $M_1$ and $\lambda_1$ are positive functions satisfying $\mathbb{E}(M_1(X_i, I_{i-1})) \leq C < \infty$ and $\lambda_1(t)$ goes to zero as $t \to 0$.

Assumption (A1) is an adaptation of conditions (A–D) of Theorem 1 of Escanciano (2007b) to the context of stationary and ergodic sequence. As argued in Escanciano (2007b) these are among the weakest conditions to ensure the weak convergence of marked empirical processes. Assumptions (A2) and (A3) are commonly used to ensure the convergence of $\sqrt{n}(\theta_n - \theta_0)$ and to validate the expansion of the process $\hat{D}_n$. These assumptions hold for most commonly used models and estimation procedures (see, for example, Koul and Stute 1999; Escanciano 2007a, 2010, for discussion and details).

From now on let $D(\mathbb{R})$ denote the space of càdlàg functions. For $x, y \in \mathbb{R}$ define

$$K(x, y) = \text{Cov}(D_n(x), D_n(y)) = \begin{pmatrix} K_1(x \wedge y) & K_{1,2}(x \wedge y) \\ K_{1,2}(x \wedge y) & K_2(x \wedge y) \end{pmatrix},$$

where $K_1$ and $K_2$ are given in Assumption (A1) and for $x \in \mathbb{R}$

$$K_{1,2}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ W^1_{\theta_0}(X_i, I_{i-1}) W^2_{\theta_0}(X_i, I_{i-1}) \mathbb{I}(X_{i-1} \leq x) \right].$$

Let $D$ denote the Gaussian process with covariance function $K$ defined above. The following result summarises the weak convergence of the processes $\hat{D}_n$ and $\hat{D}_1$.

**Theorem 2.1:** If Assumption (A1) holds then, under $H_0$, $D_n$ converges weakly to $D$. Moreover, if Assumptions (A1)–(A3) hold true, then, under $H_0$, $\hat{D}_n$ converges weakly to a centred Gaussian process $\hat{D}$ given by

$$\hat{D}(x) = D(x) - \Gamma^\top_{\theta_0}(x) \Theta,$$

where $\Theta$ is a centred multivariate Gaussian random variable with covariance matrix $\Sigma_0$, and $\Gamma_{\theta_0}(x) = (\Gamma^1_{\theta_0}(x), \Gamma^2_{\theta_0}(x))$ with

$$\Gamma^1_{\theta_0}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ \hat{m}(\theta_0, I_{i-1}) \mathbb{I}[X_{i-1} \leq x] \right] \quad \text{and}$$

$$\Gamma^2_{\theta_0}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left\{ \hat{s}^2(\theta_0, I_{i-1}) \mathbb{I}[X_{i-1} \leq x] \right\}.$$

The covariance function $K$ of $\hat{D}$ is given by

$$K(x, y) = K(x, y) - \Gamma^\top_{\theta_0}(x) G(y) - G^\top(x) \Gamma_{\theta_0}(y) + \Gamma^\top_{\theta_0}(x) \Theta_0 \Theta_0 G_{\theta_0}(y),$$

where $G(x) = \text{Cov}(\Theta, D) = (G^1(x), G^2(x))$ with

$$G^k(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ W^k_{\theta_0}(X_i, I_{i-1}) \phi^*(X_i, I_{i-1}, \theta_0) \mathbb{I}(X_{i-1} \leq x) \right] \quad \text{for } k = 1, 2.$$
Remark 2.4: The limiting covariance function $K$ is in general complicated, therefore classical statistics based on the process $\hat D_n$ do not admit distribution-free limits. To overcome this, one usually uses one of the following approaches. The first approach consists in transforming the process $\hat D_n$ in such a way to achieve a distribution-free limit and then using the transformed process to define test statistics. While the second one uses $\hat D_n$ to construct test statistics and then adopts either a numerical approximation or a re-sampling technique to estimate the non-distribution free limit. Both approaches will be discussed and compared in the rest of this manuscript. The transformation is discussed in Section 3 while the numerical and re-sampling approximations are outlined in Section 4.

2.2. Limiting law of the process $\hat D_n$ under local alternatives

In this section we establish the limiting behaviour of the process $\hat D_n$ under local alternatives $H_A$ defined as follows.

$$H_A : \mu(I_{i-1}) = m(\theta_0, I_{i-1}) + a_1(I_{i-1})/\sqrt{n} \quad \text{and} \quad \sigma^2(I_{i-1}) = \sigma^2(\theta_0, I_{i-1}) + a_2(I_{i-1})/\sqrt{n},$$

where $a_1$ is non-zero function and $a_2$ is a positive function. These two functions specify the direction of departure from the null hypothesis. We assume that $E|a_k(I_{i-1})| < \infty$ for $k = 1, 2$. The following assumption is needed to establish the limiting behaviour of $\hat D_n$ under $H_A$.

Assumption L1: Under $H_A$, the estimator $\theta_n$ of $\theta_0$ satisfies

$$n^{1/2}(\theta_n - \theta_0) = n^{-1/2} \sum_{i=1}^n \phi^*(X_i, I_{i-1}, \theta_0) + \xi_A + o_p(1),$$

where $\xi_A \in \mathbb{R}^d$ is a random/nonrandom vector and $\phi^*$ is as in Assumption (A2).

The following Theorem shows that our tests are able to detect local alternatives of the type described by $H_A$.

Theorem 2.2: Under $H_A$, if Assumptions (A1), (A3) and (L1) hold, the process $\hat D_n$ converges weakly to the process $\hat D + \Psi_A(x) - \Gamma_{\theta_0}^\top(x)\xi_A$, where

$$\Psi_A(x) = E\left\{ \left( \begin{array}{c} a_1(I_{i-1}) \\ a_2(I_{i-1}) \end{array} \right) I[X_{i-1} \leq x] \right\}.$$

Remark 2.5: Note that the local alternative hypotheses $H_A$ differ from the null in a fixed direction determined by the vector of functions $v_i = (a_1(I_{i-1}), a_2(I_{i-1}))^\top$, and that the difference between the null hypothesis $H_0$ and $H_A$ goes to zero as $n \to +\infty$. Theorem 2.2 ensures that such local alternative will be detected whenever the function $\Psi_A(x)$ is not identically zero. As pointed out in Remark 2.1, the only non-detected local alternatives are those where function $\Psi_A(x) = 0$ a.s., that is, $E[a_1(I_{i-1})|X_{i-1}] = 0$ a.s. and $E[a_2(I_{i-1})|X_{i-1}] = 0$ a.s.
3. Khmaladze transform of the process \( \hat{D}_n \)

This section presents a martingale transformation of the Khmaladze type that will be applied to the process \( \hat{D}_n \), in order to achieve a distribution-free limit for classical test statistics. It is worth mentioning that the Khmaladze transformation will ensure that the limit of the transformed process is a Brownian motion.

To define such transformation, assume that the functions \( \Gamma^k_{\theta_0}(x) \), for \( k = 1, 2 \), can be written as

\[
\Gamma^k_{\theta_0}(x) = \int_{-\infty}^{x} g_k(t) K_k(dt),
\]

where \( g_k(t) = (g_{1,k}(t), \ldots, g_{d,k}(t))^T \). For \( k = 1, 2 \) let \( A_k \) be the matrix defined by \( A_k(x) = \int_{x}^{\infty} g_k(u) g_k^T(u) K_k(du) \). In addition, assume that there exists \( x_0 < \infty \) such \( A_k(x_0) \) is non-singular. Then, it follows from the definition of \( A_k(x) \) that \( A_k(x) - A_k(x_0) \) is non-negative definite for all \( x \leq x_0 \). This implies that \( A_k(x) \) is invertible for all \( x \leq x_0 \) whenever \( A_k(x_0) \) is invertible.

In order to illustrate the calculation of the functions \( g_k \), let us consider the following example.

**Example 3.1:** Consider an ARCH(1) model given by

\[
X_t = a_0 + \sqrt{w_0 + \alpha(X_{t-1} - a_0)^2}u_t,
\]

where \( \omega_0 > 0 \), \( \alpha \geq 0 \) and \( (u_t)_{1 \leq t \leq n} \) are i.i.d random variables with mean zero and variance one and where \( u_t \) is independent of \( X_{t-1} \). In this example \( \theta = (a_0, w_0, \alpha)^T \), \( m(\theta, I_{t-1}) = a_0 \) and \( \sigma^2(\theta, I_{t-1}) = w_0 + \alpha(X_{t-1} - a_0)^2 \). One sees \( m(\theta, I_{t-1}) = (1, 0, 0)^T \) and that \( \sigma^2(\theta, I_{t-1}) = (-2\alpha(X_{t-1} - a_0), 1, (X_{t-1} - a_0)^2)^T \).

Thus

\[
\Gamma^1_{\theta}(x) = \begin{pmatrix} F(x) \\ 0 \\ 0 \end{pmatrix}, \quad \Gamma^2_{\theta}(x) = \begin{pmatrix} -2\alpha \int_{-\infty}^{x} (y - a_0) dF(y) \\ F(x) \\ \int_{-\infty}^{x} (y - a_0)^2 dF(y) \end{pmatrix},
\]

\[
K_1(x) = \mathbb{E}[(X_t - a_0)^2 1_{\{X_{t-1} \leq x\}}] = \mathbb{E}[(w_0 + \alpha(X_{t-1} - a_0)^2) u_t^2 1_{\{X_{t-1} \leq x\}}] = \int_{-\infty}^{x} (w_0 + \alpha(y - a_0)^2)^2 dF(y) \]

and \( K_2(x) = \mathbb{E}[(u_t^2 - 1)^2] \int_{-\infty}^{x} (w_0 + \alpha(y - a_0)^2)^2 dF(y) \).

Using these expressions, one finds

\[
dK_1(x) = (w_0 + \alpha(x - a_0)^2)^2 dF(x) \quad \text{and} \quad dK_2(x) = [\mathbb{E}(u_t^4) - 1](w_0 + \alpha(x - a_0)^2)^2 dF(x).
\]

Hence, the functions \( g_k \) in expression (3) are given by

\[
g_1(x) = \begin{pmatrix} 1 \\ \frac{1}{w_0 + \alpha(x - a_0)^2} \\ 0 \end{pmatrix} \quad \text{and} \quad g_2(x) = \begin{pmatrix} -2\alpha(x - a_0) \\ (w_0 + \alpha(x - a_0)^2) \frac{1}{\mathbb{E}(u_t^4) - 1} \\ (w_0 + \alpha(x - a_0)^2)^2 \frac{(w_0 + \alpha(x - a_0)^2)(x - a_0)^2}{(x - a_0)^2} \end{pmatrix}.
\]

Details on how to obtain the functions \( g_k \), in general, are given in Appendix 2.
Next, following the concept in Khmaladze (1988), for any function \( f \) consider the linear transformations \( T^k(f) \) for \( k = 1, 2 \) given by

\[
(T^k f)(x) = f(x) - \int_{-\infty}^{x} (g_k(y))^\top A_k^{-1}(y) \left[ \int_y^\infty g_k(z) f(z) \, dz \right] K_k(dy),
\]

for all \( x \leq x_0 \). The Khmaladze transform of the cumulative residual process considered here is an application of the transformation \( T^1 \) and \( T^2 \) to the processes \( \hat{D}^1_n \) and \( \hat{D}^2_n \). It is formally defined by

\[
T(\hat{D}_n(x)) = \begin{pmatrix} T^1(\hat{D}^1_n(x)) \\ T^2(\hat{D}^2_n(x)) \end{pmatrix} = \begin{pmatrix} \hat{D}^1_n(x) - \int_{-\infty}^{x} (g_1(y))^\top A_1^{-1}(y) \left[ \int_y^\infty g_1(z) \hat{D}^1_n(dz) \right] K_1(dy) \\ \hat{D}^2_n(x) - \int_{-\infty}^{x} (g_2(y))^\top A_2^{-1}(y) \left[ \int_y^\infty g_2(z) \hat{D}^2_n(dz) \right] K_2(dy) \end{pmatrix}.
\]

Easy manipulations show that, for \( k = 1, 2 \),

\[
T^k(\hat{D}^k_n(x)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W^k_{\theta_n}(X_i, I_{i-1}) \left[ I\{X_{i-1} \leq x\} - \int_{-\infty}^{\min(x,X_{i-1})} (g_k(y))^\top A_k^{-1}(y) K_k(dy) g_k(X_{i-1}) \right].
\]

Observe also that \( T^k(\hat{D}^k_n)(x) \) cannot be used in practice to build test statistics since it depends on several unknown quantities, namely, \( g_k \) and \( K_k \). Replacing \( g_k \) and \( K_k \) by their consistent estimates \( \hat{g}_k \) and \( \hat{K}_k \), respectively, one obtains an empirical version of \( T^k(\hat{D}^k_n) \) defined as follows

\[
T^k_n(\hat{D}^k_n(x)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W^k_{\theta_n}(X_i, I_{i-1}) \left[ I\{X_{i-1} \leq x\} - \int_{-\infty}^{\min(x,X_{i-1})} (\hat{g}_k(y))^\top \hat{A}_k^{-1}(y) \hat{K}_k(dy) \hat{g}_k(X_{i-1}) \right],
\]

(4)

where \( \hat{A}_k(x) = \int_{x}^{\infty} \hat{g}_k(u) \hat{g}_k^\top(u) \hat{K}_k(du) \). The following assumptions are needed to ensure the convergence of \( T_n(\hat{D}_n) = (T^1_n(\hat{D}^1_n), T^2_n(\hat{D}^2_n))^\top \).

