The Bootstrap for Testing the Equality of Two Multivariate Time Series with an Application to Financial Markets

Ángel López-Oriona and **José A. Vilar**

Research group MODES, University of A Coruña, Spain
Center of Information and Communication Technologies (CITIC)

June 27, 2022

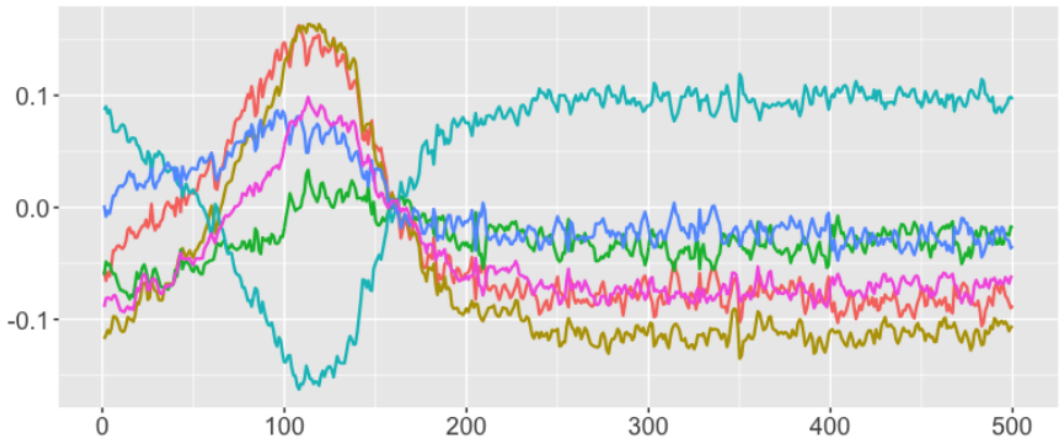
gmodes DPTO. DE MATEMÁTICAS

UNIVERSIDADE DA CORUÑA

citic

Multivariate Time Series

- A multivariate time series (MTS) is a realization of the d-variate stochastic process $\{\boldsymbol{X}_t,\, t \in \mathbb{Z}\} = \{(X_{t,1}, \ldots, X_{t,d}),\, t \in \mathbb{Z}\}$.
- We represent a MTS by means of a matrix where each row stands for a given $t \in \mathbb{Z}$ and each column stands for a variable or dimension.
- Example: 6-lead ECG of a healthy patient.



- **Measuring distance between MTS**. We want to assess discrepancy between MTS in such a way that two MTS are similar if their underlying stochastic processes are similar in terms of dependence structures.

Quantile cross-spectral density. Notion and estimate

1. **Quantile cross-covariance function of lag l.** Given a lag $l \in \mathbb{Z}$, quantile levels $(\tau, \tau') \in [0,1]^2$, and dimensions $1 \leq j_1, j_2 \leq d$, we define:

$$\gamma_{j_1,j_2}(l,\tau,\tau') = \text{Cov}\left(I\left\{X_{t,j_1} \leq q_{j_1}(\tau)\right\}, I\left\{X_{t+l,j_2} \leq q_{j_2}(\tau')\right\}\right),$$

where $q_j(\tau) = F_j^{-1}(\tau)$, with F_j being the marginal cdf of $X_{t,j}$.

2. **Quantile cross-spectral density (QCD).** Under suitable mixing conditions, the Fourier transform of the cross-covariances $\gamma_{j_1,j_2}(l,\tau,\tau')$ is well-defined:

$$\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau') = (1/2\pi)\sum_{l=-\infty}^{\infty}\gamma_{j_1,j_2}(l,\tau,\tau')e^{-il\omega}, \quad \omega \in \mathbb{R}.$$

3. **Smoothed rank-based copula cross (CCR) periodogram.** For a T-length realization of the process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$, Barunik and Kley (2019) propose to estimate $\mathfrak{f}_{j_1,j_2}(\omega,\tau,\tau')$ by smoothing the cross-periodograms based on the indicator functions $I\{\hat{F}_{T,j}(X_{t,j})\}$, where $\hat{F}_{T,j}$ denotes the ecdf of $X_{t,j}$. This estimator, denoted by $\hat{G}_{T,R}^{j_1,j_2}(\omega,\tau,\tau')$, is evaluated for every couple of components over the Fourier frequencies (Ω) and pairs of r quantile levels of a pre-fixed set $\mathcal{T}$, thus obtaining a whole representation of the process:

$$\hat{\mathfrak{f}}_{\boldsymbol{X}_t}(\Omega,\mathcal{T}) = \left\{\left(\hat{G}_{T,R}^{j_1,j_2}(\omega,\tau,\tau')\right)_{1\leq j_1,j_2\leq d}, \ \omega\in\Omega, \ \tau,\tau'\in\mathcal{T}\right\}.$$

