



# Utilizing symmetric Phi-divergence in serial independence testing

Emad Ashtari Nezhad 

*General Administration of Economic Affairs and Finance of Razavi Khorasan, Mashhad, Iran*

## Abstract

This manuscript introduces a novel class of time series independence tests based on Phi-divergence and quantile-based symbolization. We derive the asymptotic distribution of the test statistic and propose a bootstrap version. Simulations identified optimal parameter values and compared the test performance to existing methods, demonstrating higher size-corrected power for specific Phi-divergence cases. Furthermore, we investigate Rukhin and power divergence, revealing Pearsons divergence as optimal. The proposed tests were applied to financial (Tehran Stock Exchange, S&P 500) and ecological (Lynx population) datasets, effectively detecting dependence on the data and confirming the adequacy of the model through independent residuals, demonstrating the robustness and versatility of the method in diverse domains.

**Mathematics Subject Classification (2020).** 62M10, 62G10, 62F40

**Keywords.** Serial independence test, Phi-divergence, simulation studies, quantile symbolization, time series

## 1. Introduction

Time series represent observations from a stochastic process  $\{Y_t, t \in \mathbb{Z}\}$ . Time series analysis, using models like *AR*, *MA*, and *ARIMA*, is applied to identify patterns and make forecasts in fields such as economics and engineering. When evaluating these models, we commonly encounter two critical questions: first, whether the data are independent and identically distributed (iid) prior to model fitting, and second, whether the residuals are also iid after model fitting. Thus, our objective is to test the null hypothesis  $H_0 : \{Y_t, t \in \mathbb{Z}\}$ , which posits that this sequence consists of independent and identically distributed random variables.

Divergence criteria are commonly used to test the hypothesis  $H_0$ . One significant instance of this is the non-parametric test for assessing serial independence (no dependence between observations over time) through mutual information, as established by [15]. This particular test applies the *KL* distance to evaluate the differences between the estimated joint distribution and the respective marginal distributions. Similarly, Robinson [34] proposed another test based on the integrated absolute difference, while Skaug and Tjøstheim [36] implemented the Cramer-von Mises divergence measure in their approach. Furthermore, Ghoudi et al. [18] introduced test statistics grounded in the Kolmogorov-Smirnov

distance. In particular, the BDS test serves as an example of how correlation dimension measures can be used effectively to assess independence [7].

Recently, there has been a growing interest in symbolic dynamics due to their potential to create effective independence tests. An illustrative example is the  $G(m)$  test introduced by [29], which is based on permutation entropy (a measure of complexity based on ordinal patterns). Through a comparative analysis of the developed tests, it was demonstrated that the BDS test outperformed its counterparts by focusing on numerical differences between observations rather than on ranks [1, 8, 9, 29]. Furthermore, Elsinger [16] critiqued the BDS test for its method of generating overlapping  $m$ -dimensional vectors from time series data, noting that the asymptotic distribution of the  $G(m)$  test statistic does not align with the chi-square distribution. In contrast, Ashtari Nezhad et al. [3] established that the  $G(m)$  test statistic follows a weighted chi-square distribution and explored alternative approaches, such as overlap control and bootstrap methods. Their research indicated that the modified permutation entropy-based test not only improved accuracy but also significantly increased the power of the tests. In an effort to improve test accuracy through the use of symbolization, Ashtari Nezhad et al. [4] built on previous concepts by integrating quantile symbolization (discretizing time series into quantile-based symbols) with the measure  $KL$ . Their findings indicated that this test, which is based on quantile symbolization, delivered impressive accuracy in various sample sizes, effectively preserving the level of nominal significance and outperforming other competing tests.