**Assumption K1:** Assume there exists \( x_0 < \infty \) such that \( A_k(x_0) \) is invertible for \( k = 1, 2 \).

**Assumption K2:** Assume that

\[
\mathbb{E}\|W^k_{\theta_n}(X_i, I_{i-1})g_k(X_{i-1})\|, \mathbb{E}\{M_1(X_i, I_{i-1})\|g_k(X_{i-1})\|\}, \mathbb{E}\{M_2(X_0)\|g(X_0)\|\}, \mathbb{E}\{\|g(X)\|\}, \mathbb{E}\{\|g(X)g(X)^\top\| W^k_{\theta_n}(X_i, I_{i-1})^2\} \text{ and } \mathbb{E}\{\|g(X)g(X)^\top\|\} \text{ are all finite}.
\]
Assumption K3: Assume that $||\hat{g}_k - g_k||$ converges to zero in probability, and that sup$_{y \in \mathbb{R}} E \left( \int_y^\infty (g_k(z) - \hat{g}_k(z)) D_n(dz) \right)$ converges to zero in probability.

Assumption K4: For each $k = 1, 2$ and for each $j = 1, \ldots, d$,

1. $\mathbb{E}(|W^k_{i-1}(X_i, I_i, I_{i-1}) g_{j,k}(X_{i-1})|^2) < \infty$.
2. For any $\delta > 0$, $\mathbb{E}([W^k_{i-1}(X_i, I_i, I_{i-1}) g_{j,k}(X_{i-1})]^2 \mathbb{1}[|W^k_{i-1}(X_i, I_i, I_{i-1}) g_{j,k}(X_{i-1})| > \delta \sqrt{n}])$ converges to zero as $n \to \infty$.
3. $\tilde{K}_{j,k}(x) = \mathbb{E}(W^k_{i-1}(X_i, I_i, I_{i-1}) g_{j,k}(X_{i-1})^2 \mathbb{1}[X_{i-1} \leq x])$ is non-decreasing continuous function of $x$.
4. $\mathbb{E}(W^k_{i-1}(X_i, I_i, I_{i-1}) g_{j,k}(X_{i-1})^2 \mathbb{1}[x \leq X_{i-1} \leq y]\mid \mathcal{F}_{i-2}) = C^j_{k,i} \tilde{K}_{j,k}(y) - \tilde{K}_{j,k}(x)$ such that $\frac{1}{n} \sum_{i=1} n |C^j_{k,i}| = O(1)$.

The next result establishes the weak convergence of $T_n(\hat{D}_n)$.

**Theorem 3.2:** If Assumptions (A1)–(A4) and (K1)–(K4) are satisfied then, for $k = 1, 2$,

$$T^n_{k}(\hat{D}_n^k) \sim \mathbb{D}^k \text{ in } \mathbb{D}((\infty, x_0]),$$

where $\sim$ denotes convergence in distribution and $\mathbb{D}^k = \mathbb{W}^k \circ K_k$ is the $k$th component in the process $\mathbb{D}$ defined in Theorem 2.1.

**Remark 3.1:** Note that the process $\mathbb{D}$ has marginal $\mathbb{W}^1 \circ K_1$ and $\mathbb{W}^2 \circ K_2$ where $\mathbb{W}^1$ and $\mathbb{W}^2$ are Brownian motions. In general $\mathbb{W}^1$ and $\mathbb{W}^2$ are not independent. The distribution of $\int \|\mathbb{D}\|^2$ will only be distribution free if $\mathbb{W}^1$ and $\mathbb{W}^2$ are independent. This defeats the purpose of the martingale transformation in general. Though the above results clearly show that the transformation provides a distribution-free limit for testing the mean or the variance function separately. The transformation is only useful for joint testing of conditional mean and variance functions in the case of independent components of the bivariate process $\mathbb{D}$. This was the case in the work of Chen et al. (2015). In general, the components $\mathbb{D}^1$ and $\mathbb{D}^2$ will be independent if $K_{12} = 0$ which translates to a condition on the third moment of $\epsilon_i = X_i - m(\theta, I_{i-1})$. Therefore, if $\mathbb{E}(\epsilon_i^3 \mid \mathcal{F}_{i-1}) = 0$ a.s., then the components $\mathbb{D}^1$ and $\mathbb{D}^2$ are independent. This holds true for all time series with i.i.d. symmetric errors or with zero third conditional moments errors.

To make the transformation work in the case of dependent components of the process $\mathbb{D}$, requires the development of Khmaladze type transformation of pair/vector of processes to pair/vector of independent Wiener processes. Though this is very interesting research problem, it is beyond the scope of this paper and will be considered in future research. Here we will provide alternatives based on numerical approximations and resampling techniques that allow us to carry the joint test in the case of dependent components.

**Remark 3.2:** Note that the conditions imposed on the estimate $g_n$ are similar to those used in Bai (2003) and Bai and Ng (2001). Note also that the $||\hat{g}_k - g||$ converges to zero in probability can be weakened to $\int \|\hat{g}_k(x) - g(x)\|^2 F(dx)$ converges to zero in probability. But for simplicity, the proof is presented with the stronger assumption. The second condition on $\hat{g}_k$, in Assumption (K3), is usually verified using the structure of the estimator and the nature of the process $\mathbb{D}_n$, for details see the discussion in Bai (2003).
4. Test statistics

Test statistics for the hypothesis $H_0$ are obtained by considering continuous functional $G$ of the process $\hat{D}_n$ or of its Khmaladze transform $T_n(\hat{D}_n)$. That is, given a continuous functional $G$, a test statistic can be obtained using $G(\hat{D}_n)$ or $G(T_n(\hat{D}_n))$. The continuous mapping theorem provides the asymptotic behaviour of $G(\hat{D}_n)$ and $G(T_n(\hat{D}_n))$. In practice the most commonly used functionals are of Cramér-von-Mises or Kolmogorov–Smirnov types. For simplicity, the rest of this paper only focuses on Cramér-von-Mises type statistics, but the approaches discussed apply to any continuous functional. To be specific, the following marginal Cramér-von-Mises test statistics will be used

$$S^k_n = \int_{-\infty}^{\infty} \left[ \hat{G}^k_n(x) \right]^2 F_n(dx)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \{ 1 - F_n(X_{i-1} \lor X_{j-1}) \} W_{\hat{\theta}_n}^k(X_i, I_{i-1}) W_{\hat{\theta}_n}^k(X_j, I_{j-1})$$

for $k = 1, 2$, where $F_n$ denotes the empirical distribution function of the $X_i$’s. Clearly, the statistics $\{S^k_n\}_{k=1,2}$ converge in distribution to $S^k := \int_{-\infty}^{\infty} \left[ \hat{G}^k(x) \right]^2 F(dx)$ which is equivalent to $\int_{0}^{1} [\hat{G}^k(F^{-1}(u))]^2 du$.

The statistic $S^1_n$ can be used to test the conditional mean function whereas $S^2_n$ to test the conditional variance function. To conduct a joint test of mean and variance functions one must combine $S^1_n$ and $S^2_n$. To combine both statistics, Chen et al. (2015) used $S^1_n + S^2_n$ and $\max(S^1_n, S^2_n)$. Such combinations were adequate since in his case both statistics had the same limiting distribution. In our context, this is not true in general, hence to combine these test statistics we will need a more general approach such as the ones discussed in Ghoudi, Kulperger, and Rémillard (2001) and Genest and Rémillard (2004). To be specific, let

$$\hat{K}_k(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} W_{\hat{\theta}_n}^k(X_i, I_{i-1})^2 1_\{X_{i-1} \leq x\}, \quad x \in \mathbb{R} \quad \text{and} \quad k = 1, 2.$$  

Note that $\hat{K}_k$ is a consistent estimate of $K_k$. We also introduce $L_{n,k} = \sum_{i=1}^{n} \hat{K}_k(X_i)/n$ which is a consistent estimate of $L_k = \mathbb{E}(S^k) = \int K_k(x) F(dx)$.

Now we define our statistics

$$S^*_n = S^1_n/L_{n,1} + S^2_n/L_{n,2}$$

$$S^*_n = \max(S^1_n/L_{n,1}, S^2_n/L_{n,2})$$

$$S^*_n = -2[\ln(P_\nu(S^1_n)/L_{n,1}) + \ln(P_\nu(S^2_n)/L_{n,2})],$$

where $P_\nu(S^k_n)$ denotes the P-value of the statistics $S^k_n$. According to Fisher (1950, pp. 99–101), when $S^1_n$ and $S^2_n$ are independent, the statistic $S^*_n$ converges to a Chi-square distribution with 4 degrees of freedom and provides the optimal way of combining $S^1_n$ and $S^2_n$. As it will be seen, in the simulation and application sections, $S^*_n$ and $S^*_n$ are easier to compute and have similar power to $S^*_n$. 


4.1. Martingale transform based test statistics

Next, we repeat a similar approach using the transformed process $T_n(\widehat{D}_n)$ yielding

$$\tilde{S}_n^k = \int_{-\infty}^{\infty} \left( \frac{T_n^k(\widehat{D}_n^k)(x)}{(\gamma^k_n)^2} \right)^2 \tilde{K}_k(dx) = \frac{1}{n(\gamma^k_n)^2} \sum_{i=1}^{n} \left( \frac{T_n^k(\widehat{D}_n^k)(X_{i-1})}{(\gamma^k_n)^2} \right)^2 W_{\theta_n}(X_i, I_{i-1})^2,$$  \hspace{0.5cm} (7)

where $\gamma^k_n = \lim_{x \to \infty} \tilde{K}_k(x) = \frac{1}{n} \sum_{i=1}^{n} W_{\theta_n}(X_i, I_{i-1})^2$. Theorem 3.2 and the continuous mapping Theorem imply that, under the null hypothesis, $\tilde{S}_n^k$ converges in distribution to $\tilde{S}^k := \int_{0}^{1} \mathbb{W}_k(u)^2 du$, where $\mathbb{W}_k$ is a standard Brownian motion. According to Shorack and Wellner (1986), page 748, the limiting critical values of $\tilde{S}_n^k (k = 1, 2)$ are 1.2, 1.657 and 2.8 corresponding to the significance levels 10%, 5% and 1%, respectively.

Since $\tilde{S}_n^1$ and $\tilde{S}_n^2$ admit the same limiting distribution, there is no need for the normalisation by $L_1$ and $L_2$ in the definition of the combined statistics. Therefore, the combined statistics using the transformed process are defined as follows: $\tilde{S}_n^* = \tilde{S}_n^1 + \tilde{S}_n^2$, $\tilde{S}_n^* = -2[\ln(P_v(\tilde{S}_n^1)) + \ln(P_v(\tilde{S}_n^2))]$ and $\tilde{S}_n^* = \max(\tilde{S}_n^1, \tilde{S}_n^2)$. As pointed above, the test statistics $\tilde{S}_n^*$, $\tilde{S}_n^*$ and $\tilde{S}_n^*$ will be distribution-free only if $\tilde{S}_n^1$ and $\tilde{S}_n^2$ are independent. As argued earlier, this is the case if the covariance term $K_{12} = 0$ or more precisely if $\mathbb{E}\{(X_i - m(\theta_0, I_{i-1}))^3 | F_{i-1}\} = 0$ a.s.

The next subsections provides alternatives, based on the non-transformed processes, that works even when $K_{12} \neq 0$. Recalling that the test statistics $\{S_n^k\}_{k=1,2}$ given in (5) converge weakly towards $S^k = \int_{0}^{1} [\hat{D}_k(F^{-1}(u))]^2 du$. Since the process $\hat{D}_k$ has a complicated covariance, which depends on the unknown distribution $F(\cdot)$ and some unknown parameters, the $p$-values of the test statistic $S_n^k$ need to be approximated using either numerical techniques or re-sampling algorithms. A numerical approximation of the limiting distribution of $S_n^k$ and $S_n^*$, is outlined in Section 4.2 and a re-sampling algorithm is described in Section 4.3.

