Testing for the martingale hypothesis using martingale difference divergence

Luca Rolla, Alessandro Giovannelli

March 7, 2022

- Given a real valued stationary time series $\{X_t\}_{t=-\infty}^{\infty}$ we consider testing the hypothesis that, almost surely

$$E\left(X_t|X_{t-1}, X_{t-2}, ...\right) = E\left(X_t\right) \tag{1}$$

where, by expressing the information set at time t as $I_t = \{X_t, X_{t-1}, ...\}$, this is equivalent to testing

$$H_0 : E\left[X_t - E\left(X_t\right)|I_{t-1}\right] = 0 \tag{2}$$

- From a nolinear time series perspective we need to define some analytical tools that make it possible to distinguish between a martingale difference sequence and a serially uncorrelated process by testing for the existence nonlinear dependence in the mean.
- We will show how it is possible to construct a test for the martingale difference hypothesis by using the recently introduced concept of martingale difference diveregence function (Shao and Zhang, 2014)

Introduction

- Distance covariance (Szekely et al. 2007) is a measure of dependence among random variables that is able to detect any form of statistical relationship existing among them, be it linear or nonlinear.
- In its standardized form, it can be regarded as a generalization of classical Pearson correlation coefficient capable of detecting any form of dependence existing between random vectors and not limited to the case of linear dependence.
- The concept of martingale difference divergence function (Shao and Zhang, 2014; Park et al., 2015), also a recently introduced statistic, largely shares its theoretical construction with that of distance covariance.

- Distance covariance and martingale difference divergence belong to the same family of statistics, with the difference that while distance covariance is able to detect any form of dependence between two random variables, martingale difference divergence focuses on the narrower definition of conditional mean dependence.
- We shall introduce a novel test for the martingale difference hypothesis based on the concept of martingale difference divergence.

The concept of distance covariance has been introduced by Szèkely, Rizzo and Bakirov (2007) as a measure of dependence between random vectors in terms of their respective characteristic functions. Given two random vectors $\mathbf{X} \in \mathbb{R}^q$ and $\mathbf{Y} \in \mathbb{R}^p$ with q and p positive integers, we denote their joint characteristic function by $\phi_{(\mathbf{X},\mathbf{Y})}(\mathbf{t},\mathbf{s})$ and the marginal characteristic functions by $\phi_{\mathbf{X}}(t)\,\phi_{\mathbf{Y}}(s)$ for parameters $(\mathbf{t},\mathbf{s}) \in \mathbb{R}^{p+q}$. In fact, characteristic functions can be used to characterise the null hypothesis of independence between $\mathbf{X}$ and $\mathbf{Y}$:

$$H_0 : \; \phi_{(\mathbf{X},\mathbf{Y})}(\mathbf{t},\mathbf{s}) = \phi_{\mathbf{X}}(\mathbf{t})\,\phi_{\mathbf{Y}}(\mathbf{s}) \tag{3}$$

If and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent

The statistic is expressed as a weighted L_2 distance between the joint characteristic function and the product of the marginal characteristic functions of the two random vectors **X** and **Y**

$$V^2(\mathbf{X},\mathbf{Y}) = \int_{\mathbb{R}^{p+q}} \left|\phi_{(\mathbf{X},\mathbf{Y})}(\mathbf{t},\mathbf{s}) - \phi_{\mathbf{X}}(\mathbf{t})\,\phi_{\mathbf{Y}}(\mathbf{s})\right|^2 \mathcal{W}(\mathbf{t},\mathbf{s})\, d\mathbf{t} d\mathbf{s}. \tag{4}$$

$$= ||\sigma(\mathbf{t},\mathbf{s})||^2_{\mathcal{W}} \tag{5}$$

where $\mathcal{W}(\mathbf{t},\mathbf{s})$is a pre-specified weighting function.

In particular, the following form for a non-integrable weight function is chosen

$$\mathcal{W}(\mathbf{t},\mathbf{s}) = \left(c_p c_q \left|\mathbf{t}\right|_p^{1+p} \left|\mathbf{s}\right|_q^{1+q}\right)^{-1}, \tag{6}$$

where $c_d = \pi^{(1+d)/2}/\Gamma\left((1+d)/2\right)$ with Γ denoting the Gamma function and $|x|_p$ denoting the euclidean norm in $\mathbb{R}^p$.

In more general terms, with the definition of distance covariance we are actually considering the weighted norm $||\sigma(\mathbf{t},\mathbf{s})||_{\mathcal{W}}$ of the generalized covariance

$$\begin{aligned}\sigma(\mathbf{t},\mathbf{s}) &= \phi_{(\mathbf{X},\mathbf{Y})}(\mathbf{t},\mathbf{s}) - \phi_{\mathbf{X}}(\mathbf{t})\,\phi_{\mathbf{Y}}(\mathbf{s}) \\ &= E\left[\left(\exp\left(i\left\langle \mathbf{t},\mathbf{X}\right\rangle\right) - \phi_{\mathbf{X}}(\mathbf{t})\right)\left(\exp\left(i\left\langle \mathbf{s},\mathbf{X}\right\rangle\right) - \phi_{\mathbf{Y}}(\mathbf{s})\right)\right],\end{aligned}$$