Recent studies have introduced innovative nonparametric methods to detect serial dependence in time series data. Jiang et al. [25] proposed a method for time series with object value in metric spaces, utilizing distance covariance to capture non-linear dependencies. Zhou and Müller [42] introduced a framework for testing the independence between random objects in general metric spaces, employing profile association measures. Combettes [10] explored symbolic representations for time series, enhancing pattern detection through advanced symbolization techniques. Weiß and Schnurr [40] presented generalized ordinal patterns for discrete-valued time series, offering robust tools for identifying serial dependence. Additionally, Liu et al. [27] developed kernel-based joint independence tests for multivariate time series, effectively capturing higher-order dependencies. Anjali et al. [2] introduced the GSSX method, a symbolic representation based on golden sections that aims to reduce the dimensionality of time series, thus improving computational efficiency. Wang et al. [38] proposed a foundation model with data generation with series symbols to address data scarcity in time series analysis, enhancing model performance on various tasks. Betken et al. [6] utilized ordinal patterns for the detection of change points, providing a nonparametric approach to the identification of structural shifts in time series data. He et al. [22] introduced the non-parametric symbol approximation (NSAR), facilitating efficient time series classification in smart manufacturing contexts. Furthermore, Hassani et al. [21] examined the implications of white noise misapplications in time series modeling, emphasizing the importance of accurate model diagnostics. Yu et al. [41] developed a semiparametric latent ANOVA model for event-related potentials, offering insights into time series analysis in neuroscience. Mohammadi et al. [30] proposed a model-free prediction approach for time series, leveraging nonparametric techniques to enhance forecasting accuracy. Hounyo and Lin [24] studied wild bootstrap inference with multi-way clustering and serially correlated time effects, contributing to robust statistical inference methods. Lastly, Wang et al. [39] presented methods to detect state correlations between heterogeneous time series, advancing the analysis of complex temporal dependencies.

**Table 1.** Comparison of serial independence tests and methods with the proposed symmetric Phi-divergence test

Test/Method	Methodological Basis	Limitations	Proposed Test Advantage
Dionisio et al. [15]	Mutual information (KL divergence)	Asymmetric measure; tail sensitivity	Symmetric divergence; time-reversal invariant
Robinson [34]	Integrated absolute difference	Limited power for complex patterns	Higher size-corrected power via quantile symbolization
Skaug & Tjøstheim [36]	Cramer-von Mises divergence	Less robust for small samples	Improved accuracy across sample sizes
Ghoudi et al. [18]	Kolmogorov-Smirnov distance	Lower power for non-linear patterns	Enhanced power for diverse dependencies
BDS [7]	Correlation dimension	Sensitive to overlapping vectors	Robust to vector construction; symmetric measure
$G(m)$ [29]	Permutation entropy	Non-chi-square asymptotic distribution	Weighted chi-square distribution; improved accuracy
Jiang et al. [25]	Distance covariance (metric spaces)	Computationally intensive	Simpler computation; generalizes to standard series
Zhou & Müller [42]	Profile association measures	Complex implementation	Simpler quantile-based approach
Combettes [10]	Symbolic representations	Limited to specific symbolization	Flexible quantile symbolization
Weiß & Schnurr [40]	Generalized ordinal patterns	Requires ordinal transformation	Generalizes to continuous series
Liu et al. [27]	Kernel-based joint independence	High computational cost	Lower computational complexity
Anjali et al. [2]	Golden-section symbolic representation	Specific to symbolic framework	Broader applicability with quantile symbolization
Wang et al. [38]	Series-symbol data generation	Complex model training	No training required; direct testing
Betken et al. [6]	Ordinal patterns for change point	Not designed for independence	Focused independence testing
He et al. [22]	Non-parametric Symbolic Approx. (NSAR)	Application-specific	General-purpose independence testing
Hassani et al. [21]	White noise diagnostics	Not a direct independence test	Directly tests independence
Yu et al. [41]	Semiparametric latent ANOVA	Domain-specific	Broad time series applicability
Mohammadi et al. [30]	Model-free prediction	Not focused on independence	Specific independence testing
Hounyo & Lin [24]	Wild bootstrap inference	Inference-focused, not testing	Direct test with bootstrap option
Wang et al. [39]	State correlations in heterogeneous series	Specific to heterogeneous data	Generalizes to homogeneous series
Ljung-Box [26]	Autocorrelation	Limited to linear dependencies	Captures nonlinear dependencies
Runs [37]	Sequence randomness	Low power for complex dependencies	Higher size-corrected power
Proposed $DC_\phi$	Symmetric Phi-divergence with quantile symbolization	Requires parameter tuning	Generalizes methods; robust across applications