4.2. Numerical approximation of the asymptotic distribution of the test statistics

In this subsection we discuss the application of numerical integration procedure, given by Deheuvels and Martynov (1996), to approximate the asymptotic distribution of the test statistics $S_n^*$, $S_n^1$ and $S_n^2$. For an illustration of the method, we only consider the case of the approximation of the limiting distribution of the test statistic $S_n^*$. The main idea of this technique consists in using numerical integration to approximate the limiting distribution $S^* = \int_{0}^{1} [\hat{D}^1(F^{-1}(u))]^2 / L_1 + [\hat{D}^2(F^{-1}(u))]^2 / L_2 du$ by the quadratic form $Q_m = \sum_{k=1}^{m} a_k [\hat{D}^1(F^{-1}(u_k))]^2 / L_1 + [\hat{D}^2(F^{-1}(u_k))]^2 / L_2$, where $u_k$ are the quadrature nodes, $a_k$ are the quadrature coefficients and $m$ is the number of quadrature points. The easiest choice is to take $a_k = 1/m$ and $u_k = k/m$ for some large integer $m$. Because $\hat{D}$ is a Gaussian process, $Q_m$ is a quadratic form of $(2m)$ normal random variables. Imhof’s characteristic function inversion procedure (Imhof 1961), available in R package CompQuadForm, is used to compute the distribution function of $Q_m$. When applying the procedure to $S_n^1$ or $S_n^2$, $Q_m$ will be a quadratic form with $(m)$ normal random variables. Note that in the computation, before calling the imhof function of the CompQuadForm package, one needs to estimate the covariance matrix of the normal random variables in $Q_m$ and then passes the eigenvalues of this covariance matrix to the imhof function. This is
directly obtained from the estimation of covariance function $\mathcal{K}$ since each of these normal random variables is just $\hat{D}^1(F^{-1}(u_k))$ or $\hat{D}^2(F^{-1}(u_k))$ for $k \in \{1, \ldots, m\}$. The estimation of $\mathcal{K}$ is obtained by replacing $K$, $\Gamma_{\theta_0}$, $G$ and $\Sigma_0$ by their consistent estimators $\hat{K}$, $\hat{\Gamma}$, $\hat{G}$ and $\hat{\Sigma}_0$, where $\hat{K}(t)$ is a symmetric two by two matrix with $(\hat{K}_k)_{k=1,2}$ as diagonal entries $\hat{K}_k$, and $\hat{K}_{12}$ its off diagonal entry $\hat{K}_{12}$ given by

$$\hat{K}_{12}(t) = \frac{1}{n} \sum_{i=1}^{n} W_{\theta_n}^1(X_i, I_{i-1}) W_{\theta_n}^2(X_i, I_{i-1}) \mathbb{I}\{X_{i-1} \leq t\}.$$  

The other estimates are

$$\hat{G}(t) = \frac{1}{n} \sum_{i=1}^{n} W_{\theta_n}(X_i, I_{i-1}) \phi^*(X_i, I_{i-1}, \theta_n) \mathbb{I}\{X_{i-1} \leq t\},$$  

$$\Gamma_{\theta_n}(t) = \frac{1}{n} \sum_{i=1}^{n} \hat{W}_{\theta_n}(X_i, I_{i-1}) \mathbb{I}\{X_{i-1} \leq t\},$$  

and $\hat{\Sigma}_0 = \frac{1}{n} \sum_{i=1}^{n} \phi^*(X_i, I_{i-1}, \theta_n) \phi^*(X_i, I_{i-1}, \theta_n)^\top$. We also used, in the simulation and application sections, the simplest quadratures, namely $a_k = 1/m$ and $u_k = k/m$. Regrating $F^{-1}$, it is replaced by the empirical quantile function and $L_k$ is replaced by $L_{n,k} = \sum_{i=1}^{n} \hat{K}_k(X_i)/n$ ($k = 1, 2$).

**Remark 4.1:** It is worth noting that, under the independence between $S^1$ and $S^2$, the approximation of the distribution of $S^\circ$ follows from the fact that $\mathbb{P}\{S^\circ \leq x\} = \mathbb{P}\{S^1/L_1 \leq x\}\mathbb{P}\{S^2/L_2 \leq x\}$. For the statistic $S^\star$, even though under independence, its limiting distribution is a Chi-square with 4 degrees of freedom, the approximation technique is needed to calculate the $p$-values of $S^1$ and $S^2$. On the other hand, in case $S^1$ and $S^2$ are dependent, one can still use the approximation technique, described above, to calculate the $p$-value for $S^\star$, however the numerical approximation of the joint distribution of $(S^1/L_1, S^2/L_2)$ is required to be able to evaluate the $p$-value for $S^\star$ and $S^\circ$.

**4.3. Multipliers bootstrap for the test statistics**

Another approach to approximate the limiting distributions of $S^1_n, S^2_n, S^\star_n, S^\star_n$ and $S^\circ_n$ consists in using a re-sampling algorithm. To define such algorithm, let $B$ be a positive integer denoting the number of bootstrap samples. For $b = 1, \ldots, B$, let $(Z_{1,b}, \ldots, Z_{n,b})$ be a centered sequence of i.i.d. random variables with variance one independent of the sigma-field generated by the $X_i$’s. The multipliers bootstrap technique applied to the empirical process $\mathbb{D}_n$ is defined as follows

$$\mathbb{D}^\star_{n,b}(x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,b} \left\{ W_{\theta_0}(X_i, I_{i-1}) \mathbb{I}\{X_{i-1} \leq x\} - \Gamma_{\theta_0}^\top(x) \phi^*(X_i, I_{i-1}, \theta_0) \right\}.$$  

Observe that this process depends on the unknown quantities $\theta_0$ and $\Gamma_{\theta_0}$. A plug-in estimator of $\mathbb{D}^\star_{n,b}(x)$ may be obtained by replacing $\theta_0$ and $\Gamma_{\theta_0}$ by their consistent estimators $\theta_n$
and $\Gamma_{\theta_n}(x)$, defined in (8), yielding

$$\mathbb{D}_{n,b}(x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,b} \left\{ W_{\theta_n}(X_i, I_{i-1}) \mathbb{I}\{X_{i-1} \leq x\} - \Gamma_{\theta_n}^\top(x) \phi^*(X_i, I_{i-1}, \theta_n) \right\}.$$ 

The following extra assumptions are needed to establish the weak convergence of $\mathbb{D}_n$.

**Assumption M1**: There exists $\delta > 0$ such that for all $\theta$ satisfying $\|\theta - \theta_0\| \leq \delta$ one has $\|W_b^k(X_i, I_{i-1}) - W_{\theta_0}^k(X_i, I_{i-1})\| \leq \|\theta - \theta_0\| M_3(X_i, I_{i-1})$ where $M_3$ is a positive function satisfying $\mathbb{E}(M_3(X_i, I_{i-1})) \leq C < \infty$.

**Assumption M2**: Let $\delta > 0$ such that for all $\theta$ satisfying $\|\theta - \theta_0\| \leq \delta$ one has

$$\|\phi^*(X_i, I_{i-1}, \theta) - \phi^*(X_i, I_{i-1}, \theta_0) - (\theta - \theta_0)^\top \dot{\phi}^*(X_i, I_{i-1}, \theta_0)\| \leq \|\theta - \theta_0\| \lambda_3(\|\theta - \theta_0\|) M_4(X_i, I_{i-1})$$

where $M_4$ and $\lambda_3$ are positive functions satisfying $\mathbb{E}(M_4(X_i, I_{i-1})) \leq C < \infty$ and $\lim_{t \to 0} \lambda_3(t) = 0$.

**Assumption M3**: Suppose that $E\|\dot{\phi}^*(X_i, I_{i-1}, \theta_0)\dot{\phi}^*(X_i, I_{i-1}, \theta_0)^\top\| \leq C < \infty$.

The next result summarises the asymptotic behaviour of the bootstrapped process $\mathbb{D}_{n,b}$.

**Theorem 4.1**: Suppose that Assumptions (A1)–(A4) and (M1)–(M3) hold true, then, for $b = 1, \ldots, B$, $\mathbb{D}_{n,b}$ converges in distribution to independent copies of $\mathbb{D} = \mathbb{D} - \Gamma_{\theta_0}^\top \Theta$.

Note that the bootstrapped version of the statistic $S_n^k (k = 1, 2)$ is given by

$$S_{n,b}^k := \int_{-\infty}^{\infty} \mathbb{D}_{n,b}^k(x)^2 F_n(dx) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{i,b} Z_{j,b} \mathcal{M}_{i,j}^k,$$

where

$$\mathcal{M}_{i,j}^k = W_{\theta_n}^k(X_i, I_{i-1}) W_{\theta_n}^k(X_j, I_{j-1})(1 - F_n(X_{i-1} \wedge X_{j-1})) - W_{\theta_n}^k(X_i, I_{i-1}) \phi^*(X_i, I_{i-1}, \theta_n)^\top \mathcal{L}_n^k(X_{i-1}) - W_{\theta_n}^k(X_j, I_{j-1}) \phi^*(X_i, I_{i-1}, \theta_n)^\top \mathcal{L}_n^k(X_{j-1}) + \phi^*(X_i, I_{i-1}, \theta_n)^\top AN \phi^*(X_i, I_{i-1}, \theta_n),$$

where $\mathcal{L}_n^k(t) := \int_{-\infty}^{\infty} \Gamma_{\theta_n}^k(u) F_n(du)$ and $\mathcal{N} := \int \Gamma_{\theta_n}^k(u) \Gamma_{\theta_n}^k(u)^\top F_n(du)$.

Set, for $k = 1, 2$, $L_{n,k} = n^{-1} \sum_{i=1}^{n} \hat{k}_k(X_i)$. To obtain a bootstrapped version of $S_n^*$ and $S_n^\circ$, one uses $S_{n,b}^1 = S_{n,b}^1/L_{n,1} + S_{n,b}^2/L_{n,2}$ and $S_{n,b}^\circ = \max\{S_{n,b}^1/L_{n,1}, S_{n,b}^2/L_{n,2}\}$, respectively.

To apply the multipliers bootstrap procedure to approximate the $p$-value of any of the Cramèr–von Mises statistics $S_n^1$, $S_n^2$, $S_n^\circ$, one proceeds according to the following algorithm.

- Estimate $\theta_0$ by $\theta_n$ and compute the test statistics $S_i^n$ for $i \in \{1, 2, *, \circ\}$. 
Table 1. CPU times in seconds for 1000 simulation replicates for different testing procedures.

| Procedure                      | Model A0 | Model A1 |
|--------------------------------|----------|----------|
| Approximation                  | 84.7     | 85.5     |
| Khmaladze transform            | 159.2    | 158.3    |
| Multipliers bootstrap          | 603.3    | 436.6    |
| Parametric bootstrap           | 25007.5  | 24476.7  |

- For $b \in \{1, \ldots, B\}$
  - Generate $(Z_{1,b}, \ldots, Z_{n,b})^\top$ as a sequence of i.i.d normal $(0, 1)$ random variables.
  - Compute the bootstrapped Cramér–von Mises statistics $S_{n,b}^i$ for $b = 1, \ldots, B$ and $i \in \{1, 2, \star, \circ\}$.
- Estimate the $p$-value, $P_{V_i}$ of $S_{n,b}^i$ by $P_{V_i} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{S_{n,b}^i > S_{n,b}^i\}$ for $i \in \{1, 2, \star, \circ\}$.

For the test statistics $S_{n,b}^\star$, first compute $S_{n,b}^\star = -2 \ln(P_{V_1}) - 2 \ln(P_{V_2})$. Then, for each $b_0 = 1, \ldots, B$, compute $S_{n,b_0}^\star = -2 \ln(P_{V_{1,b_0}}) - 2 \ln(P_{V_{2,b_0}})$ where $P_{V_{k,b_0}} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{S_{n,b}^k > S_{n,b_0}^k\}$ for $k = 1, 2$. The $p$-value of the statistics $S_{n,b}^\star$ is estimated by $P_{V_{\star}} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{S_{n,b}^\star > S_{n,b}^\star\}$. Note that the $p$-values obtained using the bootstrap procedure described above are valid for all situations including the case of dependent components of the process $D$. However, in the case where the process $D$ has independent components, one can obtain $P_{V_{\star}}$ using the Chi-square distribution with 4 degrees of freedom.

**Remark 4.2:** Note that in the above algorithm $\theta$ is estimated only once and the matrices $M^k$ with $k = 1, 2$ are only computed once. This makes the multipliers Bootstrap extremely fast as each bootstrap iteration involves only the generation of a vector $Z_b$ of $n$ i.i.d normal mean zero and variance 1 random variables and then the statistics $S_{n,b}^k$ is computed as $S_{n,b}^k = Z_b^\top M^k Z_b / n$. This is much faster than the parametric bootstrap which requires re-estimating the parameters and recalculating all the quantities including the matrices $M^1$ and $M^2$ for each bootstrap iteration. To compare the computation times, we implemented the parametric bootstrap for models A0 and A1 of Section 5.3 and we determined the CPU time required to run 1000 simulation replicates with sample size $n = 300$ and $B = 500$ bootstrap samples on a Dell 5820 workstation.

Table 1 shows that in term of speed, the multipliers bootstrap is a bit slower than the Khmaladze transform and the numerical approximation, but it is definitely more than 41 times faster than parametric bootstrap.

Note that, the theoretical validity of the parametric bootstrap was discussed in Escanciano (2008). It was implemented here following steps (1–4) of the same paper. Its finite sample performances, for models A0 and A1, were similar to those of the multipliers bootstrap.

**5. Finite sample performance**

This section presents several simulation experiments carried out to assess the power of the proposed test statistics. The first experiment, described in Section 5.1, is designed to assess the fit of a pure ARCH(1) model. The second experiment, outlined in Section 5.2,
assesses the power of our test statistics for detecting departure from GARCH(1,1) model. The third experiment, given in Section 5.3, studies the behaviour of the tests when fitting and AR(1)-GARCH(1,1) model. The fourth experiment, presented in Section 5.4, outlines the main findings of our testing procedures applied to the case of stochastic differential equation models. Two sub-experiments are considered, in the first one we test if the model has a linear drift and a constant diffusion while in the second one we test if the model has linear drift and a diffusion proportional to the square root of the series.