The standardized form of the statistic, called distance correlation, is defined as the nonnegative square root of the ratio between distance covariance and the product of the so-called distance variances, $V(\mathbf{X}, \mathbf{X})$ and $V(\mathbf{Y}, \mathbf{Y})$ as follows

$$R^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \frac{V^2(\mathbf{X},\mathbf{Y})}{\sqrt{V^2(\mathbf{X},\mathbf{X}) V^2(\mathbf{Y},\mathbf{Y})}}, & V^2(\mathbf{X}, \mathbf{X})\, V^2(\mathbf{Y}, \mathbf{Y}) > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

where $R(\mathbf{X}, \mathbf{Y}) \in [0, 1]$ under the assumption $E\left(|\mathbf{X}|_p + |\mathbf{Y}|_q\right) < \infty$, with $R^2(\mathbf{X}, \mathbf{Y}) = 0$ in the case of independence.

We shall recall the following lemma contained in Szekely et al. (2007):

Lemma (Szekely et al. (2007))

If $0 < \alpha < 2$, *then for all* x *in* $\mathbb{R}^d$

$$\int_{\mathbb{R}^d} \frac{1 - \cos\langle z, x\rangle}{|z|_d^{d+\alpha}} dt = C(d, \alpha)|x|^{\alpha},$$

where

$$C(d, \alpha) = \frac{2\pi\Gamma(1 - \alpha/2)}{\alpha 2^{\alpha}\Gamma((d + \alpha)/2)},$$

and $\Gamma(\cdot)$ *is the complete gamma function.*

This together with the assumptions made about the weight function allows them to find an explicit form for the integral.

If we assume that $E\left(|\mathbf{X}|_p^2 + |\mathbf{Y}|_q^2\right) < \infty$ then distance covariance takes the form

$$V^2(\mathbf{X},\mathbf{Y}) = E\left|\mathbf{X}-\mathbf{X}'\right|_p\left|\mathbf{Y}-\mathbf{Y}'\right|_q + E\left|\mathbf{X}-\mathbf{X}'\right|_p E\left|\mathbf{Y}-\mathbf{Y}''\right|_q - 2E\left|\mathbf{X}-\mathbf{X}'\right|_p \tag{8}$$

with $(\mathbf{X}',\mathbf{Y}')$and $(\mathbf{X}'',\mathbf{Y}'')$ being independent copies of $(\mathbf{X},\mathbf{Y})$.

This motivates the construction of the empirical version of the statistic. Given the random sample $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, .., n$, obtained from the joint distribution of the random vectors $\mathbf{X}$ and $\mathbf{Y}$, the sample distance covariance first the $n \times n$ Euclidean distance matrices with elements $(a_{ij}) = \left(\left|\mathbf{X}_i - \mathbf{X}_j\right|_p\right)$ and $(b_{ij}) = \left(\left|\mathbf{Y}_i - \mathbf{Y}_j\right|_q\right)$ are computed. The elements of the original euclidean distance matrices are transformed as follows

$$A_{ij} = a_{ij} - \overline{a}_{i.} - \overline{a}_{.j} + \overline{a}_{..}, \quad B_{ij} = b_{ij} - \overline{b}_{i.} - \overline{b}_{.j} + \overline{b}_{..} \tag{9}$$

Accordingly, the sample distance covariance is defined as the nonnegative square root of

$$\widehat{V}^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{i,j}^{n} A_{ij} B_{ij} \ . \tag{10}$$

If we assume that $E\left|\mathbf{X}\right|_q < \infty$ and $E\left|\mathbf{Y}\right|_p < \infty$ both $\widehat{V}^2(\mathbf{X}, \mathbf{Y})$ and $\widehat{R}^2(\ \mathbf{X}, \mathbf{Y})$ converge almost surely to their population counterparts.

As an extension of distance covariance, Shao and Zhang (2014) introduced the so-called martingale difference divergence For two random random vectors $\mathbf{X} \in R^q$ and $Y \in \mathbb{R}$ we want to measure the conditional mean independence of Y on $\mathbf{X}$, that is,

$$H_0 : E\left(Y|\mathbf{X}\right) = E\left(Y\right), \text{ almost surely.} \tag{11}$$

In particular, martingale difference divergence is defined as the nonnegative square root of

$$\begin{aligned} MDD^2\left(Y|\mathbf{X}\right) &= \int_{\mathbb{R}^q} \left|E\left(Y \cdot \exp\left(i\left\langle \mathbf{t}, \mathbf{X}\right\rangle\right)\right) - E\left(Y\right)\phi_{\mathbf{X}}\left(\mathbf{t}\right)\right|^2 \mathcal{W}\left(\mathbf{t}\right) d\mathbf{t} \quad (12) \\ &= \frac{1}{c_q}\int_{\mathbb{R}^q} \frac{\left|\sigma\left(\mathbf{t}\right)\right|^2}{\left|\mathbf{t}\right|_q^{1+q}} dt \\ &= \left|\left|\sigma\left(\mathbf{t}\right)\right|\right|_{\mathcal{W}}^2 \quad (13) \end{aligned}$$

where the nonintegrable weight function $\mathcal{W}\left(\cdot\right)$ is the same weight function defined by Szekely et al.(2007). The statistic will equal zero if and only if Y is conditionally mean independent with respect to the random vector $\mathbf{X}$.