Various tests have been proposed to assess serial independence in time series, ranging from traditional divergence-based methods to recent symbolic dynamics approaches. As summarized in Table 1, we compare these methods with our proposed test, emphasizing its superior robustness and power. This study introduces a novel class of statistical tests grounded in Phi-divergence, building on and extending the methodology established by [4]. The proposed framework leverages the symmetric properties of Phi-divergence to address critical limitations of traditional divergence measures, particularly in the context of time series analysis. A key innovation of our approach is its emphasis on time-reversal invariance (robustness to the direction of time), a property essential for robust dependency detection in applications where the direction of time is not inherently meaningful. By incorporating symmetric divergence measures, such as Jensen-Shannon and total variation divergences, our test statistic offers a balanced and versatile tool for independence testing, advancing the growing literature on permutation entropy-based methods.

The detection of dependencies in time series has seen significant progress through entropy-based approaches, which provide robust frameworks for analyzing complex datasets. Early methods, such as those relying on Kullback-Leibler (KL) divergence, focused on asymmetric measures to quantify dependencies [4]. Although effective in certain contexts, these approaches are limited by their sensitivity to the direction of divergence, which can obscure symmetric relationships in time series data. Subsequent advances, notably by [3], introduced permutation-entropy-based tests that improved accuracy by addressing these limitations, demonstrating that test statistics follow a weighted chi-square distribution and incorporating bootstrap techniques for improved performance. Further refinements by integrated quantile symbolization [4] with KL divergence, achieving high precision at varying sample sizes.

Despite these developments, asymmetric divergence measures such as KL divergence are less effective in scenarios where time-reversal invariance is critical, such as when assessing whether a time series exhibits symmetric behavior under temporal reversal. In such cases, the direction of divergence is irrelevant and asymmetric measures may not capture the full structure of dependencies. Our proposed framework overcomes this challenge employing symmetric Phi-divergence, which generalizes the methodology of [4] and ensures invariance to time reversal. This symmetry is particularly valuable in time series analysis, where processes like the AR(1) model may exhibit directional biases that symmetric measures (see Example 3.3).

In Section 2, we provide an explanation of symmetric phi-divergence. Section 3 focuses on the compatibility of the proposed test, detailing its methodology and key features. In Section 4, we perform simulations to evaluate the proposed test against competing methods. Section 5 demonstrates the application of the proposed test, while Section 6 wraps up the article with concluding remarks.

## 2. Symmetric Phi-divergence

Phi-divergence is a valuable tool used to evaluate the differences between probability density functions or distribution functions and has proven effective in developing tests for independence. This measure can be mathematically defined as follows:

$$D_\phi(f, f_0) = \int_{\mathbb{R}^m} f_0(\mathbf{x}) \phi\left(\frac{f(\mathbf{x})}{f_0(\mathbf{x})}\right) d\mu, \quad (2.1)$$

where the distributions  $F_0$  and  $F$  are absolutely continuous with respect to the measure  $\mu$  and have corresponding densities  $f_0(\mathbf{x})$  and  $f(\mathbf{x})$ . Here,  $\mathbf{x}$  represents an  $m$ -dimensional vector, and  $\phi(x) : [0, \infty) \rightarrow \mathbb{R}$  is a convex function that satisfies specific conditions, including  $\phi(1) = 0$ ,  $0\phi(0/0) = 0$ , and  $0\phi(u/0) = u \lim_{t \rightarrow \infty} \phi(t)/t$  [31].

By choosing appropriate  $\phi$  functions, a variety of established divergence measures can be formulated, including Kullback-Leibler [4, 11], Pearson  $\chi^2$  [12], Balakrishnan-Sanghvi [5], Triangular divergence [31], Cressie-Read [12], Minimum Discrimination Information [31], Jeffreys distance [31], Hellinger distance [31], Total Variation [31], Jensen-Shannon [32], Power divergence [12], and Rukhin divergence [35]. In the following, we describe these measures, focusing on their mathematical characteristics, sensitivity to distributional features, and typical use cases, followed by a comparative discussion of their distinctions.