### 5.1. Testing ARCH(1) model

The purpose of the simulation experiment considered here is to test whether the data is generated according to one of the models

\[
(M_0) \quad X_t = \sqrt{h_t} \epsilon_t, \quad \text{where} \quad h_t = 1.1 + 0.5X_{t-1}^2, \\
(M_1) \quad X_t = \sqrt{h_t} \epsilon_t, \quad \text{where} \quad h_t = 1.1 + 0.5X_{t-1}^2 + 0.5X_{t-1}, \\
(M_2) \quad X_t = \sqrt{h_t} \epsilon_t, \quad \text{where} \quad h_t = 1.1 + 0.5X_{t-1}^2 + 0.5 \text{sign}(X_{t-1}), \\
(M_3) \quad X_t = \sqrt{h_t} \epsilon_t, \quad \text{where} \quad h_t = 1.1 + 0.5X_{t-1}^2 + X_{t-1}, \\
(M_4) \quad X_t = \sqrt{h_t} \epsilon_t, \quad \text{where} \quad h_t = 1.1 + 0.5X_{t-1}^2 + \text{sign}(X_{t-1}), \\
(M_5) \quad X_t = 0.2X_{t-1} + \sqrt{h_t} \epsilon_t, \quad \text{where} \quad h_t = 1.1 + 0.5X_{t-1}^2 + 0.5X_{t-1}. 
\]

A pure ARCH(1) model is fitted to the data. Then tests described in Section 4 are applied using the \( \phi^* \) corresponding to maximum likelihood estimator of the ARCH model with parameters \( \alpha_0 \) and \( \alpha_1 \). We generated series of length \( n = 100, 300 \) following models \( M_0 - M_5 \). Results given in Table 2 summarise a Monte-Carlo simulation study with 2000 replications of tests with 5% significance levels. Note that \( M_0 \) corresponds to the null

### Table 2. Percentage of rejection of the null hypothesis (ARCH(1) model) when the data are generated according to models \( M_0 \) to \( M_5 \) and \( n = 100 \) and 300.

| n   | DGP  | Transformation technique | Multipliers procedure | Numerical approximation |
|-----|------|--------------------------|----------------------|------------------------|
|     |      | \( S_1 \) | \( S_2 \) | \( S_3 \) | \( S_4 \) | \( S_5 \) | \( S_6 \) | \( S_7 \) | \( S_8 \) |
| \( M_0 \) | 100  | 3.7 | 4.2 | 3.9 | 4.0 | 4.0 | 4.3 | 3.6 | 3.7 | 3.8 | 4.0 | 4.7 | 3.3 | 3.6 | 3.8 | 3.8 |
| \( M_1 \) | 100  | 4.5 | 20.8 | 15.0 | 14.5 | 15.1 | 5.3 | 21.2 | 12.0 | 10.8 | 14.0 | 5.5 | 18.4 | 11.6 | 10.2 | 13.3 |
| \( M_2 \) | 100  | 4.3 | 20.9 | 14.9 | 14.9 | 14.8 | 4.8 | 15.2 | 9.5 | 7.9 | 11.8 | 5.2 | 14.2 | 9.0 | 7.6 | 11.3 |
| \( M_3 \) | 100  | 3.8 | 43.4 | 30.7 | 32.9 | 29.4 | 5.2 | 61.0 | 42.2 | 44.7 | 44.5 | 5.3 | 60.4 | 42.6 | 44.5 | 43.1 |
| \( M_4 \) | 100  | 3.1 | 47.4 | 35.1 | 35.9 | 34.3 | 3.6 | 43.5 | 24.8 | 24.1 | 32.5 | 3.9 | 42.6 | 24.5 | 24.0 | 31.0 |
| \( M_5 \) | 100  | 12.2 | 23.2 | 23.1 | 21.5 | 23.2 | 16.8 | 22.3 | 23.6 | 19.3 | 26.5 | 17.2 | 22.8 | 24.1 | 19.7 | 26.9 |
| \( M_0 \) | 300  | 3.9 | 5.5 | 4.6 | 5.0 | 4.7 | 4.4 | 5.0 | 4.5 | 4.5 | 4.8 | 4.4 | 4.6 | 4.3 | 4.7 | 4.4 |
| \( M_1 \) | 300  | 3.7 | 45.3 | 35.9 | 37.0 | 35.1 | 4.7 | 57.7 | 44.3 | 44.5 | 45.3 | 4.9 | 57.8 | 44.3 | 45.1 | 44.3 |
| \( M_2 \) | 300  | 3.8 | 40.2 | 31.7 | 31.6 | 30.5 | 4.6 | 38.0 | 23.8 | 22.2 | 29.3 | 4.1 | 37.9 | 23.6 | 21.2 | 27.6 |
| \( M_3 \) | 300  | 3.5 | 63.1 | 54.6 | 56.4 | 53.9 | 4.3 | 92.1 | 88.9 | 89.8 | 87.0 | 4.2 | 92.2 | 89.3 | 90.0 | 86.9 |
| \( M_4 \) | 300  | 3.6 | 68.0 | 59.2 | 60.4 | 58.4 | 4.3 | 76.0 | 64.6 | 64.7 | 66.9 | 4.4 | 75.9 | 64.7 | 64.9 | 66.7 |
| \( M_5 \) | 300  | 27.5 | 44.6 | 53.0 | 47.5 | 54.3 | 43.3 | 56.0 | 69.3 | 62.4 | 71.0 | 43.1 | 57.0 | 70.8 | 64.7 | 72.8 |
hypothesis and, as shown in the table, the tests maintain their levels quite well. Tests based on Khamaladze transform are in general a bit less powerful than those based on the original process. As expected, Table 2 shows that $S_1^1$ and $\tilde{S}_1^1$ have no power detecting $M_1$–$M_4$ since the mean is correctly specified for these models. The power of the combined statistics $S_1^n$, $S_2^n$ and $S_3^n$ (or $\tilde{S}_1^n$, $\tilde{S}_2^n$ and $\tilde{S}_3^n$) remains similar in general. The powers obtained using the multipliers procedure and those derived from numerical approximation are very close. This shows that both techniques provide very good estimation of the $p$-values of these test statistics. For models $M_1$–$M_4$, the statistic $S_2^n$ is clearly the source of the detection. For model $M_5$ the situation is bit different since both mean and variance functions are incorrectly specified. We notice that both $S_1^n$ and $S_2^n$ have some power, but the combined statistics are more powerful in this case. This conclusion is more evident when $n = 300$.

5.2. Testing GARCH(1,1) model

This section presents the result of a simulation study in which we assess the power of the tests discussed in this manuscript, in detecting departure from a GARCH(1,1) model. That is we wish to test if the mean and variance functions are those of a GARCH(1,1) or not. We use the same settings as in Experiment 2 of Escanciano (2010). More precisely, we generated the data from an AR(1)-GARCH(1,1) model with autoregressive coefficient $a_1$ varying from $-0.9$ to $0.9$. We fitted a GARCH(1,1) and recorded the percentage of rejection. As in all simulations reported in this manuscript we used 2000 Monte-Carlo simulation iterations. For this experiment we used a sample size $n = 100$ an in Escanciano (2010). The results are reported in Table 3. Note that $a_1 = 0$ corresponds to the null hypothesis in this case. Table 3 shows that the 5% level is respected quite well in general. It also shows that the type of alternative considered here is mainly detected by the contribution of component $S_1^n$ or $\tilde{S}_1^n$ to the combined statistics. This makes all combined statistics quite powerful in

| $n$ | $a_1$ | $S_1^n$ | $S_2^n$ | $S_3^n$ | $S_4^n$ | $S_5^n$ | $S_6^n$ | $S_7^n$ | $S_8^n$ | $S_9^n$ | $S_{10}^n$ | $S_{11}^n$ | $S_{12}^n$ | $S_{13}^n$ |
|-----|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 100 | -0.9 | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   |
|     | -0.7 | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   |
|     | 0.0  | 63.2  | 4.4   | 4.4   | 4.3   | 4.3   | 4.3   | 4.3   | 4.3   | 4.3   | 4.3   | 4.3   | 4.3   | 4.3   |
|     | 0.3  | 68.1  | 5.4   | 5.4   | 5.3   | 5.3   | 5.3   | 5.3   | 5.3   | 5.3   | 5.3   | 5.3   | 5.3   | 5.3   |
|     | 0.5  | 99.0  | 2.5   | 2.5   | 2.5   | 2.5   | 2.5   | 2.5   | 2.5   | 2.5   | 2.5   | 2.5   | 2.5   | 2.5   |
|     | 0.7  | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   |
|     | 0.9  | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   |
| 300 | 0.0  | 4.8   | 4.7   | 4.7   | 4.7   | 4.7   | 4.7   | 4.7   | 4.7   | 4.7   | 4.7   | 4.7   | 4.7   | 4.7   |
|     | 0.3  | 99.2  | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   |
|     | 0.5  | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   |
|     | 0.7  | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   |
|     | 0.9  | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   | 100   |
detecting such alternatives. The components $S_n^2$ and $\tilde{S}_n^2$ have no power against this type of alternatives. This is expected since these components were designed to detect change in the variance function. In this context, statistics based on the transformed process seem to be a bit more powerful than those based on the original process. Comparing our results to Figure 1 in Escanciano (2010), we notice that the power of the tests presented here is slightly better.

5.3. Testing AR(1)-GARCH(1,1) model

In this section we present the result of a study testing the null hypothesis $A_0$ that the mean and variance functions are those of an AR(1)-GARCH(1,1) model. The same five alternatives considered in Experiment 3 in Escanciano (2010) are used here. To be specific, the data are generated according to one of the following alternatives, whereas AR(1)-GARCH(1,1) is the model fitted to the data. The test statistics proposed in this manuscript are applied and the percentage of rejection, in a 2000 replicates Monte-Carlo simulation, is reported in Table 4. The alternatives $A_1$–$A_5$ considered are defined as follows:

$A_0$: The null hypothesis AR(1)-GARCH(1,1) model: $X_t = 0.02 + 0.02X_{t-1} + \varepsilon_t$ where $\varepsilon_t = \sqrt{h_t} u_t$ with $h_t = 0.08 + 0.1\varepsilon_{t-1}^2 + 0.85h_{t-1}$ and $u_t$ i.i.d Normal random variables with mean zero and variance one.

$A_1$: ARMA(1,1)-GARCH(1,1) model: $X_t = 0.02 + 0.02X_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$ and $\varepsilon_t$ is as in model $A_0$. 

Figure 1. Interest rate for a maturity of 1 months (% per year) from July 1964 to April 1989.
Table 4. Percentage of rejection of the null hypothesis (AR(1)-GARCH(1,1) model) when the data are generated according to A0–A5.

| n  | DGP     | Transformation technique | Multipliers procedure | Numerical approximation |
|----|---------|--------------------------|-----------------------|------------------------|
|    |         | $S_1^*$                  | $S_2^*$               | $S_3^*$               |
|    |         | $S_4^*$                  | $S_5^*$               | $S_6^*$               |
|    |         | $S_7^*$                  | $S_8^*$               | $S_9^*$               |
| 300| A0      | 4.6 3.9 4.6 4.5 4.9     | 2.6 5.4 4.4 5.3 4.5  | 2.3 5.3 4.5 5.1 4.2   |
|    | A1      | 3.6 4.1 3.4 3.8 3.6     | 1.4 4.3 3.2 3.5 2.9  | 1.3 3.8 2.9 3.4 2.8  |
|    | A2      | 100 5.8 99.8 99.9 99.8  | 77.8 31.9 61.6 47.2 78.8 | 77.7 31.9 63.1 47.3 78.4 |
|    | A3      | 6.2 6.0 7.1 6.8 7.1     | 55.5 51.6 71.8 64.9 78.1 | 55.5 52.0 71.7 64.9 76.9 |
|    | A4      | 68.3 5.3 59.0 60.3 57.7 | 13.6 27.6 30.5 27.2 32.2 | 13.0 27.4 31.1 27.3 31.7 |
|    | A5      | 100 15.4 100 100 98.2  | 91.3 18.1 87.2 87.4 89.0 | 91.4 18.2 87.7 86.7 89.4 |
| 600| A0      | 4.0 4.9 3.7 3.8 3.6     | 2.2 5.0 4.2 4.5 3.3  | 2.2 4.7 3.8 4.4 3.1  |
|    | A1      | 100 11.2 100 100 100    | 98.2 63.7 96.0 91.0 98.4 | 98.1 63.6 96.1 91.0 98.4 |
|    | A3      | 4.8 13.7 10.8 11.2 10.6 | 80.1 69.6 92.6 90.5 94.6 | 80.7 70.0 92.6 90.4 94.8 |
|    | A4      | 88.4 6.4 84.0 84.4 83.3 | 30.7 41.7 51.6 44.0 56.6 | 29.5 41.1 50.8 43.3 55.9 |
|    | A5      | 100 40.9 100 100 100    | 97.6 50.2 97.5 97.2 97.9 | 97.6 51.0 97.5 97.3 97.8 |

A2: TAR model: $X_t = 0.6X_{t-1} + \varepsilon_t$ if $X_{t-1} \leq 1$ and $X_t = -0.5X_{t-1} + \varepsilon_t$ if $X_{t-1} > 1$ with $

A3: EGARCH(1,1) model: $X_t = \sqrt{h_t}u_t$ where $\ln(h_t) = 0.025 + 0.5\ln(h_{t-1}) + 0.25(|u_{t-1}| - \sqrt{2/\pi}) - 0.8u_{t-1}$ and $u_t$ is as in model A0.

A4: Bilinear model: $X_t = 0.6X_{t-1} + 0.7u_{t-1}X_{t-2} + u_t$ where $u_t$ is as in model A0.

A5: Nonlinear Moving average model: $X_t = 0.8u^2_{t-1} + u_t$ where $u_t$ is as in model A0.