Auto distance covariance

Zhou (2012) Auto-distance covariance (correlation) can be seen as an extension of classical auto-covariance (correlation) to the analysis of non-linear and non-gaussian time series: makes possible to distinguish between serially independent (i.i.d.) random variables and variables that are and uncorrelated but display some form of dependence in higher moments.

- Let $\{\mathbf{X}_t\}$ be a d-dimensional stationary time series: the auto-distance covariance at lag $j \in \mathbb{N}^+$ is defined as the nonnegative square root of the L_2 norm

$$
\begin{aligned}
V_{\mathbf{X}}^2(j) &= \int_{\mathbb{R}^{2d}} |\sigma_j(\mathbf{u}, \mathbf{v})|^2 \mathcal{W}(\mathbf{u}, \mathbf{v}) \, d\mathbf{u} d\mathbf{v} && (14) \\
&= \frac{1}{c_d^2} \int_{\mathbb{R}^{2d}} \frac{|\sigma_j(\mathbf{u}, \mathbf{v})|^2}{|\mathbf{u}|^{d+1} |\mathbf{v}|^{d+1}} d\mathbf{u} d\mathbf{v} \\
&= ||\sigma_j(\mathbf{u}, \mathbf{v})||_{\mathcal{W}}^2 && (15)
\end{aligned}
$$

- Since $\phi_j(u, v) = \phi(u)\phi(v)$ for all u, v if and only if $\{X_t\}$ and $\{X_{t-|j|}\}$ are independent, the covariance term $\sigma_j(u, v)$ equals zero if and only if the variable in consideration is independent from its past (at lag j) value.

In the univariate case, under the assumption that X_t is a strictly stationary α-mixing process and $E(X_t) < \infty$, Fokianos and Pitsillou (2017) proved the convergence almost surely of $\widehat{V}_X^2(\cdot)$ to its population counterpart for a fixed lag value j.

Auto martingale Difference Divergence

In the case of a d-dimensional stationary time series $\{\mathbf{X}_t\}$, the AMDD at lag $j \in \mathbb{N}$ is defined as follows

$$\begin{aligned} MDD_X^2(j) &= \int_{\mathbb{R}^d} |\sigma_j(\mathbf{v})|^2 \mathcal{W}(\mathbf{v})\, d\mathbf{v}, \\ &= \frac{1}{c_d} \int_{\mathbb{R}^d} \frac{|\sigma_j(\mathbf{v})|^2}{|\mathbf{v}|^{d+1}} d\mathbf{v} \qquad (16) \\ &= ||\sigma_j(v)||_{\mathcal{W}}^2 \qquad (17) \end{aligned}$$

where the generalized covariance term $\sigma_j(\mathbf{v})$ is defined in this case as $\sigma_j(v) = Cov(\mathbf{X}_t, e^{i\mathbf{v}\mathbf{X}_{t-j}})$.

Univariate case

From now on we will focus on the univariate time series case

$$V_X^2(j) = \int_{\mathbb{R}^2} |\sigma_j(u,v)|^2 \mathcal{W}(u,v)\, dudv \qquad (18)$$
$$= \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{|\sigma_j(u,v)|^2}{|u|^2 |v|^2} dudv$$

$$MDD_X^2(j) = \int_{\mathbb{R}} |\sigma_j(\mathbf{v})|^2 \mathcal{W}(\mathbf{v})\, d\mathbf{v},$$
$$= \frac{1}{\pi} \int_{\mathbb{R}} \frac{|\sigma_j(v)|^2}{|v|^2} dv \qquad (19)$$

Martingale difference hypthesis

- In particular, $MDD_X(j)$ equals zero for a fixed $j \in \mathbb{N}^+$ if and only if X_t is conditionally mean independent of X_{t-j}, i.e., $E(X_t|X_{t-j}) = E(X_t)$ is satisfied. In this sense the statistic can help us determine if a given process is a martingale difference sequence.
- In general terms we say that a sequence of variables $\{X_t\}$ satisfies the martingale difference hypothesis if and only if $E(X_t|X_{t-j}) = E(X_t)$ almost surely for all $j \in \mathbb{N}^+$.

- It is relevant to determine whether a given process is a martingale difference sequence or not even though it is a white noise process. In fact, while the existence of linear dependence (non-null autocorrelation) necessarily implies the presence of dependence in the conditional mean (i.e. the martingale difference hypothesis is rejected), the converse is not true, as we could observe processes that are uncorrelated despite exhibiting some form of nonlinear dynamics in the conditional mean.
- AMDD can be considered a useful graphical tool when determining whether a time series is a martingale difference sequence or not

Consider the following GARCH(1,1) model

$$\begin{cases} X_t = \varepsilon_t \sigma_t, \\ \sigma_t^2 = 0.001 + 0.09 X_{t-1}^2 + 0.90 \sigma_{t-1}^2, \end{cases} \tag{20}$$

where $\{\varepsilon_t\}$ sequence of i.i.d. standard normal random variables. This is an uncorrelated martingale difference sequence (observations are conditionally mean independent upon their past values).