**Kullback-Leibler (KL).** The Kullback-Leibler (KL) divergence, defined by  $\phi(x) = x \log(x)$ , measures the loss of information when approximating  $f$  with  $f_0$ . It is asymmetric, meaning  $D_\phi(f, f_0) \neq D_\phi(f_0, f)$ , reflecting its directional nature in information theory [11, Page 18]. KL divergence is highly sensitive to differences in the tails of distributions due to the logarithmic term, which amplifies large ratios  $f(\mathbf{x})/f_0(\mathbf{x})$ . This makes it ideal for applications in information theory, such as model comparison in machine learning or serial independence testing [4]. Its asymmetry and potential for infinite values (when  $f_0(\mathbf{x}) = 0$  but  $f(\mathbf{x}) > 0$ ) require careful application.

**Pearson  $\chi^2$  (PE).** The Pearson  $\chi^2$  divergence, given by  $\phi(x) = \frac{1}{2}(x - 1)^2$ , is symmetric, satisfying  $D_\phi(f, f_0) = D_\phi(f_0, f)$ . Its quadratic form emphasizes large deviations between  $f(\mathbf{x})$  and  $f_0(\mathbf{x})$ , particularly in the central regions of distributions [12]. Symmetry and differentiability make it a standard choice for hypothesis testing and goodness-of-fit tests, such as in multinomial models. Its sensitivity to central deviations rather than tails distinguishes it from other measures.

**Balakrishnan-Sanghvi (BS).** The Balakrishnan-Sanghvi divergence, defined by  $\phi(x) = \frac{(x-1)^2}{(x+1)^2}$ , is symmetric and bounded in  $[0, 1]$ . The bounded nature improves the robustness to extreme values, making it suitable for analyzing attribute-based distances in categorical data [5]. The denominator  $(x + 1)^2$  reduces the impact of large deviations, distinguishing it from the Pearson divergence.

**Triangular Divergence (TD).** The triangular divergence, with  $\phi(x) = \frac{(1-x)^2}{1+x}$ , is asymmetric and emphasizes differences in small probabilities due to the denominator  $1 + x$ . This property makes it appropriate for applications that focus on rare events, such as reliability analysis [31]. Its asymmetry limits its use in contexts that require bidirectional comparisons.

**Cressie-Read (CR).** The Cressie-Read divergence, given by  $\phi(x) = \frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}$ , forms a general family that includes other divergences as special cases (e.g., Pearson when  $\lambda = 1$ , KL as  $\lambda \rightarrow 0$ ). Its properties, such as symmetry, depend on the parameter  $\lambda$ , offering flexibility for goodness-of-fit tests in statistical inference [12].

**Minimum Discrimination Information (MD).** The MD divergence, with  $\phi(x) = -\log(x) + x - 1$ , is asymmetric and equivalent to the KL divergence up to first-order approximations. It is used in information-theoretic contexts where minimizing discrimination information is critical, such as in time series analysis [31]. Like KL, it is sensitive to differences in the tails, limiting its robustness in some settings.

**Jeffreys Distance (JD).** The Jeffreys distance, defined by  $\phi(x) = (x - 1) \log(x)$ , is a symmetric version of the KL divergence. This symmetry makes it suitable for bidirectional comparisons, such as in statistical inference for time series models [31]. Its logarithmic term ensures sensitivity to tail differences, similar to KL divergence.

**Hellinger Distance (HE).** The Hellinger distance, with  $\phi(x) = \frac{1}{2}(\sqrt{x}-1)^2$ , is symmetric and bounded in  $[0, 1]$ . The square root formulation dampens large deviations, enhancing robustness to outliers. It is widely used in statistical applications that require robust measures, such as in hypothesis testing [31].

**Total Variation (TV).** The divergence of total variation, given by  $\phi(x) = |1 - x|$ , is a symmetric metric that measures the maximum difference between the distributions. Its strength as a metric makes it robust for applications that require strict bounds, such as robust hypothesis testing in statistical models [31]. However, it is non-differentiable at  $x = 1$ , which can complicate optimization.