Comparing with Table 2 in Escanciano (2010), one notes that combined tests introduced in this manuscript are more powerful than those considered in Escanciano (2010). The only exception being the case of the alternative A1: ARMA(1,1)-GARCH(1,1), where our tests were not able to detect such alternative while those in Escanciano (2010) had reasonable power. The reason our statistics do not detect A1, is that this model satisfies the condition stated in Remark 2.1. To be specific, first consider the model A1* defined by $X_t = 0.02 + 0.02X_{t-1} + 0.5(\varepsilon_{t-1} - \mathbb{E}(\varepsilon_{t-1}|X_{t-1})) + \varepsilon_t$ and $\varepsilon_t$ is as in model A0. The difference between A0 and A1* given by $d_t = 0.5(\varepsilon_{t-1} - \mathbb{E}(\varepsilon_{t-1}|X_{t-1}))$ satisfies $\mathbb{E}(d_t|X_{t-1}) = 0$. Hence, our test statistics will not detect the difference between A0 and A1*. Now for A1, when we fit AR(1) instead of ARMA(1,1) to the $X_t$, what really happens is that the $\varepsilon_{t-1}$ term of the ARMA(1,1) gets replaced by it projection on the space of $X_{t-1}$ which is equivalent to replacing $\varepsilon_{t-1}$ by it linear regression estimate on $X_{t-1}$. Hence, the difference between the true model and the fitted model is proportional to the residuals of such regression which, in theory, has mean zero and is independent of $X_{t-1}$.

We also notice that tests based on the transformed process are bit more powerful in the case of alternatives A2, A4 and A5. On the other hand these tests fall way behind in detecting A3. This seems to concord with the first simulation experiment, where we have observed that tests based on the transformed process had a bit more power in detecting a change in the mean function. In fact, for the three alternatives A2, A4 and A5 there is only a change in the conditional mean function. However, when we consider alternative A3, the conditional mean function is still that of an AR(1) while the change occurred in the conditional variance function.
5.4. Testing stochastic differential equation models

In this part, we are interested in testing whether the data fits a specified stochastic differential equation (SDE) model. In finance continuous time models are widely used to study the dynamic of some financial products such as asset prices, interest rate and bonds. Continuous-time modelling in finance was introduced by Bachelier (1900) on the theory of speculation and really started with Merton (1970) seminal work. Since then several models were developed which can be formulated by the following general SDE

\[ dX_t = \mu_\theta(X_t)\,dt + \sigma(X_t)\,dW_t, \quad (10) \]

where \( W_t \) is a standard Brownian motion. The drift \( \mu_\theta(\cdot) \) and the diffusion \( \sigma^2(\cdot) \) are known functions. Several well-known models in financial econometrics, (including Black and Scholes 1973; Vasicek 1977; Cox, Ingersoll Jr, and Ross 1980; Chan, Karolyi, Longstaff, and Sanders 1992; Aït-Sahalia 1996, among others), can be written under the form (10) with a specific form of drift and diffusion functions. Below the list of SDE models considered here.

N1: Vasicek: \( dX_t = 3(10 - X_t)\,dt + 5dW_t, x_0 = 0.03, \)
N2: Hyperbolic: \( dX_t = 5\frac{X_t}{\sqrt{1 + X_t^2}}\,dt + 5dW_t, x_0 = 3, \)
N3: CIR: \( dX_t = (1 + 4.5X_t)\,dt + 0.75\sqrt{X_t}dW_t, x_0 = 3, \)
N4: CKLS1: \( dX_t = 1.5(1 - X_t)\,dt + 1.5X_t^{0.8}dW_t, x_0 = 5, \)
N5: CKLS2: \( dX_t = 1.5(1 - X_t)\,dt + 0.5X_t^{1.5}dW_t, x_0 = 5, \)
N6: Aït-Sahalia: \( dX_t = (1 + 15X_t + 0.25X_t^{-1} - 2X_t^2)\,dt + 0.5X_t^{1.5}dW_t, x_0 = 5. \)

In practice, the diffusion process \( \{X_t\} \) is observed at instants \( \{t = i\Delta| i = 0, \ldots, n\} \), where \( \Delta > 0 \) is generally small, but fixed as \( n \) increases. For instance the series could be observed hourly, daily, weekly or monthly. Therefore, we may model these discretely observed measurements using time series models. In fact, despite the fact that (10) is written in a continuous-time form, one often uses the following Euler discretisation scheme to get

\[ X_{t+\Delta} - X_t = \mu_\theta(X_t)\Delta + \sigma(X_t)\,(W_{t+\Delta} - W_t), \quad t = 0, \Delta, 2\Delta, \ldots, \quad (11) \]

as an approximation that facilitates computational and theoretical derivation. The accuracy of such Euler discretisation is studied in Jacod and Protter (1998). The purpose of this simulation study is to generate processes satisfying the SDEs described by models N1–N6 above, then use the procedures described in Section 4 to test whether the data fits a specific type of SDE models. Two types of null hypotheses will be considered. The first one is an homoscedastic model corresponding to Vasicek model with \( \mu \) of the form \( a + bX_{t-1} \) for some parameters \( (a, b) \in \mathbb{R}^2 \) and \( \sigma = c \) for some parameter \( c > 0 \). The second one is an heteroscedastic CIR model where \( \mu \) is of the form \( a + bX_{t-1} \) and \( \sigma \) equals \( c\sqrt{X_{t-1}} \).

We suppose the process is observed over the time interval \([0, T]\) and \( n \) corresponds to the number of instants where the process is observed. The sampling mesh in such case is \( \Delta = T/n \). In order to assess the sensitivity of our test to the sampling frequency from the underlying process, we consider \( T = 1 \) and \( \Delta = 0.002, 0.005 \) and \( 0.01 \), corresponding to a total sampling instants \( n = 500, 200 \) and \( 100 \), respectively. Such framework supposes
that collecting more observations means shortening the time interval between successive existing observations, not lengthening the time period over which data are recorded.

The data is generated according to one of SDE models listed above. The parameters are estimated, after discretisation, using the maximum pseudo-likelihood method (see Aït-Sahalia 2002). Tables 5 and 6 report the percentages of rejection of the two null hypotheses considered. The results are obtained using 1000 Monte-Carlo replications for different values of sampling mesh Δ.

### Table 5. Percentage of rejection of the null hypothesis \( H_0 \) (Vasicek type Model) for different sample sizes \( n = 100, 200 \) and 500.

| \( n \) | DGP | \( S_{n1}^1 \) | \( S_{n1}^2 \) | \( S_{n1}^* \) | \( S_{n2}^1 \) | \( S_{n2}^2 \) | \( S_{n2}^* \) | \( S_{n1}^1 \) | \( S_{n1}^2 \) | \( S_{n1}^* \) | \( S_{n2}^1 \) | \( S_{n2}^2 \) | \( S_{n2}^* \) | \( S_{n1}^1 \) | \( S_{n1}^2 \) | \( S_{n1}^* \) | \( S_{n2}^1 \) | \( S_{n2}^2 \) | \( S_{n2}^* \) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | N1 | 5.1 | 4.2 | 3.8 | 4.1 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N2 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N3 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N4 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N5 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N6 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |

### Table 6. Percentage of rejection of the null hypothesis \( H_0 \) (CIR type Model) for different sample sizes \( n = 100, 200 \) and 500.

| \( n \) | DGP | \( S_{n1}^1 \) | \( S_{n1}^2 \) | \( S_{n1}^* \) | \( S_{n2}^1 \) | \( S_{n2}^2 \) | \( S_{n2}^* \) | \( S_{n1}^1 \) | \( S_{n1}^2 \) | \( S_{n1}^* \) | \( S_{n2}^1 \) | \( S_{n2}^2 \) | \( S_{n2}^* \) | \( S_{n1}^1 \) | \( S_{n1}^2 \) | \( S_{n1}^* \) | \( S_{n2}^1 \) | \( S_{n2}^2 \) | \( S_{n2}^* \) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | N1 | 5.1 | 4.2 | 3.8 | 4.1 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N2 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N3 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N4 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N5 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
| | N6 | 4.2 | 4.1 | 3.8 | 4.2 | 4.7 | 5.2 | 3.7 | 3.5 | 3.6 | 4 | 100 | 100 | 100 | 100 | 5.5 | 100 | 100 | 100 | 99.6 |
For the first sub-experiment, Table 5 shows that the tests are quite powerful in detecting change in the diffusion term. Alternatives N3–N6 are rejected with high probability even for \( n = 100 \). The three combined statistics are more or less similar in term of their detection power. As expected, the statistics \( S_n \) or \( \tilde{S}_n \) do no detect change in the diffusion term. All statistics fail to detect the alternative N2 even for large sample sizes. A closer look at the trajectories of the process N2 revealed that these trajectories are identical to those of a linear drift. The reason is essentially due to the discretisation used to generate the data and estimate the parameters of the SDE given in (11). In our simulation, we are using \( \Delta = 0.002, 0.005 \) and 0.01. Within this small intervals, the quantity \( X_t / \sqrt{1 + X_t^2} \) is equivalent to its first-order Taylor expansion. We also noticed in our application to interest rate data that model D1, of the same class as N1, and model D2, of the same class as N3, provide similar results.

For the second sub-experiment, reported in Table 6, tests based on the original process and using either the multipliers procedure or the numerical approximation do respect their levels more appropriately than those based on the transformed process. The three combined statistics have the power to detect all alternatives considered here with N5 and N6 being easier to detect. This could be explained by the large change in the diffusion functions between N5 or N6 and the null hypothesis in this case N3.

6. Application to interest rate data

This section presents an application of the testing methodology to interest rate dynamics. Monthly values of interest rate for a maturity of 1 month (% per year) from July 1964 to April 1989 are considered. The monthly interest rate \((r_t)\) dynamic is displayed in Figure 1. Figure 2 is a scatter plot of monthly changes in interest rate against the previous month’s rate and it clearly shows an obvious heteroscedasticity, with the range of changes increasing significantly as the level of interest rates increases. An augmented Dickey-Fuller test of non stationarity of interest rate data is performed and the non stationarity hypothesis is rejected, with \( p \)-value equal to 0.01, for 1 to 12 months lagged series.

![Figure 2](image-url)  
*Figure 2.* Monthly changes in interest rate against interest rate on preceding month.
Table 7. P-values of the test statistics applied to different models (D1–D7) fitted to the monthly interest rate data.

| Test                      | D1    | D2    | D3    | D4    | D5    | D6    | D7    |
|---------------------------|-------|-------|-------|-------|-------|-------|-------|
| Transformation technique  |       |       |       |       |       |       |       |
|  \( \hat{S}_1 \)         | 0.1432| 0.1108| 0.3589| 0.1548| 0.1574| 0.1779| 0.5179|
|  \( \hat{S}_2 \)         | 0.0001| 0.0001| 0.0000| 0.0018| 0.0096| 0.7873| 0.7937|
|  \( \hat{S}_\ast \)      | 0.0001| 0.0002| 0.0002| 0.0028| 0.0128| 0.3911| 0.7779|
|  \( \hat{S}_\circ \)      | 0.0001| 0.0002| 0.0001| 0.0036| 0.0191| 0.3241| 0.7676|
|  \( \hat{S}_\bullet \)    | 0.2840| 0.3080| 0.5130| 0.5370| 0.6410| 0.8170| 0.9850|
| Multipliers procedure     |       |       |       |       |       |       |       |
|  \( \hat{S}_1 \)         | 0.2840| 0.3080| 0.5130| 0.5370| 0.6410| 0.8170| 0.9850|
|  \( \hat{S}_2 \)         | 0.0000| 0.0001| 0.0000| 0.0001| 0.0010| 0.6077| 0.6126|
|  \( \hat{S}_\ast \)      | 0.0000| 0.0002| 0.0000| 0.0002| 0.0019| 0.7812| 0.8343|
|  \( \hat{S}_\circ \)      | 0.0000| 0.0001| 0.0000| 0.0001| 0.0010| 0.7169| 0.6908|
|  \( \hat{S}_\bullet \)    | 0.0001| 0.0003| 0.0001| 0.0008| 0.0056| 0.8490| 0.9088|

The aim here is to verify if the interest rate data described above fits a selected diffusion model. The following models will be considered as candidates

(D1) Vasicek: \( dr_t = (\alpha + \beta r_t)dt + \sigma dW_t \),

(D2) Hyperbolic: \( dr_t = \alpha \frac{r_t}{\sqrt{1+r_t^2}}dt + \sigma dW_t \),

(D3) Aït-Sahalia 1: \( dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^{-1} + \alpha_3 r_t^2)dt + \sigma dW_t \),

(D4) CIR: \( dr_t = (\alpha + \beta r_t)dt + \sigma \sqrt{r_t} dW_t \),

(D5) CKLS 1: \( dr_t = \kappa (\alpha - r_t)dt + \sigma r_t^{0.8} dW_t \),

(D6) CKLS 2: \( dr_t = \kappa (\alpha - r_t)dt + \sigma r_t^{1.5} dW_t \),

(D7) Aït-Sahalia 2: \( dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^{-1} + \alpha_3 r_t^2)dt + \sigma r_t^{1.5} dW_t \).