A novel dissimilarity measure between MTS

- A dissimilarity measure between two MTS of length T, $\overline{\boldsymbol{X}}_t^{(1)}$ and $\overline{\boldsymbol{X}}_t^{(2)}$, could be established by comparing their representations in terms of $\hat{\mathrm{f}}_{\boldsymbol{X}_t^{(1)}}$ and $\hat{\mathrm{f}}_{\boldsymbol{X}_t^{(2)}}$.
- We represent the MTS $\overline{\boldsymbol{X}}_t^{(i)}$ by means of a vector $\widehat{\boldsymbol{\Psi}}^{(i)}$ constructed as:
  1. Flattening the elements of the estimated array $\hat{\mathrm{f}}_{\boldsymbol{X}_t^{(i)}}$.
  2. Taking real and imaginary parts separately and concatenating them.
- Dissimilarity measure between the series $\overline{\boldsymbol{X}}_t^{(1)}$ and $\overline{\boldsymbol{X}}_t^{(2)}$ (López-Oriona & Vilar, 2021):

$$\widehat{d}_{QCD}(\overline{\boldsymbol{X}}_t^{(1)}, \overline{\boldsymbol{X}}_t^{(2)}) = ||\widehat{\boldsymbol{\Psi}}^{(1)} - \widehat{\boldsymbol{\Psi}}^{(2)}||^2.$$

- The distance $\widehat{d}_{QCD}(\overline{\boldsymbol{X}}_t^{(1)}, \overline{\boldsymbol{X}}_t^{(2)})$ is a sample quantity. A hypothesis test can be defined by considering the corresponding population quantity as follows:

$$H_0 : d_{QCD}(\boldsymbol{X}_t^{(1)}, \boldsymbol{X}_t^{(2)}) = 0 \quad \text{vs} \quad H_1 : d_{QCD}(\boldsymbol{X}_t^{(1)}, \boldsymbol{X}_t^{(2)}) > 0,$$

where $\boldsymbol{X}_t^{(1)}$ and $\boldsymbol{X}_t^{(2)}$ are the generating stochastic processes of $\overline{\boldsymbol{X}}_t^{(1)}$ and $\overline{\boldsymbol{X}}_t^{(2)}$, respectively. $\widehat{d}_{QCD}(\overline{\boldsymbol{X}}_t^{(1)}, \overline{\boldsymbol{X}}_t^{(2)})$ is the test statistic.

A Frequency Domain Bootstrap Test (FDB) (I)

- Dette and Paparoditis (2009) consider the general class of hypotheses

$$H_0 : \int_{-\pi}^{\pi} ||\varphi(g(\omega), \omega)||^2 d\omega = 0 \text{ vs } H_1 : \int_{-\pi}^{\pi} ||\varphi(g(\omega), \omega)||^2 d\omega > 0, \quad (1)$$

  where $g(\omega)$ denotes the spectral density matrix of a multivariate process and φ is a suitable vector-valued function.

- A test statistic for the previous test can be obtained by considering $\int_{-\pi}^{\pi} ||\varphi(\hat{g}(\omega), \omega)||^2 d\omega$, where $\hat{g}(\omega)$ is a proper estimate of $g(\omega)$.

- It can be shown that, by choosing the multivariate process and the function φ in a specific way, (1) can be written as

$$H_0 : \int_{-\pi}^{\pi} d^2_{QCD,\omega}(\boldsymbol{X}_t^{(1)}, \boldsymbol{X}_t^{(2)})\, d\omega = 0 \text{ vs } H_1 : \int_{-\pi}^{\pi} d^2_{QCD,\omega}(\boldsymbol{X}_t^{(1)}, \boldsymbol{X}_t^{(2)})\, d\omega > 0,$$

  where the subscript ω indicates that the distance is computed by choosing only the frequency ω.