**Jensen-Shannon (JS).** The Jensen-Shannon divergence, defined by

$$\phi(x) = \frac{1}{2} \left\{ x \log(x) - (x+1) \log\left(\frac{x+1}{2}\right) \right\},$$

is symmetric and bounded in  $[0, 1]$ . Its connection to mutual information makes it popular in applications like entropy-based testing for time series dependence [32, Page 147].

**Power Divergence (PD).** The power divergence, with  $\phi(x) = \frac{x^{\nu+1} - x - \nu(x-1)}{\nu(\nu+1)}$  for  $\nu \neq 0, -1$ , generalizes several divergences (e.g., Pearson at  $\nu = 1$ , KL as  $\nu \rightarrow 0$ ). Its properties depend on  $\nu$ , providing flexibility for tailored statistical tests, such as multinomial goodness-of-fit [12].

**Rukhin Divergence (RU).** The Rukhin divergence, defined by  $\phi(x) = \frac{(1-x)^2}{2(\nu+(1-\nu)x)}$  for  $0 \leq \nu \leq 1$ , is a weighted version of Pearson divergence. Its adaptability through the parameter  $\nu$  makes it suitable for specific statistical applications, such as mixture parameter estimation [35].

The choice of Phi-divergence depends on the application requirements. Symmetric measures such as Jeffrey's distance, Jensen-Shannon distance, and Hellinger distance are ideal for bidirectional comparisons in statistical inference or entropy-based testing [32]. For robustness, total variation and Hellinger distance excel, as their formulations mitigate large deviations; the direct measurement of maximum difference of total variation is particularly robust in hypothesis testing under minimal assumptions [31]. Pearson and Cressie-Read divergences, being differentiable, facilitate optimization in statistical inference, especially for goodness-of-fit tests [12]. KL divergence, foundational in information theory, is less robust due to its asymmetry and tail sensitivity, but Jeffreys distance addresses this by enforcing symmetry [31]. For hypothesis testing, Pearson or power divergence (with  $\nu = 1$ ) are preferred for their statistical properties and computational ease [12]. In categorical data analysis, the bounded Balakrishnan-Sanghvi divergence is advantageous for its robustness to varying sample sizes [5]. Sensitivity to specific distributional features tails (KL, JD), central deviations (PE), or small probabilities (TD) further guides the selection of an appropriate measure.

However, a critical limitation of many commonly used Phi-divergences, such as Kullback-Leibler and triangular divergences, is their lack of symmetry, meaning that  $D_\phi(f, f_0) \neq D_\phi(f_0, f)$ . This asymmetry can introduce directional bias in applications where fairness, balance, or reciprocal comparison is essential. To overcome this issue and ensure an unbiased bidirectional assessment, a symmetrized divergence can be constructed by defining the function  $\eta(x) = \frac{1}{2} \left( \phi(x) + x\phi\left(\frac{1}{x}\right) \right)$ . The resulting symmetric divergence is given by [31]:

$$D_\eta(f, f_0) = \frac{D_\phi(f, f_0) + D_\phi(f_0, f)}{2}. \quad (2.2)$$

In statistical inference, symmetrization is not merely a mathematical refinement but is essential for tasks requiring impartiality and robustness, particularly in time series analysis

where time-reversal invariance plays a critical role. By eliminating directional dependence, symmetric divergences, such as the time-reversal invariant measure  $D_\eta$  used in serial independence testing, provide a principled and reliable framework for hypothesis testing and model selection, ensuring interpretability and fairness while robustly analyzing temporal dependencies across both directions of time [33]. In the following section, after introducing our test statistic based on  $D_\eta(f, f_0)$ , we will evaluate its performance through a simulation example to investigate this capability.

### 3. The proposed test: Methodology and core characteristics

To develop our independence test, we first formalize the dependence structure that we aim to detect. The concept of  $m$ -dependence plays a central role in characterizing the memory of the process:

**Definition 3.1** ([19]). The stochastic process  $\{Y_t, t \in \mathbb{Z}\}$  is defined as  $m$ -dependent if every two consecutive vectors are separated by at least  $m$  time units. This means that any two vectors, such as  $(Y_t, Y_{t+1}, \dots, Y_{t+i})$  and  $(Y_{t+j}, Y_{t+j+1}, \dots, Y_{t+r})$ , are independent for all nonnegative integers  $i, j$ , and  $r$  whenever  $j - i > m$ .