The procedure consists in fitting the interest rate data to each of the models outlined above and then applying the statistics described in Section 4 to see how well the model fits the data. Table 7 reports the p-values resulting from applying the statistics discussed in Section 4 after fitting each of the models (D1)–(D7) to the monthly interest rate data. As expected, all the models with constant conditional volatility (D1-D3) are clearly rejected even when we considered different drift functions. Models (D4) and (D5) are also rejected, implying that the rate in the diffusion function of the monthly interest rate is not 0.5 or 0.8. On the other hand, all tests cannot reject the (D6) and (D7) implying that the diffusion of the interest rate is proportional to \( X_t^{1.5} \). This conclusion is in perfect concordance with existing findings in the literature (see for instance Aït-Sahalia 1999). In fact, the rate of 1.5 was the recommended choice in Aït-Sahalia (1996), in a study of the same interest rate data.

7. Conclusion

Though Khmaldze matringale transform is quite popular as it simplifies the asymptotic behaviour. However, its finite sample performance is in general far from being the best compared to re-sampling or numerical approximation of the original statistics. We also
notice that even the levels of tests based on Khmaladze transform were sensitive to the estimation of the function \(g\) defined in (3). In our implementation, outlined in Appendix 2, we estimated the function \(g\) using a nonparametric estimate of the derivative of the function \(\Gamma_{\theta_0}\). We noticed that the levels of tests based on Khmaladze transform were sensitive to the bandwidth parameter utilised in the estimation of the derivatives. Levels of tests obtained by numerical approximation or re-sampling of the original statistics are stable and close to their target values. Tests, using the original cumulative residual process, are in general more powerful than the ones based on martingale transformation. Moreover, as mentioned earlier, \(p\)-values of the combined statistics based on the transformed process are only computable when the components \(S^1_n\) and \(S^2_n\) are independent. Whereas, \(p\)-values for the multipliers bootstrap procedure for both \(S^*_n\) and \(S^\circ_n\) are obtained in the same manner whether the components \(S^1_n\) and \(S^2_n\) are independent or not. For the approximation technique, the algorithm described in the manuscript provides the \(p\)-values of \(S^*_n\) whether the components \(S^1_n\) and \(S^2_n\) are independent or not. For \(S^\circ_n\), we used the independence of \(S^1_n\) and \(S^2_n\) to simplify our computation, but the process can be generalised to the dependent case by approximating the distribution of the maximum of two dependent quadratic forms of normal random variables. The computations involved in the approximation techniques are extremely fast. As mentioned in the paper, the computation involved in the multipliers bootstrap are quite fast since the parameters are estimated only once and that each bootstrap iteration just requires generating \(n\) i.i.d centred normal random variables with variance one. The combined statistics \(S^*_n\), \(S^\circ_n\) and \(S^\bullet_n\) have similar power behaviour in general, but \(S^*_n\) is much easier to compute and its \(p\)-values are easier to obtain. We, therefore, recommend practitioners to use tests based on the original process more often with a multipliers procedure or a numerical approximation technique. Among these statistics \(S^*_n\) would be the easiest to implement. The tests introduced in this manuscript are in general a bit more powerful than those in Escanciano (2010), but there are alternatives, such as A1, where tests in Escanciano (2010) clearly outperform those discussed here.

The procedures discussed here may be generalise to the case of multivariate time series where \(X_t \in \mathbb{R}^d\). The procedure based on the multipliers bootstrap would be the easiest to generalise. In fact, it will not require any modification it suffices to adjust the definition of the process and the functions \(\phi^*\) and \(\Gamma_{\theta_0}\) to the multivariate case. The generalisation of the numerical approximation and that of the martingale transform would need more technical work.

**Acknowledgements**

We thank the editor, associate editor and the two referees for their valuable and constructive comments which helped improve the manuscript substantially.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work received funding from CY Advanced Studies of CY Cergy Paris University and was partly developed during author’s stay at CY Advanced Studies whose support is gratefully acknowledged. The work is also funded in part by the United Arab Emirates University UPAR grant program.
References

Aït-Sahalia, Y. (1996), 'Testing Continuous-Time Models of the Spot Interest Rate', *Review of Financial Studies*, 9(2), 385–426.

Aït-Sahalia, Y. (1999), 'Transition Densities for Interest Rate and Other Nonlinear Diffusions', *The Journal of Finance*, 54(4), 1361–1395.

Aït-Sahalia, Y. (2002), 'Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach', *Econometrica*, 70(1), 223–262.

Andrews, D.W. (1992), 'Generic Uniform Convergence', *Econometric Theory*, 8(2), 241–257.

Auestad, B., and Tjøstheim, D. (1990), 'Identification of Nonlinear Time Series: First Order Characterization and Order Determination', *Biometrika*, 77(4), 669–687.

Bachelier, L (1900), 'Théorie de la Spéculation', Annales Scientifiques de l’École Normale Supérieure, 17, 21–86. doi:10.24033/asens.476.

Bai, J. (2003), 'Testing Parametric Conditional Distributions of Dynamic Models', *Review of Economics and Statistics*, 85(3), 531–549.

Bai, J., and Ng, S. (2001), 'A Consistent Test for Conditional Symmetry in Time Series Models', *Journal of Econometrics*, 103(1–2), 225–258.

Benjamini, Y., and Hochberg, Y. (1995), 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.

Black, F., and Scholes, M. (1973), 'The Pricing of Options and Corporate Liabilities', *Journal of Political Economy*, 81(3), 637–654.

Chan, K.C., Karolyi, G.A., Longstaff, F.A., and Sanders, A.B. (1992), 'An Empirical Comparison of Alternative Models of the Short-Term Interest Rate', *The Journal of Finance*, 47(3), 1209–1227.

Chen, M., and An, H.Z. (1997), 'A Kolmogorov–Smirnov Type Test for Conditional Heteroskedasticity in Time Series', *Statistics & Probability Letters*, 33(3), 321–331.

Chen, S.X., and Gao, J. (2011), 'Simultaneous Specification Testing of Mean and Variance Structures in Nonlinear Time Series Regression', *Econometric Theory*, 27(4), 792–843.

Chen, Q., Zheng, X., and Pan, Z. (2015), 'Asymptotically Distribution-Free Tests for the Volatility Function of a Diffusion', *Journal of Econometrics*, 184(1), 124–144.

Cox, J.C., Ingersoll Jr, J.E., and Ross, S.A. (1980), 'An Analysis of Variable Rate Loan Contracts', *The Journal of Finance*, 35(2), 389–403.

Deheuvels, P., and Martynov, G.V. (1996), 'Cramér-von Mises-Type Tests with Applications to Tests of Independence for Multivariate Extreme-Value Distributions', *Communications in Statistics – Theory and Methods*, 25(4), 871–908.

Dette, H., Neumeyer, N., and Kiefer, J.V. (2007), 'A New Test for the Parametric Form of the Variance Function in Non-parametric Regression', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5), 903–917.

Diebolt, J. (1990), 'Testing the Functions Defining a Nonlinear Autoregressive Time Series', *Stochastic Processes and Their Applications*, 36(1), 85–106.

Drost, F.C., and Nijman, T.E. (1993), 'Temporal Aggregation of GARCH Processes', *Econometrica*, 61(4), 909–927.

Durlauf, S.N. (1991), 'Spectral Based Testing of the Martingale Hypothesis', *Journal of Econometrics*, 50(3), 355–376.

Escanciano, J.C. (2007a), 'Model Checks Using Residual Marked Empirical Processes', *Statistica Sinica*, 17(1), 115–138.

Escanciano, J.C. (2007b), 'Weak Convergence of Non-stationary Multivariate Marked Processes with Applications to Martingale Testing', *Journal of Multivariate Analysis*, 98(7), 1321–1336.
Escanciano, J.C. (2008), ‘Joint and Marginal Specification Tests for Conditional Mean and Variance Models’, *Journal of Econometrics*, 143(1), 74–87.

Escanciano, J.C. (2010), ‘Asymptotic Distribution-Free Diagnostic Tests for Heteroskedastic Time Series Models’, *Econometric Theory*, 26(3), 744–773.

Escanciano, J.C., and Mayoral, S. (2010), ‘Data-Driven Smooth Tests for the Martingale Difference Hypothesis’, *Computational Statistics & Data Analysis*, 54(8), 1983–1998.

Escanciano, J.C., Pardo-Fernández, J.C., and Van Keilegom, I. (2018), ‘Asymptotic Distribution-Free Tests for Semiparametric Regressions with Dependent Data’, *The Annals of Statistics*, 46(3), 1167–1196.

Fisher, R.A. (1950), *Statistical Methods for Research Workers* (11th ed.), London: Oliver and Boyd.

Genest, C., and Rémillard, B. (2004), ‘Tests of Independence Or Randomness Based on the Empirical Copula Process’, *Test*, 13, 335–369.

Ghoudi, K., Kulperger, R.J., and Rémillard, B. (2001), ‘A Nonparametric Test of Serial Independence for Time Series and Residuals’, *Journal of Multivariate Analysis*, 79, 191–218.

González-Manteiga, W., and Crujeiras, R.M. (2013), ‘An Updated Review of Goodness-of-Fit Tests for Regression Models’, *Test*, 22(3), 361–411.

Hafner, C.M., and Kyriakopoulou, D. (2021), ‘Exponential-Type Garch Models with Linear-in-Variance Risk Premium’, *Journal of Business & Economic Statistics*, 39(2), 589–603.

Imhof, J.-P. (1961), ‘Computing the Distribution of Quadratic Forms in Normal Variables’, *Biometrika*, 48(3/4), 419–426.

Jacod, J., and Protter, P. (1998), ‘Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations’, *The Annals of Probability*, 26(1), 267–307.

Khmaladze, E. (1988), ‘An Innovation Approach to Goodness-of-Fit Tests in $r^m$’, *The Annals of Statistics*, 16(4), 1503–1516.

Koul, H.L., and Song, W. (2010), ‘Conditional Variance Model Checking’, *Journal of Statistical Planning and Inference*, 140(4), 1056–1072.

Koul, H.L., and Stute, W. (1999), ‘Nonparametric Model Checks for Time Series’, *The Annals of Statistics*, 27(1), 204–236.

Laïb, N. (1999), ‘Nonparametric Testing for Correlation Models with Dependent Data’, *Journal of Nonparametric Statistics*, 12(1), 53–82.

Laïb, N. (2003), ‘Non-parametric Testing of Conditional Variance Functions in Time Series’, *Australian & New Zealand Journal of Statistics*, 45(4), 461–475.

Laïb, N., and Chebana, F. (2011), ‘A Simultaneous Test for Conditional Mean and Conditional Variance Functions in Time Series Models with Martingale Difference Innovations’, *Statistical Methodology*, 8(2), 221–241.

Laïb, N., and Louani, D. (2002), ‘On the Conditional Homoscedasticity Test in Autoregressive Model with Arch Error’, *Communications in Statistics – Theory and Methods*, 31(7), 1179–1202.

McKeague, I.W., and Zhang, M. -J. (1994), ‘Identification of Nonlinear Time Series From First Order Cumulative Characteristics’, *The Annals of Statistics*, 22(1), 495–514.

Merton, R.C (1970), ‘A Dynamic General Equilibrium Model of the Asset Market and Its Application to the Pricing of the Capital Structure of the Firm’, *Sloan School of Management Working Paper*, No. 497-70.

Ngatchou-Wandji, J. (2002), ‘Weak Convergence of Some Marked Empirical Processes: Application to Testing Heteroscedasticity’, *Journal of Nonparametric Statistics*, 14(3), 325–339.

Pardo-Fernández, J.C., Jiménez-Gamero, M.D., and El Ghouch, A. (2015), ‘Tests for the Equality of Conditional Variance Functions in Nonparametric Regression’, *Electronic Journal of Statistics*, 9(2), 1826–1851.

Polonik, W., and Yao, Q. (2008), ‘Testing for Multivariate Volatility Functions Using Minimum Volume Sets and Inverse Regression’, *Journal of Econometrics*, 147(1), 151–162.

Shorack, G., and Wellner, J. (1986), *Empirical Processes with Applications to Statistics*, John Wiley & Sons.

Simes, R.J. (1986), ‘An Improved Bonferroni Procedure for Multiple Tests of Significance’, *Biometrika*, 73(3), 751–754.

Stute, W. (1997), ‘Nonparametric Model Checks for Regression’, *The Annals of Statistics*, 25(2), 613–641.
Stute, W., Quindimil, M.P., Manteiga, W.G., and Koul, H. (2006), ‘Model Checks of Higher Order Time Series’, Statistics & Probability Letters, 76(13), 1385–1396.
Stute, W., Thies, S., and Zhu, L.-X. (1998), ‘Model Checks for Regression: An Innovation Process Approach’, The Annals of Statistics, 26(5), 1916–1934.
Vasicek, O. (1977), ‘An Equilibrium Characterization of the Term Structure’, Journal of Financial Economics, 5(2), 177–188.
Wang, L., and Zhou, X.-H. (2005), ‘A Fully Nonparametric Diagnostic Test for Homogeneity of Variances’, Canadian Journal of Statistics, 33(4), 545–558.
Wooldridge, J.M. (1990), ‘A Unified Approach to Robust, Regression-Based Specification Tests’, Econometric Theory, 6(1), 17–43.

Appendices

Appendix 1. Proofs

We start by proving the next Lemma which is used repeatedly in the proofs. It establishes a uniform law of large number result needed to show that the convergence is uniform in $x$ for both Theorems 3.2 and 4.1.