A Frequency Domain Bootstrap Test (FDB) (II)

- **Step 1**. Given the set of Fourier frequencies $\{\omega_k = 2\pi k/T,\ 1 \leq k \leq (T-1)\}$, generate two sets of $(T-1)$ independent matrices

$$I^{(i)*}(\omega_k) = \widehat{G}(\omega_k)^{1/2} W_k^* \widehat{G}(\omega_k)^{1/2}, \quad i = 1, 2,$$

where $\widehat{G}(\omega_k)$ is given by

$$\widehat{G}(\omega_k) = \frac{1}{2}\left(\widehat{G}^{(1)}(\omega_k) + \widehat{G}^{(2)}(\omega_k)\right),$$

with $\hat{G}^{(i)}$ being the matrix of smoothed CCR-periodograms based on the MTS of $\overline{\boldsymbol{X}_t}^{(i)}$ and

$$W_k^* \sim \begin{cases} W^C(1, I_{dr}) & 1 \leq k \leq T/2, \\ W^R(1, I_{dr}) & k \in \{T/2\}, \\ W_{T-k}^* & T/2 \leq k \leq (T-1), \end{cases}$$

where W^C and W^R denote the complex and real Wishart distributions and I_{dr} stands for the identity matrix of order dr.

A Frequency Domain Bootstrap Test (FDB) (III)

- **Step 2**. Construct the bootstrap quantile cross-spectral density matrices estimates, $\widehat{G}^{(i)*}(\omega_k)$, $i = 1, 2$, $k = 1, \ldots, T$, by smoothing $I^{(i)*}(\omega_k)$ in the form

$$\widehat{G}^{(i)*}(\omega_k) = \frac{2\pi}{T} \sum_{s=1}^{T-1} W_T(\omega_k - \omega_s) I^{(i)*}(\omega_s),$$

  where W_T is a combination of weight functions depending on a smoothing parameter $h_T \to \infty$ when $T \to \infty$.

- **Step 3**. Based on $\widehat{G}^{(1)*}$ and $\widehat{G}^{(2)*}$, construct the bootstrap version $\widehat{d}^{*}_{QCD}$ of $\widehat{d}_{QCD}$.

- **Step 4**. Repeat Steps 1-3 a large number B of times to obtain the bootstrap replicates $\widehat{d}^{(1)*}_{QCD}, \ldots, \widehat{d}^{(B)*}_{QCD}$.

- **Step 5**. The decision rule consists of rejecting H_0 if $\widehat{d}_{QCD}(\overline{\boldsymbol{X}}_t^{(1)}, \overline{\boldsymbol{X}}_t^{(2)}) > q^{*}_{1-\alpha}$, with $q^{*}_{1-\alpha}$ being the $1-\alpha$ bootstrap empirical quantile for a significance level $\alpha \in (0, 1)$.

A Test based on the Moving Blocks Bootstrap (MBB) (Kunsch, 1989) and (Liu, 1992)

- **Step 1**. Fix a positive integer, b, representing the block size, and take $k = \lceil T/b \rceil$.
- **Step 2**. For $i \in \{1, 2\}$, define the block $\boldsymbol{B}_j^{(i)} = \{\boldsymbol{X}_j^{(i)}, \ldots, \boldsymbol{X}_{j+b-1}^{(i)}\}$, for $j = 1, \ldots, q$, with $q = T - b + 1$. Let $\overline{\boldsymbol{B}} = \{\boldsymbol{B}_j^{(1)}, \ldots, \boldsymbol{B}_q^{(1)}, \boldsymbol{B}_j^{(2)}, \ldots, \boldsymbol{B}_q^{(2)}\}$ be the set of all blocks, those coming from $\overline{\boldsymbol{X}}_t^{(1)}$ and those coming from $\overline{\boldsymbol{X}}_t^{(2)}$.
- **Step 3**. Draw two sets of k blocks, $\boldsymbol{B}^{(i)*} = \{\boldsymbol{B}_1^{(i)}, \ldots, \boldsymbol{B}_k^{(i)}\}$, $i = 1, 2$, with equiprobable distribution from the set $\overline{\boldsymbol{B}}$.
- **Step 4**. Construct the pseudo-time series $\overline{\boldsymbol{X}}_t^{(i)*}$ by considering the first T temporal components of $\boldsymbol{B}^{(i)*}$, $i = 1, 2$. Compute the bootstrap version $\widehat{d}_{QCD}^*$ of $\widehat{d}_{QCD}$ based on $\overline{\boldsymbol{X}}_t^{(1)*}$ and $\overline{\boldsymbol{X}}_t^{(2)*}$.
- **Step 5**. Repeat Steps 3 and 4 a large number B of times to obtain the bootstrap replicates $\widehat{d}_{QCD}^{(1)*}, \ldots, \widehat{d}_{QCD}^{(B)*}$.
- **Step 6**. The decision rule consists of rejecting H_0 if $\widehat{d}_{QCD}(\overline{\boldsymbol{X}}_t^{(1)}, \overline{\boldsymbol{X}}_t^{(2)}) > q_{1-\alpha}^*$.