We can also introduce the bilinear model

$$X_t = \varepsilon_t + 0.15\varepsilon_{t-1}X_{t-1} + 0.05\varepsilon_{t-1}X_{t-2}, \tag{21}$$

where $\{\varepsilon_t\}$ is again a sequence of i.i.d. standard normal random variables. What we have in this case is a time series that is uncorrelated but is not a martingale difference sequence (we have the unraveling of some degree of serial dependence in the conditional mean).

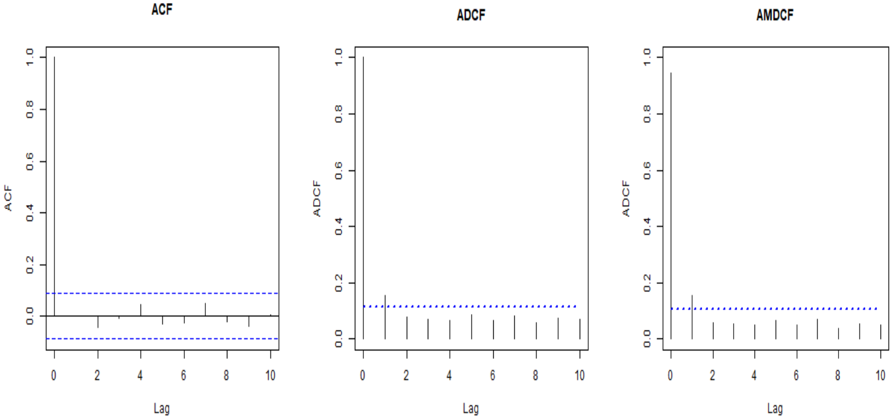

Figure: The ACF, AMDCF and ADCF plots for the BILINEAR model.

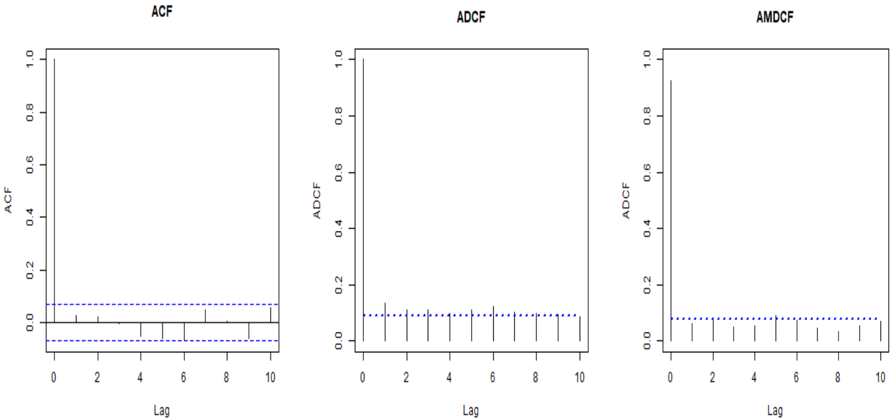

Figure: The ACF, AMDCF and ADCF plots for the GARCH model.

Generalized spectral density

The construction of the test is based on the concept of generalized spectral density (Hong 1999): given the generalized auto-covariance defined already as $\sigma_j\left(u,v\right) = Cov\left(e^{iu'X_t}, e^{iv'X_{t-j}}\right) = \phi_j\left(u,v\right) - \phi\left(u\right)\phi\left(v\right)$, under the assumptions that grant the existence of the Fourier transform of $\sigma_j\left(u,v\right)$ (making the condition $\sup_{u,v\in(-\infty,\infty)}\sum_{j=-\infty}^{\infty}\left|\sigma_j\left(u,v\right)\right|$ satisfied), the generalized spectral density function takes the form

$$f\left(\omega,u,v\right) = \frac{1}{2\pi}\sum_{-\infty}^{\infty}\sigma_j\left(u,v\right)e^{-ij\omega}, \quad \omega\in\left[-\pi,\pi\right] \tag{22}$$

The generalized spectral density $f\left(u,v,\omega\right)$ can capture all pairwise dependencies in contrast to standardized spectral density which on the other hand, can determine only linear dependencies.

- Assumption (1) : $\{X_t\}$ is a strictly stationary α-mixing process with mixing coefficients $\alpha(j)$, $j \geqslant 1$.
- Assumption (2) : $E|X_t| < \infty$.
- Assumption (3) : The mixing coefficients of $\{X_t\}$, $\alpha(j)$, satisfy
  - $\sum_{j=-\infty}^{\infty} \alpha(j) < \infty$;
  - $\alpha(j) = O\left(1/j^2\right)$.
- Assumption (4) : The kernel function $k(.)$ is defined such that $k : \mathbb{R} \longrightarrow [-1, 1]$, is symmetric and continuous at zero and at all but a finite number of points. Moreover it holds that $k(0) = 1$, $\int_{-\infty}^{\infty} k^2(z)\, dz < \infty$ and $|k(z)| \leq C|z|^{-b}$ for large z and $b > 1/2$.