Lemma A.1: If $(X_1, ..., n, \text{is strictly stationary and ergodic series satisfying } \mathbb{E}|X_1| < \infty$ and if $(X_i)$ is strictly stationary and ergodic series then, for $n \to \infty$, we have

(i) $\sup_{x \in \mathbb{R}} \left\{ \frac{1}{n} \sum_{i=1}^{n} X_i \mathbb{I}(X_i - 1 \leq x) - \mathbb{E}[X_1 \mathbb{I}(X_0 \leq x)] \right\}$ converges to zero a.s.,

(ii) If $g_k$ and $A_k$ satisfy the conditions of Theorem 3.2 then

\[
\sup_{x \leq x_0} \left\| \frac{1}{n} \sum_{i=1}^{n} X_i \int_{-\infty}^{\min(x, X_i-1)} (g_k(y))^\top A_k^{-1}(y) K_k(dy) \right\|
\]

converges to zero a.s. for $k = 1, 2$ and for every $x_0$ that guarantees the invertibility of $A_k(x_0)$.

Proof: The proof of (i) and (ii) are quite similar. Here, only the proof of (ii) is given. For (i) one repeats the same steps. One can also see that (i) is similar to (Koul and Stute 1999, Eq. 4.1). Though (ii) can be deduced from of ULLN (cf. Andrews 1992) if $x$ belongs to a compact set. Since this is not the case here, a direct proof using a Glivenko–Cantelli type argument shall be given next. First, recall that $A_k(x)$ is invertible for all $x \leq x_0$ whenever $A_k(x_0)$ is invertible and that one can easily verify that $\|A_k^{-1}(x)\| \leq \|A_k^{-1}(x_0)\| < \infty$. One also sees that

\[
\left\| \int_{-\infty}^{\min(x, X_0)} (g_k(y))^\top A_k^{-1}(y) K_k(dy) \right\| \leq \left[ \int_{-\infty}^{\min(x, X_0)} \|(g_k(y))^\top K_k(dy) \right] \|A_k^{-1}(x_0)\| < \infty,
\]

since $g_k$ satisfies the conditions of Theorem 3.2. The LLN for stationary ergodic sequence yields the almost sure convergence of ii) for every fixed $x \leq x_0$. To prove that the convergence is uniform in $x \leq x_0$ one uses a Glivenko–Cantelli type argument applied using $\eta^*(x) = \mathbb{E}[X_1 \int_{-\infty}^{\min(x, X_0)} \|(g_k(y))^\top A_k^{-1}(y) K_k(dy) \]$. Note that $\eta^*$ is a continuous increasing function and for all $x \leq x_0$ one has $\eta^*(x) \leq \eta^*(x_0) < \infty$. Therefore for $\varepsilon > 0$ there exist a finite partition $-\infty = t_0 < t_1 < \ldots t_k = x_0$ such that $0 \leq \eta^*(t_{j+1}) - \eta^*(t_j) \leq \varepsilon$. To ease presentation, let

\[
\eta_n(x) = \frac{1}{n} \sum_{i=1}^{n} X_i \int_{-\infty}^{\min(x, X_i-1)} (g_k(y))^\top A_k^{-1}(y) K_k(dy) \]

and
Straightforward computations show the covariance function of
\( \| \eta_n(x) - \eta_n(x_0) \| + \| \eta_n(t_j) - \eta(t_j) \| + \| \eta(t_j) - \eta(x) \| \). Using its definition one sees that
\[
\| \eta_n(x) - \eta_n(t_j) \| \leq \left| \frac{1}{n} \sum_{i=1}^{n} |x_i| \right| \int_{-\infty}^{\min(t_{j+1}, x_{i-1})} (g_k(y))^{\top} A_k^{-1}(y) K_k(dy)
\]
\[
- \eta^*(t_{j+1}) + \eta^*(t_j) \right| + |\eta^*(t_{j+1}) - \eta^*(t_j)|.
\]
One also easily verifies that \( \| \eta(t_j) - \eta(x) \| \leq |\eta^*(t_{j+1}) - \eta^*(t_j)| \). Therefore
\[
\sup_{x \leq x_0} |\eta_n(x) - \eta(x)| \leq \| \eta_n(t_j) - \eta(t_j) \| + \varepsilon + \max_{j=0, \ldots, k} \left| \frac{1}{n} \sum_{i=1}^{n} |x_i| \right|
\]
\[
\times \int_{\min(t_{j+1}, x_{i-1})}^{\min(t_{j}, X_{i-1})} \| (g_k(y))^{\top} A_k^{-1}(y) \| K_k(dy) - \eta^*(t_{j+1}) + \eta^*(t_j) \right|.
\]
The pointwise LLN and the fact that \( k \) is finite imply that the first and last terms in the above inequality converge almost surely to zero. Since \( \varepsilon \) was arbitrary, one concludes that \( \sup_{x \leq x_0} |\eta_n(x) - \eta(x)| \) converges to zero almost surely.

### A.1 Proof of Theorem 2.1

The weak convergence of \( D_n \) follows from Theorem 1 in Escanciano (2007b). For \( \hat{D}_n \), direct manipulations show that
\[
\hat{D}_n(x) = D_n(x) - \Gamma_{\theta_0}^{\top}(x) \sqrt{n}(\theta_n - \theta_0) + B_{n,1}(x) + B_{n,2}(x),
\]
where
\[
B_{n,1}(x) = \left[ \frac{1}{n} \sum_{i=1}^{n} W_{\theta_0}^{\top}(X_i, I_{i-1}) I[X_{i-1} \leq x] - \Gamma_{\theta_0}^{\top}(x) \right] \sqrt{n}(\theta_n - \theta_0),
\]
and
\[
B_{n,2}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ W_{\theta_n}(X_i, I_{i-1}) - W_{\theta_0}(X_i, I_{i-1}) - W_{\theta_0}^{\top}(X_i, I_{i-1})(\theta_n - \theta_0) \right] I[X_{i-1} \leq x].
\]
Note that \( \|B_{n,1}\| \) converges in probability to zero by Lemma A.1, the definition of \( \Gamma_{\theta_0} \), and the fact that \( \sqrt{n}(\theta_n - \theta_0) = O_p(1) \). By Assumption (A3), the term \( \|B_{n,2}\| \) is bounded by \( \sqrt{n}\|\theta_n - \theta_0\|\lambda_1(\|\theta_n - \theta_0\|) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_1(X_i, I_{i-1}) \) which converges to zero in probability by the LLN and the fact that \( \sqrt{n}\|\theta_n - \theta_0\| = O_p(1) \). Therefore, \( \hat{D}_n(x) \) is asymptotically equivalent to \( D_n(x) - \Gamma_{\theta_0}^{\top}(x) \sqrt{n}(\theta_n - \theta_0) \). Calling on Assumption (A2), one verifies that \( \sqrt{n}(\theta_n - \theta_0) \) is tight and converges to \( \Theta \) and that \( (D_n, \sqrt{n}(\theta_n - \theta_0)) \) converge jointly to \( (D, \Theta) \). Hence \( \hat{D}_n \) converges to \( \hat{D} = D - \Gamma_{\theta_0}^{\top}(x)\Theta \). Straightforward computations show the covariance function of \( \hat{D} \) is precisely given by (2).

### A.2 Proof of Theorem 2.2

The proof follows the same approach as that of Theorem 2.1. Precisely, one writes
\[
\hat{D}_n(x) = D_n(x) + B_{n,0}(x) - B_{n,1}(x) + B_{n,2}(x),
\]
where

\[
B_{n,0}(x) = \frac{1}{n} \sum_{i=1}^{n} \left( \begin{array}{c} a_1(I_{i-1}) \\ a_2(I_{i-1}) \end{array} \right) \mathbb{I}[X_{i-1} \leq x],
\]

\[
B_{n,1}(x) = \left[ \frac{1}{n} \sum_{i=1}^{n} \hat{W}_{\theta_0}^\top(X_i, I_{i-1}) \mathbb{I}[X_{i-1} \leq x] \right] \sqrt{n}(\theta_n - \theta_0),
\]

and

\[
B_{n,2}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ W_{\theta_0}(X_i, I_{i-1}) - W_{\theta_0}(X_i, I_{i-1}) - \hat{W}_{\theta_0}^\top(X_i, I_{i-1})(\theta_n - \theta_0) \right] \mathbb{I}[X_{i-1} \leq x].
\]

\(B_{n,0}\) converges uniformly to \(\Psi_A(x)\) by Lemma A.1. Assumption (L1) and Lemma A.1, yield that \(B_{n,1}\) converges to \(\Gamma_\theta^\top(x)(\Theta + \xi_A)\). From Assumption (A3), one concludes that the term \(B_{n,2}\) is uniformly bounded by \(\sqrt{n}\|\theta_n - \theta_0\|\lambda_1(\|\theta_n - \theta_0\|) \sum_{i=1}^{n} M_1(X_i, I_{i-1})\) which goes to zero in probability since 
\(\mathbb{E}[M_1(X_i, I_{i-1})] < \infty\) and by Assumption (L3), \(\sqrt{n}\|\theta_n - \theta_0\| = O_P(1)\) and \(\|\theta_n - \theta_0\| = o_P(1)\).

**A.3 Proof of Theorem 3.2**

The proof of Theorem 3.2 is as follows. First, in Lemma A.2 we establish, for \(k = 1, 2\), that \(T_k^k(\hat{D}_n^k)\) is asymptotically equivalent to \(T_k(\hat{D}_n)\). Second, using the continuous mapping theorem, one concludes that \(T_k(\hat{D}_n^k)\) converges to \(T_k(\hat{D}^k)\). Then the proof is completed by showing that \(T_k(\hat{D}_n^k)\) is equal to \(T_k(\hat{D}^k)\) which has the same law as \(\hat{D}^k\).

**Lemma A.2:** Under the assumptions of Theorem 3.2, for \(k = 1, 2\), one gets

\[
\sup_{x \leq x_0} \|T_k^k(\hat{D}_n^k)(x) - T_k(\hat{D}_n^k)(x)\| \text{ converges to zero in probability.}
\]

**Proof:** To prove the Lemma, observe that

\[
T_k^k(\hat{D}_n^k)(x) - T_k(\hat{D}_n^k)(x) = \int_{-\infty}^{x} \hat{g}_k^\top(y) A_k^{-1}(y) \left( \int_{y}^{\infty} \hat{g}_k(z) \hat{D}_n^k(dz) \right) K_k(dy) - \int_{-\infty}^{x} \hat{g}_k^\top(y) \hat{A}_k^{-1}(y) \left( \int_{y}^{\infty} \hat{g}_k(z) \hat{D}_n^k(dz) \right) \hat{K}_k(dy).
\]

First, one establishes that \(\psi^k_n\) is tight and that \(\sup_y |\psi^k_n(y) - \hat{\psi}^k_n(y)|\) converges to zero in probability, where \(\psi^k_n(y) = \int_{y}^{\infty} g_k(z) \hat{D}_n^k(dz)\) and \(\hat{\psi}^k_n(y) = \int_{y}^{\infty} \hat{g}_k(z) \hat{D}_n^k(dz)\). For the tightness of \(\psi^k_n\), set \(\psi^0_n(y) = \int_{y}^{\infty} g_k(z) \hat{D}_n(dz) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{\theta_0}(X_i, I_{i-1}) g_k(X_{i-1}) \mathbb{I}[X_{i-1} > y]\). Note that \(\psi^0_n\) is a marked empirical process its tightness follows from Escanciano (2007b) and Assumption (K4). Next, using the same decomposition as in the proof of Theorem 2.1, one sees that

\[
\sup_{x \in \mathbb{R}} |\psi^k_n(x) - \psi^0_n(x) + \frac{1}{n} \sum_{i=1}^{n} g_k(X_{i-1}) \hat{W}_{\theta_0}(X_i, I_{i-1}) \mathbb{I}[X_{i-1} > x] \sqrt{n}(\theta_n - \theta)| = o_P(1).
\]

The above statement with the tightness of \(\psi^0_n\) and assumptions (A3-A4, K4) imply the tightness of \(\psi^k_n\). To give an upper bound of \(\sup_y |\psi^k_n(y) - \hat{\psi}^k_n(y)|\), use the decomposition in the proof of Theorem 2.1, which shows that

\[
\sup_{y} |\psi^k_n(y) - \hat{\psi}^k_n(y)|
\]
which converges to zero by Assumptions (A1)–(A4), (K1)–(K4) and the LLN.

Next, it will be shown that $\sup_{x} \| \mathcal{A}_k(x) - A_k(x) \|$ converges to zero in probability. Observe that

$$
\| \mathcal{A}_k(x) - A_k(x) \| = \left\| \int_{\mathbb{R}} \hat{g}_k(t) g_k^\top(t) K_k(dt) - \int_{\mathbb{R}} \hat{g}_k(t) \hat{g}_k^\top(t) \hat{K}_k(dt) \right\|
$$

The first term above is equal

$$
\left\| \frac{1}{n} \sum_{i=1}^{n} \hat{g}_k(X_{i-1}) g_k^\top(X_{i-1}) W_{\theta_0}^k(X_i, I_{i-1}) \mathbb{I}[X_{i-1} > x] - \int_{\mathbb{R}} g_k(t) \hat{g}_k^\top(t) K_k(dt) \right\|
$$

which converges to zero a.s. by Lemma (A.1) and Assumption (K2).

The second term is bounded by $(\| \hat{g}_k - g_k \| + \| g_k \|) \| \hat{g}_k - g_k \| \int_{\mathbb{R}} \hat{K}_k(dt)$ which goes to zero in probability by Assumption (K3).