A Test based on the Stationary Bootstrap (SB) (Politis & Romano, 1994)

- **Step 1**. Fix a positive real number $p \in [0, 1]$.
- **Step 2**. Consider the set $\widetilde{\boldsymbol{X}_t} = \{\overline{\boldsymbol{X}_t}^{(1)}, \overline{\boldsymbol{X}_t}^{(2)}\}$. Draw randomly two temporal observations from $\widetilde{\boldsymbol{X}_t}$. Each one of these observations are of the form $\boldsymbol{X}_{j^i}^{(k^i)}$ for some $k^i = 1, 2$ and $j^i = 1, \ldots, T$, $i = 1, 2$.
- **Step 3**. For $i = 1, 2$, given the last observation $\boldsymbol{X}_{j^i}^{(k^i)}$, define the next bootstrap replication in $\overline{\boldsymbol{X}_t}^{(i)*}$ as $\boldsymbol{X}_{j^i+1}^{(k^i)}$ with probability $1 - p$, and drawn from the set $\widetilde{\boldsymbol{X}_t}$ with probability p. When $j^i = T$, the selected observation is $\boldsymbol{X}_1^{(2)}$ if $k^i = 1$ and $\boldsymbol{X}_1^{(1)}$ if $k^i = 2$.
- **Step 4**. Repeat Step 3 until the pseudo-series $\overline{\boldsymbol{X}_t}^{(1)*}$ and $\overline{\boldsymbol{X}_t}^{(2)*}$ contain T observations. Compute the bootstrap version $\widehat{d}^*_{QCD}$ of $\widehat{d}_{QCD}$ based on $\overline{\boldsymbol{X}_t}^{(1)*}$ and $\overline{\boldsymbol{X}_t}^{(2)*}$.
- **Step 5**. Repeat Steps 3 and 4 a large number B of times to obtain the bootstrap replicates $\widehat{d}^{(1)*}_{QCD}, \ldots, \widehat{d}^{(B)*}_{QCD}$.
- **Step 6**. The decision rule consists of rejecting H_0 if $\widehat{d}_{QCD}(\overline{\boldsymbol{X}_t}^{(1)}, \overline{\boldsymbol{X}_t}^{(2)}) > q^*_{1-\alpha}$.

Simulation study (I)

- Three simulated scenarios were considered, with the parameter δ indicating deviation from the null hypothesis.

- **Scenario 1**. **Vector autoregressive (VAR) models**.

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0.1+\delta & 0.1+\delta \\ 0.1+\delta & 0.1+\delta \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix},$$

  with $\delta \in \{0, 0.1, 0.2, 0.3\}$.

- **Scenario 2**. **Threshold autoregressive (TAR) models**.

$$\begin{pmatrix} X_{t,\mathbf{1}} \\ X_{t,\mathbf{2}} \end{pmatrix} = \begin{pmatrix} (0.9-\delta) X_{t-\mathbf{1},\mathbf{2}} I_{\{|X_{t-\mathbf{1},\mathbf{1}}| \leq \mathbf{1}\}} + (\delta - 0.3) X_{t-\mathbf{1},\mathbf{1}} I_{\{|X_{t-\mathbf{1},\mathbf{1}}| > \mathbf{1}\}} \\ (0.9-\delta) X_{t-\mathbf{1},\mathbf{1}} I_{\{|X_{t-\mathbf{1},\mathbf{2}}| \leq \mathbf{1}\}} + (\delta - 0.3) X_{t-\mathbf{1},\mathbf{2}} I_{\{|X_{t-\mathbf{1},\mathbf{2}}| > \mathbf{1}\}} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,\mathbf{1}} \\ \epsilon_{t,\mathbf{2}} \end{pmatrix},$$

  with $\delta \in \{0, 0.2, 0.4, 0.6\}$.

Simulation study (II)

- **Scenario 3**. Dynamic conditional correlation (DCC) models. Let $(X_{t,1}, X_{t,2})^\intercal = (\sigma_{t,1}\epsilon_{t,1}, \sigma_{t,2}\epsilon_{t,2})^\intercal$. The data-generating process consists of two Gaussian **GARCH** models (Engle, 2002):

$$\sigma^2_{t,1} = 0.01 + 0.05X^2_{t-1,1} + 0.94\sigma^2_{t-1,1},$$

$$\sigma^2_{t,2} = 0.5 + 0.2X^2_{t-1,2} + 0.5\sigma^2_{t-1,2},$$

$$\begin{pmatrix}\epsilon_{t,1}\\ \epsilon_{t,2}\end{pmatrix} \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho_t\\ \rho_t & 1\end{pmatrix}\right].$$

  where $\rho_t = 0.9 - \delta$, with $\delta \in \{0, 0.4, 0.8, 1.2\}$.