Hong shows that $f\left(u,v,\omega\right)$ can be consistently estimated by

$$\widehat{f}_n\left(\omega,u,v\right)=\frac{1}{2\pi}\sum_{j=-(n-1)}^{n-1}\left(1-\frac{|j|}{n}\right)^{1/2}k\left(j/p\right)\widehat{\sigma}_j\left(u,v\right)e^{-ij\omega} \tag{23}$$

where $\widehat{\sigma}_j\left(u,v\right)=\widehat{\phi}_j\left(u,v\right)-\widehat{\phi}_j\left(u\right)\widehat{\phi}_j\left(v\right)$ and $\widehat{\phi}_j\left(u,v\right)=\left(n-|j|\right)^{-1}\sum_{t=|j|+1}^{n}e^{i\left(uX_t+vX_{t-|j|}\right)}$ is the empirical joint characteristic function. Finally, the variable $p=p_n$ is a bandwidth or lag order parameter and $k\left(\cdot\right)$ is a symmetric kernel function or "lag window".

Under the assumption of serial independence, $f(\omega, u, v)$ becomes the flat spectrum

$$f_0(\omega, u, v) = \frac{1}{2\pi}\sigma_0(u, v), \quad \omega \in [-\pi, \pi],$$

which can be equivalently estimated by $\widehat{f}_0(\omega, u, v) = (2\pi)^{-1}\widehat{\sigma}_0(u, v)$.
In order to detect serial dependence we can compare the two estimators $\widehat{f}_n(u, v, \omega)$ and $\widehat{f}_0(u, v, \omega)$: in Hong (1999) this is done by means of a weighted L_2 distance

$$\begin{aligned} H_{99} &= \int\int_{-\pi}^{\pi} \left|\widehat{f}_n(\omega, u, v) - \widehat{f}_0(\omega, u, v)\right|^2 d\omega dW(u, v) \qquad (24) \\ &= \left(\frac{\pi}{2}\right)^{-1} \sum_{j=1}^{n-1} k^2(j/p)(1 - j/n) \int \left[\widehat{\sigma}_j(u, v)^2\right] dW(u, v) \\ &= \left(\frac{\pi}{2}\right)^{-1} \sum_{j=1}^{n-1} k^2(j/p)(1 - j/n) ||\sigma_j(u, v)||_W^2 \end{aligned}$$

where the second equality follows from Parseval's identity and $W(u, v)$ is a specified symmetric weighting function.

The test for the hypothesis of serial independence introduced by Fokianos and Pitsillou (2017) is based on this principle: assuming that the weighting function $W(u,v)$ is equal to the weight function introduced in Szekely et al. (2007), that is, in the case of univariate series, $W(u,v) = \mathcal{W}(u,v) = \left(\pi^2 |u|^2 |v|^2\right)^{-1}$, the original test H_{99} can be re-expressed in terms of distance covariance

$$T_n = \sum_{j=1}^{n-1} (n-j)\, k^2(j/p)\, \widehat{V}_X^2(j),$$

where this simplified form of the test is computationally efficient as it avoids the computation of any integral.

While the generalized spectrum in its base form is used to capture any form of serial dependence, its partial derivatives at zero can be used to test for an array of more specific hypotheses as for example serial uncorrelatedness, MDH, conditional homoscedasticity, conditional symmetry and many others. To this purpose we can compare, in terms of their regularized L_2 distance, the derivative estimators

$$\widehat{f}_n^{(0,m,l)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-(n-1)}^{n-1} (1 - |j|/n)^{1/2} k(j/p) \widehat{\sigma}_j^{(m,l)}(u, v) e^{-ij\omega}$$

$$\widehat{f}_n^{(0,m,l)}(\omega, u, v) = \frac{1}{2\pi} \widehat{\sigma}_0^{(m,l)}(u, v) \tag{25}$$

where $\widehat{\sigma}_j^{(m,l)}(u, v) = \partial^{m+l} \widehat{\sigma}_j(u, v) / \partial^m u \partial^l v$. Specific aspects of dependence can be considered depending on the order (m, l) : this peculiar property derives from the various specifications that the term of generalized covariance assumes according to the order of differentiation.

In particular, when the order of differentiation is taken to be $(m, l) = (1, 0)$, the spectrum can be used to construct a test for the martingale difference hypothesis $E\left[(X_t - \mu)\,|X_{t-j}, j > 0\right] = 0$ *a.s.*, where $\mu = E\left(X_t\right)$.