As pointed earlier, for all $x \leq x_0$ one has $A_k(x) - A_k(x_0)$ non-negative definite implying that $A_k(x)$ is invertible whenever $A_k(x_0)$ is invertible and that $\| A_k^{-1}(x) \| \leq \| A_k^{-1}(x_0) \| < \infty$. Using the above result for $\hat{A}_k - A_k$ and classical algebraic manipulations one also easily see that

$$
\sup_{x \leq x_0} \| \hat{A}_k^{-1}(x) - A_k^{-1}(x) \| \overset{p}{\to} 0.
$$

To complete the proof of the Lemma A.2 note that

$$
T_n^k(\hat{\mathbb{D}}^k_n)(x) - T^k(\hat{\mathbb{D}}^k_n)(x) = \int_{-\infty}^{x} g_k^\top(y) A_k^{-1}(y) \psi_n^k(y) K_k(dy) - \int_{-\infty}^{x} \hat{g}_k^\top(y) \hat{A}_k^{-1}(y) \hat{\psi}_n^k(y) \hat{K}_k(dy).
$$

Direct computations show that

$$
T_n^k(\hat{\mathbb{D}}^k_n)(x) - T^k(\hat{\mathbb{D}}^k_n)(x) = \int_{-\infty}^{x} g_k^\top(y) A_k^{-1}(y) \psi_n(y)(K_k(dy) - \hat{K}_k(dy))
$$

$$
+ \int_{-\infty}^{x} \left[ g_k^\top(y) A_k^{-1}(y) \psi_n^k(y) - \hat{g}_k^\top(y) \hat{A}_k^{-1}(y) \hat{\psi}_n^k(y) \right] \hat{K}_k(dy).
$$

Hence

$$
\sup_{x \leq x_0} | T_n^k(\hat{\mathbb{D}}^k_n)(x) - T^k(\hat{\mathbb{D}}^k_n)(x) |
$$

$$
\leq \sup_{x \leq x_0} \left| \int_{-\infty}^{x} g_k^\top(y) A_k^{-1}(y) \psi_n(y)(K_k(dy) - \hat{K}_k(dy)) \right|
$$

$$
+ \sup_{x \leq x_0} \left| \int_{-\infty}^{x} \left[ g_k^\top(y) A_k^{-1}(y) \psi_n^k(y) - \hat{g}_k^\top(y) \hat{A}_k^{-1}(y) \hat{\psi}_n^k(y) \right] \hat{K}_k(dy) \right|.
$$
Using Lemma (4.1) in Koul and Stute (1999) and the fact that \( \psi_n \) is tight, one concludes that the first term in the above expression goes to zero in probability. Term 2 is bounded by \( \sum_{\ell=1}^3 \tilde{C}_{\ell,n} \) where

\[
\tilde{C}_{1,n} = \| \hat{\psi}_n^k - \psi_n^k \| \sup_{x \leq x_0} \| A_k^{-1}(x) \| \int_{-\infty}^{\infty} \| \hat{g}_k^\top(y) \| \hat{K}_k(dy),
\]

\[
\tilde{C}_{2,n} = \sup_{x \leq x_0} \| \hat{A}_k^{-1}(x) - A_k^{-1}(x) \| (\| \hat{\psi}_n^k - \psi_n^k \| + \| \psi_n^k \|) \int_{-\infty}^{\infty} \| \hat{g}_k^\top(y) \| \hat{K}_k(dy).
\]

and

\[
\tilde{C}_{3,n} = \sup_{x \leq x_0} (\| \hat{A}_k^{-1}(x) - A_k^{-1}(x) \| + \| A_k^{-1}(x) \| (\| \hat{\psi}_n^k - \psi_n^k \| + \| \psi_n^k \|)) \int_{-\infty}^{\infty} \| \hat{g}_k^\top(y) - g_k^\top \| \hat{K}_k(dy).
\]

Since \( \psi_n^k \) is tight and \( \| \psi_n^k \| = O_p(1) \), the proof is complete by noting that, using the assumptions and the above results, \( \tilde{C}_{\ell,n} \) for \( \ell = 1, \ldots, 3 \) converge to zero in probability.

The transformation \( T^k \) is linear and continuous which implies that \( T^k(\tilde{D}^k_n) \) converges weakly to \( T^k(\mathbb{D}^k) = T^k(\mathbb{D}^k - \Theta^\top \Gamma_{\theta_0}^k) = T^k(\Gamma_{\theta_0}^k) = T^k(\mathbb{D}^k) \) since direct computations show that \( T^k(\Gamma_{\theta_0}^k) = 0 \). Straightforward computations, (similar to those in Stute et al. 1998), enable us to verify that \( T^k(\mathbb{D}^k) \) has the same distribution as \( \mathbb{D}^k = \mathbb{W}_k \circ K_k \) where \( \mathbb{W}_k \) is standard Brownian motion.

### A.4 Proof of Theorem 4.1

Theorem 4.1 will be shown by establishing parts (a) and (b) below

(a) \( \tilde{D}_{n,b} \) is asymptotically equivalent to \( \mathbb{D}^*_{n,b} \)

(b) \( \mathbb{D}^*_{n,b} \) converge weakly to independent copies of \( \mathbb{D} \)

#### Proof of (a): 

Straightforward computations show that \( \tilde{D}^k_n(x) - \mathbb{D}^*_{n,b}(x) = \sum_{k=1}^5 R_{n,k}(x) \), where

\[
R_{n,1}(x) = \left( \frac{1}{n} \sum_{i=1}^n Z_{i,b} \hat{W}_b(X_i, I_{i-1}) \I_{|X_{i-1}| \leq x} \right) \sqrt{n} (\theta_n - \theta_0),
\]

\[
R_{n,2}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,b} \left[ W_b(X_i, I_{i-1}) - W_b(X_i, I_{i-1}) - \hat{W}_b(X_i, I_{i-1}) (\theta_n - \theta_0) \right] \I_{|X_{i-1}| \leq x},
\]

\[
R_{n,3}(x) = [\Gamma_{\theta_n}(x) - \Gamma_{\theta_0}(x)]^\top \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,b} \phi^*(X_i, X_{i-1}, \theta_0) \right),
\]

\[
R_{n,4}(x) = \Gamma_{\theta_0}^\top(x) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,b} \phi^*(X_i, X_{i-1}, \theta_0) ^\top (\theta_n - \theta_0), \text{ and}
\]

\[
R_{n,5}(x) = \Gamma_{\theta_0}^\top(x) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,b} [\phi^*(X_i, X_{i-1}, \theta_n) - \phi^*(X_i, X_{i-1}, \theta_0) - \phi^*(X_i, X_{i-1}, \theta_0) ^\top (\theta_n - \theta_0)].
\]

The term \( R_{n,1} \) converges to zero in probability by Lemma A.1, while term \( R_{n,2} \) converges in probability to zero uniformly for all \( x \in \mathbb{R} \) following the same steps as for the term \( B_{n,2} \) in the proof of Theorem 2.1. By Assumption (M1), \( \sup_x \| \Gamma_{\theta_n}(x) - \Gamma_{\theta_0}(x) \| \leq \| \theta_n - \theta_0 \| C \) which goes to zero in probability. One also sees that \( \sup_x \| \Gamma_0(x) \| \leq \mathbb{E} \| W_0(X_i, I_{i-1}) \| < \infty \). Next, since \( \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{i,b} \phi^*(X_i, X_{i-1}, \theta_0) = O_p(1) \) by the multiplier central limit theorem, then the term \( R_{n,3} \)
converge uniformly to zero in probability. On the other side
\[ \| R_{n,1} \| \leq \| \theta_n - \theta_0 \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,b} \hat{\phi}^*(X_i, X_{i-1}, \theta_0)^\top \left( \| \Gamma_{\theta_n} - \Gamma_{\theta_0} \| + \| \Gamma_{\theta_0} \| \right) = o_P(1), \]

since \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{i,b} \hat{\phi}^*(X_i, X_{i-1}, \theta_0) = O_P(1) \) by Assumption (M3). Finally, by Assumption (M2),
\[ \| R_{n,5} \| \leq \frac{1}{\sqrt{n}} \| \theta_n - \theta_0 \| \lambda_3 (\| \theta_n - \theta_0 \|) \left( \frac{1}{n} \sum_{i=1}^{n} |Z_{i,b}| M_4(X_i, X_{i-1}) \right) \]
\[ \times \left( \| \Gamma_{\theta_n} - \Gamma_{\theta_0} \| + \| \Gamma_{\theta_0} \| \right) = o_P(1). \]
Combining these results, yields \( \sup\| \mathbb{D}_{n,b}^*(x) - \mathbb{D}_{n,b}^*(x) \| = o_P(1) \) which completes the proof of part (a).

**Proof of (b):**
One notices that the multiplier central limit theorem yields the weak convergence of \( \mathbb{D}_{n,b}^* \). It just remain to show that the asymptotic covariance operator of the process \( \mathbb{D}_{n,b}^* \) is the same as that of the limiting process \( \mathbb{D}_{\cdot} \).

Observe that
\[ \text{Cov} \left( \mathbb{D}_{n,b}^*(x), \mathbb{D}_{n,b}^*(y) \right) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{E}(V_i(x) V_j(y)) \text{E}(Z_{i,b} Z_{j,b}), \]
where \( V_i(x) = W_{\theta_0}(X_i, I_{i-1}) \mathbb{1}[X_{i-1} \leq x] - \phi^*(X_i, I_{i-1}, \theta_0)^\top \Gamma_{\theta_0}(x) \). One has only to consider the terms with \( i \neq j \) in the above sums, since for \( i = j, \text{E}(Z_{i,b} Z_{j,b}) = 0 \). Recalling that \( \text{E}(\|Z_{i,b}\|^2) = 1 \), one sees that the above covariance reduces to \( \text{Cov}(\mathbb{D}_{n,b}^*(x), \mathbb{D}_{n,b}^*(y)) = \frac{1}{n} \sum_{i=1}^{n} \text{E}(V_i(x) V_i(y)) = \text{E}(V_1(x) V_1(y)) \). Straightforward computations show that \( \text{E}(V_1(x) V_1(y)) = K(x, y) - \Gamma_{\theta_0}^\top(x) G(y) - G^\top(x) \Gamma_{\theta_0}(y) + \Gamma_{\theta_0}^\top(x) \Sigma_0 \Gamma_{\theta_0}(y) \), which is the same as the covariance of the process \( \mathbb{D}_{\cdot} \).

**Appendix 2. Implementation of tests based on Khmaladze transform**
Here, we outline the key steps in the algorithm used to compute the test statistics based on Khmaladze transform identified in Equation (7). The only challenging step in (7) is the computation of \( T_n^k(\hat{\mathbb{D}}_{\cdot}) \). Note that Equation (4) shows that to find \( T_n^k(\hat{\mathbb{D}}_{\cdot}) \) we need \( \hat{g}_k, \hat{A}_k \) and \( \hat{K}_k \). An estimate of \( \hat{K}_k \) is given by Equation (6). To obtain \( \hat{g}_k \), the estimate of the function \( g_k \), we first estimated the functions \( \Gamma_{\theta_0}^1(x) \) and \( \Gamma_{\theta_0}^2(x) \) using \( \Gamma_{\theta_0}^1(x) = \frac{1}{n} \sum_{i=1}^{n} \hat{m}(\theta_n, \tilde{I}_{i-1}) \mathbb{1}[X_{i-1} \leq x] \) and \( \Gamma_{\theta_0}^2(x) = \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}^2(\theta_n, \tilde{I}_{i-1}) \mathbb{1}[X_{i-1} \leq x] \), respectively.

Then we used a histogram type estimate of \( \hat{d}^k \hat{K}_k \) for the estimation of \( \hat{g}_k \). That is \( \hat{g}_k(x) = \frac{\hat{K}_k(x+h_n)-\hat{K}_k(x-h_n)}{\hat{K}_k(x+h_n)-\hat{K}_k(x-h_n)} \), where \( h_n \) goes to zero. In our simulation \( h_n \propto n^{-1/3} \) was used. We also estimated \( \frac{d\hat{K}_k}{dF_n} = \hat{K}_k(x+h_n)-\hat{K}_k(x-h_n) \) and for any function \( \varphi \) we evaluated the integral
\[ \int \varphi(x) d\hat{K}_k = \int \varphi(x) \frac{d\hat{K}_k}{dF_n} dF_n = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i) \hat{K}_k(X_i+h_n) - \hat{K}_k(X_i-h_n) \]
\[ \frac{F_n(X_i+h_n) - F_n(X_i-h_n)}{F_n(X_i+h_n) - F_n(X_i-h_n)}. \]
In particular the estimate of \( \hat{A}_k(x) \) is
\[ \hat{A}_k(x) = \sum_{i=1}^{n} \mathbb{1}[X_i \geq x] \hat{g}_k(X_i) \hat{g}_k(X_i)^\top \frac{\hat{K}_k(X_i+h_n) - \hat{K}_k(X_i-h_n)}{F_n(X_i+h_n) - F_n(X_i-h_n)}. \]

Note that to compute (7), we only need to evaluate the functions \( \hat{g}_k(\cdot) \) and \( \hat{A}_k(\cdot) \) at the points \( X_i \) for \( i = 1, \ldots, n \). We also implemented kernel type estimators of \( g_k \). It turned out that the bandwidth choice is more important than the kernel/histogram choice. The choice of bandwidth seems to have an effect over the levels of these tests. The bandwidth selected in our experiments gave levels close to nominal values under the null hypothesis.