- Two MTS $\overline{\boldsymbol{X}}_t^{(1)}$ and $\overline{\boldsymbol{X}}_t^{(2)}$ are generated to apply the proposed tests. $\overline{\boldsymbol{X}}_t^{(1)}$ is associated with $\delta = 0$ while $\overline{\boldsymbol{X}}_t^{(2)}$ is associated with all the possible values for δ (one at a time).

- A number of $B = 200$ bootstrap replicates were used. The simulation procedure was repeated 200 times. Different values for the series length, T, were considered. Average rejection rates were calculated at a significance level $\alpha = 0.05$.

Results (I)

- **Null hypothesis** ($\alpha = 0.05$).

	Scenario	1	2	3
$T = 500$	BFD	0.025	0.020	0.035
	MBB	0.080	0.070	0.080
	SB	0.055	0.055	0.050
$T = 1000$	BFD	0.060	0.030	0.060
	MBB	0.070	0.095	0.130
	SB	0.040	0.060	0.060

- **Alternative hypothesis. Scenario 1** ($\alpha = 0.05$).

	Method	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$
$T = 100$	BFD	0.060	0.395	0.905
	MBB	0.160	0.575	0.980
	SB	0.100	0.465	0.960
$T = 200$	BFD	0.060	0.610	0.990
	MBB	0.185	0.790	0.995
	SB	0.095	0.695	0.990
$T = 300$	BFD	0.115	0.730	1
	MBB	0.225	0.835	1
	SB	0.130	0.770	1

Results (II)

- **Alternative hypothesis. Scenario 2** ($\alpha = 0.05$).

	Method	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.6$
$T = 100$	BFD	0.340	0.695	0.950
	MBB	0.540	0.775	0.990
	SB	0.325	0.690	0.910
$T = 200$	BFD	0.595	0.875	0.985
	MBB	0.780	0.925	1
	SB	0.625	0.885	0.985
$T = 300$	BFD	0.765	0.965	1
	MBB	0.885	0.990	1
	SB	0.805	0.955	1

- **Alternative hypothesis. Scenario 3** ($\alpha = 0.05$).