We can write therefore

$$\int \int_{-\pi}^{\pi} \left| \widehat{f}_n^{(1,0)}(\omega, 0, v) - \widehat{f}_0^{(1,0)}(\omega, 0, v) \right|^2 d\omega dW(v) \tag{26}$$

$$= \left(\frac{\pi}{2}\right)^{-1} \sum_{j=1}^{n-1} k^2(j/p)(1 - j/n) \int \left| \widehat{\sigma}_j^{(1,0)}(0, v) \right|^2 dW(v)$$

$$= \left(\frac{\pi}{2}\right)^{-1} \sum_{j=1}^{n-1} k^2(j/p)(1 - j/n) \left|\left| \widehat{\sigma}_j^{(1,0)}(v) \right|\right|_W^2,$$

where the second equality follows from Parseval's identity; the estimator $\widehat{\sigma}_j^{(1,0)}(0, v) = (n - j)^{-1} \sum_{t=j+1}^{n} X_t \left(e^{ivX_{t-j}} - \widehat{\phi}_j(0, v)\right)$ is consistent for $\sigma_j^{(1,0)}(0, v) = Cov\left(X_t, e^{ivX_{t-|j|}}\right) = E\left(X_t e^{ivX_{t-|j|}}\right) - E\left(X_t\right)\phi(v)$.

We can now discuss the connection of MDD with the original martingale test introduced by Hong and how we can use MDD to simplify Hong's framework.Assuming that $W(v)$ is the nonintegrable weight function $\mathcal{W}(v)$ introduced MDD allows us to evaluate the norm term $||\sigma_j(v)||_{\mathcal{W}}$: accordingly, we formulate our test statistic as

$$M_n = \sum_{j=1}^{n-1} (n-j)\, k^2 (j/p)\, \widehat{MDD}_X^2 (j)\,. \tag{27}$$

- Assumption (1) : $\{X_t\}$ is a strictly stationary α-mixing process with mixing coefficients $\alpha(j)$, $j \geqslant 1$.
- Assumption (2) : $E\left|X_t^2\right| < \infty$.
- Assumption (3) : The mixing coefficients of $\{X_t\}$, $\alpha(j)$, satisfy
    - $\sum_{j=-\infty}^{\infty} \alpha(j) < \infty$;
    - $\alpha(j) = O\left(1/j^2\right)$.
- Assumption (4) : The kernel function $k(.)$ is defined such that $k : \mathbb{R} \longrightarrow [-1, 1]$, is symmetric and continuous at zero and at all but a finite number of points. Moreover it holds that $k(0) = 1$, $\int_{-\infty}^{\infty} k^2(z)\, dz < \infty$ and $|k(z)| \leq C|z|^{-b}$ for large z and $b > 1/2$.

Suppose Assumption (1) and (2) hold true. Then for all $j = 0, \pm 1, \pm 2, ...$we have

$$\widehat{MDD}^2_X(j) \xrightarrow{A.S.} MDD^2_X(j)$$

almost surely, as $n \to \infty$.

Based on Hong (1999) we can define the standardized version of the statistic

$$M_{99} = \left[\sum_{j=1}^{n-1} (n-j)\, k^2 (j/p)\, \widehat{MDD}_X^2 (j) - \widehat{C}_0 \right] \cdot \left(\widehat{D}_0 \right)^{-1/2} \tag{28}$$

In particular, the terms used for the standardization, $\widehat{C}_0$ and $\widehat{D}_0$ are defined as follows

$$\widehat{C}_0 = \widehat{R}_1 (0) \sum_{j=1}^{n-1} k^2 (j/p) \int \widehat{\sigma}_0 (v, -v)\, dW(v) \tag{29}$$

$$\widehat{D}_0 = 2 \widehat{R}_1^2 (0) \sum_{j=1}^{n-2} k^4 (j/p) \int \left| \widehat{\sigma}_0 (v, v') \right|^2 dW(v)\, dW(v'),$$

where the integrals can be again explicitated through distance covariance.

Theorem

Suppose Assumptions 2 and 4 hold, and $p = cn^{\lambda}$, with $c > 0$, $\lambda \in (0,1)$. Then, assuming X_t is independently and identically distributed

$$\frac{M_n - \widehat{R}_1(0) \sum_{j=1}^{n-1} k^2(j/p) \int \widehat{\sigma}_0(v,-v)\, d\mathcal{W}(v)}{\left\{ 2\widehat{R}_1^2(0) \sum_{j=1}^{n-2} k^4(j/p) \int |\widehat{\sigma}_0(v,v')|^2\, d\mathcal{W}(v)\, d\mathcal{W}(v') \right\}^{1/2}} \xrightarrow{d} N(0,1)$$

as $n \longrightarrow \infty$, in distribution.

Theorem

Suppose that Assumptions 1, 3(i), and 4 hold and $p = cn^{\lambda}$ *for* $c > 0$ *and* $\lambda \in (0,1)$. *Then ,*

$$\frac{\sqrt{p}}{n} M_{99} \xrightarrow{P} \frac{\frac{\pi}{2} \int \int \left| f^{(1,0)}(\omega, 0, v) - f_0^{(1,0)}(\omega, 0, v) \right| d\omega d\mathcal{W}(v)}{\left[D_0 \int_0^{\infty} k^4(z)\, dz \right]^{1/2}}$$

as $n \longrightarrow \infty$, *in probability.*

- We implement the wild bootstrap procedure described in Fokianos and Pitsillou (2017) to approximate the limit distribution of degenerate U-and V-statistics and relies on generating sequences of auxiliary random variables $\left\{W^{*}_{t,n}\right\}_{t=1}^{n-|j|}$ to obtain the bootstrap value of the test statistic.