To test  $H_0$  using the sample  $\{Y_t; t = 1, 2, \dots, n\}$ , an  $m$ -dimensional block can be formed as  $\mathbf{Y}_t = (Y_t, Y_{t+1}, \dots, Y_{t+m-1})^T$ , for  $t = 1, 2, \dots, n - m + 1$ . If we define  $Q_{\frac{1}{d}}$  as the  $\frac{1}{d}$ -quantile of the variables  $\{Y_t; t = 1, 2, \dots, n\}$ , the null hypothesis can be rewritten as:

$$H_0 : P(\mathbf{Y}_t \in \mathbf{A}_i) = \left(\frac{1}{d}\right)^m, \quad i = 1, 2, \dots, d^m, \quad (3.1)$$

where  $\mathbf{A}_i = A_{i_1} \times A_{i_2} \times \dots \times A_{i_m}$ , with  $A_i$ s belonging to the set

$$\vartheta = \{(-\infty, Q_{\frac{1}{d}}], (Q_{\frac{1}{d}}, Q_{\frac{2}{d}}], \dots, (Q_{\frac{d-1}{d}}, \infty)\}.$$

#### 3.1. Formulation of test statistics

By substituting  $f(\mathbf{y}_t)$  and  $f_0(\mathbf{y}_t)$  (which equals  $\frac{1}{d^m}$  under  $H_0$  in  $\mathbf{A}_i$ ) into  $D_\eta(f_0, f)$ , equation (2.2), we get:

$$DQ_\phi = \frac{1}{2} \sum_{i=1}^{d^m} \left\{ \int_{\mathbf{A}_i} \frac{1}{d^m} \phi(d^m f(\mathbf{y}_t)) d\mu + \int_{\mathbf{A}_i} f(\mathbf{y}_t) \phi\left(\frac{1}{d^m f(\mathbf{y}_t)}\right) d\mu \right\}. \quad (3.2)$$

We can define  $\{\delta_t; t = 1, 2, \dots, n\}$  such that

$$\delta_t = \begin{cases} 1 & Y_t \in (-\infty, Q_{\frac{1}{d}}], \\ 2 & Y_t \in (Q_{\frac{1}{d}}, Q_{\frac{2}{d}}], \\ \vdots & \\ d & Y_t \in (Q_{\frac{d-1}{d}}, \infty). \end{cases}$$

Each symbol  $c_i$  corresponds to an  $m$ -dimensional vector containing  $\delta_t$  values. To map this transformation, the function  $\delta_t$  is applied to each element of the vector  $\mathbf{Y}_t$ . The set  $G = \{c_1, c_2, \dots, c_R\}$  consists of symbols constructed as  $m$ -dimensional vectors, where each component is selected from the discrete set  $\{1, 2, \dots, d\}$ , and  $R = d^m$ .

We can then define a function  $g : \mathbb{R}^m \rightarrow G$  on the vector  $\mathbf{Y}_t$ , applying  $\delta_t$  to each of its components. This creates a discrete mapping that represents different states or categories within the data. Thus, each symbol  $c_i$  corresponds to a vector formed from the sequence  $(\delta_t, \delta_{t+1}, \dots, \delta_{t+m-1})^T$ , capturing the behavior of the data over the interval from  $t$  to  $t + m - 1$ .

The probability of a symbol  $c_i$ , denoted  $p_{c_i}$ , is defined as  $P(\mathbf{Y}_t \in \mathbf{A}_i)$ . An estimate of  $p_{c_i}$  is given by  $\hat{p}_{c_i} = \frac{1}{K} \sum_{t=1}^K W_{c_i,t}$ , where  $K = n - m + 1$ , and:

$$W_{c_i,t} = \begin{cases} 1 & g(\mathbf{Y}_t) = c_i; \\ 0 & g(\mathbf{Y}_t) \neq c_i. \end{cases}$$

If the  $Y_t$  values are identically distributed, then it  $W_{c_i,t}$  follows a Bernoulli distribution with a success probability of  $p_{c_i}$ . Consequently, the probabilities  $p_{c_i}$  must satisfy the condition  $\sum_{i=1}^R p_{c_i} = 1$ .