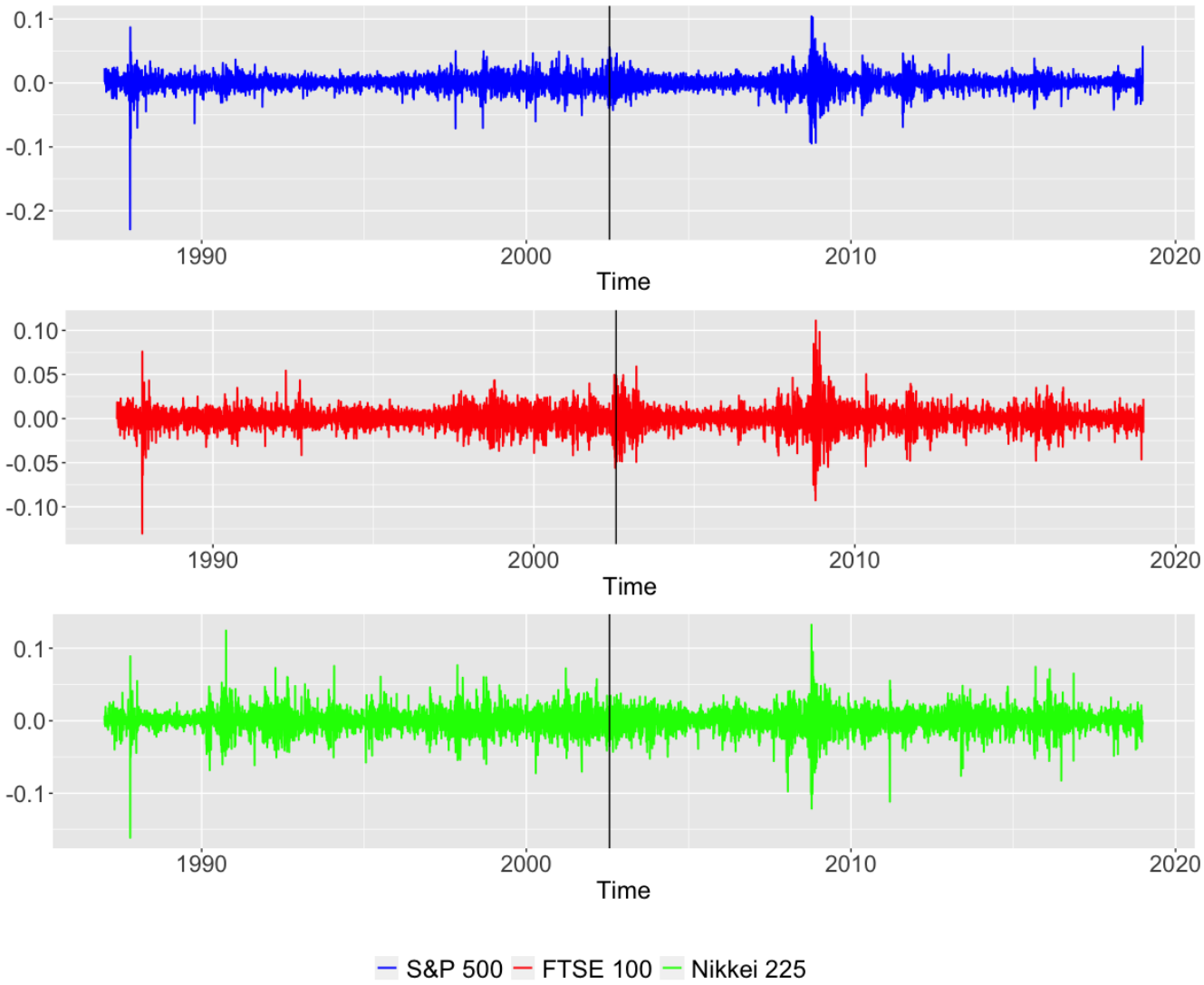
	Method	$\delta = 0.4$	$\delta = 0.8$	$\delta = 1.2$
$T = 100$	BFD	0.030	0.150	0.895
	MBB	0.080	0.395	0.950
	SB	0.055	0.230	0.870
$T = 200$	BFD	0.050	0.505	0.965
	MBB	0.185	0.725	1
	SB	0.080	0.455	0.965
$T = 300$	BFD	0.135	0.730	1
	MBB	0.255	0.840	1
	SB	0.155	0.695	1

¿Did the Dotcom Bubble Crash Change the Global Market Behaviour?

- **Dotcom bubble**. Rapid rise in U.S. technology stock equity valuations exacerbated by investments in Internet-based companies during the bull market in the late 1990s.
- **Bubble crash**. Between 2001 and 2002, with equities entering a bear market. The Nasdaq index fell by 76.81 %.
- **Data**. Three well-known stock market indexes, the S&P 500 (U.S), the FTSE 100 (Europe) and the Nikkei 225 (Japan). Two different periods: from 1987 to 2002 and from 2002 to 2018.
- **Question**. Did the three stock market indexes show the same behaviour before and after the dotcom bubble crash?.

Original Series (Stock Prices)

Transformed Series (Stock Returns)



Results

- The three bootstrap tests were applied to test the equality of generating processes given both MTS. A number of $B = 500$ bootstrap replicates was considered.
- The p-values obtained by means of methods BFD, MBB and SB were all 0, indicating rejection of the null hypothesis.
- **Conclusion**. The doctom bubble crash in the early 2000s provoked a permanent change in the behaviour of the global economy.

Conclusions

- The problem of testing the equality of two multivariate time series was addressed.
- Three bootstrap tests were proposed: (i) a frequency domain bootstrap test (FDB), (ii) a test based on the moving blocks bootstrap (MBB) and (iii) a test based on the stationary bootstrap (SB).
- The procedures were analysed in a broad simulation study in terms of both size and power. Overall, FDB and SB showed the best results.
- The tests were applied to check whether or not the dotcom bubble crash of 2000s provoked a permanent change in the global market economy. The answer was yes.

References

Dette, H., & Paparoditis, E. (2009). Bootstrapping frequency domain tests in multivariate time series with an application to comparing spectral densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(4), 831–857.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, 1217–1241.

Liu, R. Y. (1992). Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap*.

López-Oriona, Á., & Vilar, J. A. (2021). Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series. *Expert Systems with Applications*, *185*, 115677.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, *89*(428), 1303–1313.

Thank you for your attention!