- Remember now that sample martingale difference divergence takes the form $\widehat{MDD}^2_X(j) = (n-|j|)^{-2} \sum_{r,l=1+|j|}^{n} A_{rl} B_{rl}$ where, based on the sample $\left\{X_t, X_{t-|j|} : t = 1+|j|, ..., n\right\}$, we calculate the double centered Euclidean distance matrices $A = (A_{rl})$ and $B = (B_{rl})$. If we define the $((n-|j|) \times 1)$ unitary vector as $e_{n-|j|}$ we can equivalently express the statistic as

$$\widehat{MDD}^2_X(j) = (n-|j|)^{-2} \left(e_{n-|j|}\right)^T * A \circ B * \left(e_{n-|j|}\right) \tag{30}$$

where $\circ$ stands for the element-wise product operator for two matrices.

Size Properties

Under the null hypothesis that the univariate process $\{X_t\}$ is a MDS, we consider the following models:

1) GARCH(1,1) process:

$$\begin{cases} X_t = \varepsilon_t \sigma_t \\ \sigma_t^2 = 0.001 + 0.09 X_{t-1}^2 + 0.9 \sigma_{t-1}^2 \end{cases} \tag{31}$$

2) ARCH(2) process:

$$\begin{cases} X_t = \varepsilon_t \sigma_t \\ \sigma_t^2 = 0.5 + 0.8 X_{t-1}^2 + 0.1 X_{t-2}^2 \end{cases} \tag{32}$$

with $\{\varepsilon_t\}$, $\{u_t\}$ two distinct sequences of iid standard normal random variables.

We can estimate the distribution of the test statistic M_n by that of

$$M_n^* = \sum_{j=1}^{n-1} (n-j)\, k^2 (j/p) \frac{1}{(n-|j|)^2} \left(W_{n-|j|}^*\right)^T * A \circ B * \left(W_{n-|j|}^*\right) \quad (33)$$

where $W_{n-|j|}^*$ is the $((n-|j|) \times 1)$ auxiliary variables' vector. For a number of bootstrap replicates equal to B the botstrap p-values are then obtained as the ratio $\sum_{b=1}^{B} \mathbb{I}\left(T_n^* \geqslant T_n\right)/B$, where $\mathbb{I}(\cdot)$ is the indicator function.

n		500						1000					
p		n^(1/5)		n^(2/5)		n^(3/5)		n^(1/5)		n^(2/5)		n^(3/5)	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T	BAR	4.4	8.5	5.7	11.5	6.3	10.6	5.5	10.1	4.2	8.6	5.7	10.8
	PAR	4.5	9.3	5.9	11.3	6.1	10.8	4.9	9.2	4.5	9.8	//	//

Table: Empirical size, GARCH(1,1), Independent Wild Boootstrap ($l_n = 0$)

The table shows the empirical rejection probabilities (for 5% and 10% nominal levels) obtained for Parzen and Bartlett kernel functions and different values of bandwidth parameter, $p = n^{1/5}, n^{2/5}$ and $n^{3/5}$ respectively, where n indicates the sample dimension chosen for the simulated process. The p-values are obtained on 499 bootstrap replications for each of the 1000 simulations.

n		500						1000					
p		n^(1/5)		n^(2/5)		n^(3/5)		n^(1/5)		n^(2/5)		n^(3/5)	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
T	BAR	6.2	9.8	6.0	9.8	5.0	9.2	3.0	7.6	4.6	9.8	6.2	9.8
	PAR	5.0	8.4	5.8	11.2	8.0	13.6	5.7	10.4	4.8	9.8	5.8	10.3

Table: Empirical size, ARCH(2), Independent Wild Boootstrap ($l_n = 0$)

In general, we can say that, in the case of independent wild bootstrap, the empirical size properties of the test are appropriate.

Conversely, under the alternative, we study the following non-martingale difference sequences

4) Bilinear Process:

$$X_t = \varepsilon_t + 0.15\varepsilon_{t-1}X_{t-1} + 0.05\varepsilon_{t-1}X_{t-2}, \tag{34}$$

5) Threshold autoregressive model of order on (TAR(1)):

$$X_t = \begin{cases} -1.5X_{t-1} + \varepsilon_t, & X_{t-1} < 0 \\ 0.5X_{t-1} + \varepsilon_t, & X_{t-1} \geq 0 \end{cases} \tag{35}$$

with $\{\varepsilon_t\}$ sequence of iid standard normal random variables for both models.

		M_n, Wild Ind.Boot. $(l_n = 0)$		
p		$p = n^{1/5}$	$p = n^{2/5}$	$p = n^{3/5}$
n	100	26	19	15
	300	44	39	21
	600	79	66	33
	800	93	80	50
	1000	100	100	55

Table: Empirical power levels, Bilinear Model, Parzen Kernel

We report the empirical power level for the sample sizes $n = (100, 300, 600, 800, 1000)$. The results are obtained employing a parzen kernel for different bandwidth parameter values $p = n^{1/5}, n^{2/5}$ and $n^{3/5}$. The p-values are obtained on 499 bootstrap replications for each of the 1000 simulations.

		M_n, Wild Ind.Boot. ($l_n = 0$)		
p		$p = n^{1/5}$	$p = n^{2/5}$	$p = n^{3/5}$
n	100	89	68	15
	300	100	80	21
	600	100	84	33
	800	100	90	50
	1000	100	100	55

Table: Empirical power levels, TAR Model, Parzen Kernel

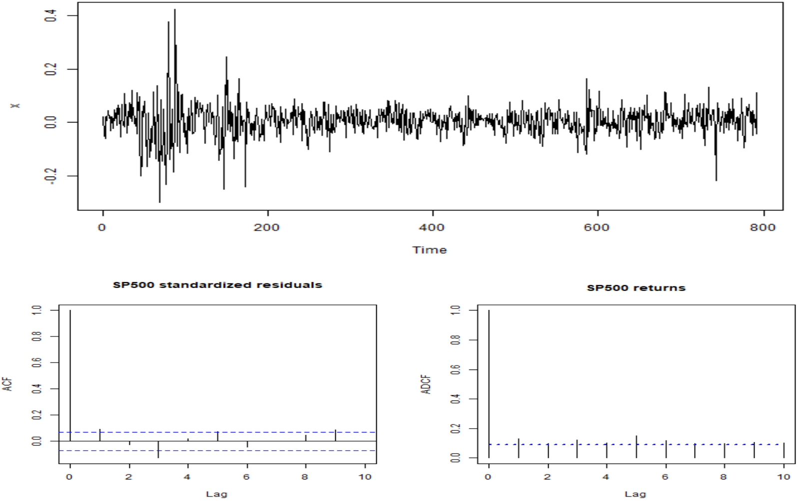

Figure: S&P 500 excess returns series (upper panel), ACF plot and ADCF plot .

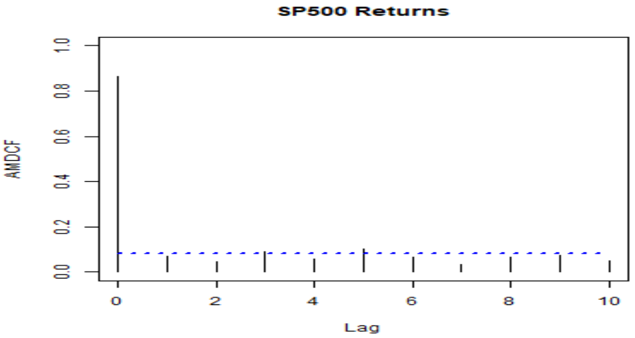

Figure: S&P 500 excess returns series (upper panel), ACF plot and AMDCF plot .

p	T_n	M_n
4	0.014	0.86
15	0.000	0.284
55	0.008	0.286

Table: P-values for statistics T_n and M_n for S&P 500 return series.

- Doukhan, P. and Louhichi, S. (1999), “A New Weak Dependence Condition and Applications to Moment Inequalities”, *Stochastic Processes and Their Applications*, 84, 313–342.
- Durlauf, S. 1991, ”Spectral-based test for the martingale hypothesis”, *Journal of Econometrics,* 50, 1-19.
- Deo, R. S. 2000, ”Spectral tests of the Martingale Hypothesis under conditional heteroskedasticity”, *Journal of Econometrics,* 99, 291-315.
- Escanciano, C. J. and Velasco, C. (2006), ”Generalized spectral tests for the martingale difference hypothesis”, *Journal of Econometrics,* 134, 151-185.
- Fokianos, K. and Pitsillou, M. (2017), ”Consistent testing for pairwise dependence in time series”, *Technometrics*, 59, 262–270.
- Fokianos, K. and Pitsillou, M. (2018), ”Testing independence for multivariate time series via the auto-distance correlation matrix”, *Biometrika*, 105, 337–352.

Lee, C.E. and Shao, X. (2018). "Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series", *Journal of the American Statistical Association*, 113, 216-229.

Leucht, A., and Neumann, M. H. (2013), "Dependent Wild Bootstrap for Degenerate U- and V-statistics", *Journal of Multivariate Analysis*, 117, 257–280.

Shao, X. and Zhang, J. (2014), "Martingale difference correlation and its use in high-dimensional variable screening", *Journal of the American Statistical Association*, 109, 1302–1318.

Székely, G. J., Rizzo,M. L., and Bakirov, N. K. (2007), "Measuring and Testing Dependence by Correlation of Distances", *The Annals of Statistics*, 35, 2769-2794.

Tsay, R. S. (2005), "Analysis of financial time series" (2nd ed.), Hoboken:NJ:Wiley Series in Probability and Statistics. [266,267]

Wang, G., Shao, X. (2021) "Testing for the martingale difference hypothesis in multivariate time series models", *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2021.1889568

Zhou, Z. (2012), "Measuring Nonlinear Dependence in Time-Series, a Distance Correlation Approach", *Journal of Time Series Analysis*, 33, 438–457.