Thus, we can approximate the integral  $\int_{\mathbf{A}_i} f(\mathbf{y}_t) \phi\left(\frac{1}{d^m f(\mathbf{y}_t)}\right)$  by using  $\hat{p}_{c_i} \phi\left(\frac{1}{d^m \hat{p}_{c_i}}\right)$  and estimate  $\int_{\mathbf{A}_i} \frac{1}{d^m} \phi\left(d^m f(\mathbf{y}_t)\right)$  by  $\frac{1}{d^m} \phi\left(d^m \hat{p}_{c_i}\right)$ . Using these estimates, we can compute an approximation of  $DQ_\phi$  based on quantile symbolization as follows:

$$DC_\phi = \frac{1}{2} \sum_{i=1}^{d^m} \left\{ \frac{1}{d^m} \phi\left(d^m \hat{p}_{c_i}\right) + \hat{p}_{c_i} \phi\left(\frac{1}{d^m \hat{p}_{c_i}}\right) \right\}.$$

**Remark 3.2.** The following two asymmetric test statistics can be derived from  $DC_\phi$  using the following relationship:

$$DC1_\phi = \frac{2K}{\phi''(1)} \sum_{i=1}^{d^m} \frac{1}{d^m} \phi\left(d^m \hat{p}_{c_i}\right),$$

$$DC2_\phi = \frac{2K}{\phi''(1)} \sum_{i=1}^{d^m} \hat{p}_{c_i} \phi\left(\frac{1}{d^m \hat{p}_{c_i}}\right).$$

By choosing  $\phi(x) = x \log(x)$  in  $DC1_\phi$ , we obtain a test statistic that aligns with the one proposed by [4]. In accordance with the specified assumptions, when  $d = 2$ , this statistic closely resembles the test for spatial data introduced in [28].

Time-Reversal Invariance refers to a property of a statistical measure or process that remains unchanged when the direction of time is reversed (i.e., when the sequence of observations is inverted). For a time series  $\{Y_t; t = 1, 2, \dots, n\}$ , a measure is invariant in time-reversal if it yields the same result for  $Y_t$  and its reversed series  $Y_{n-t+1}$  [20]. In the following, we will examine the property of  $DC_\phi$  through a practical example.

**Example 3.3.** To evaluate the time-reversal invariance properties of divergence measures, we conducted a simulation study comparing the  $DC1_{KL}$  and  $DC2_{KL}$  divergences, as well as the symmetric KL divergence ( $DC_{KL}$ ), for an autoregressive (AR(1)) process. The study aimed to quantify the sensitivity of these three measures to time reversal, providing information on their suitability for symmetry-based analyses. By generating multiple realizations of a time series and computing divergences for both the original and time-reversed sequences, we analyzed the distribution of the resulting differences. The results, visualized in Figure 1, include all three divergence measures, where the symmetric KL divergence ( $DC_\phi$ ) is also known as Jeffrey's divergence.

The simulation was designed to model a stationary time series with moderate temporal dependence. We generate  $n_{\text{sim}} = 100$  independent realizations of an AR(1) process defined by  $Y_t = 0.6Y_{t-1} + \epsilon_t$ , where  $\epsilon_t \sim N(0, 1)$ , and each series had length  $n = 200$ . The autoregressive coefficient  $\phi = 0.6$  was selected to ensure a realistic level of autocorrelation. Each time series was transformed into a symbolic representation using an  $m = 2$ -dimensional embedding to capture pairwise dependencies, quantized into  $L = 4$  intervals based on empirical quartiles.

For each realization, we computed divergence values for both the original series  $\{Y_t; t = 1, 2, \dots, n\}$  and the time-reversed series  $\{Y_{n-t+1}, t = 1, 2, \dots, n\}$ . The three divergence measures  $DC1_{KL}$ ,  $DC2_{KL}$ , and  $DC_{KL}$  were evaluated, where  $DC_{KL}$  represents the special

case of  $DC_\phi$  with  $\phi(u) = u \log u$ . To assess time-reversal invariance, we computed the differences  $\Delta_1 = DC1_{KL}(Y_t) - DC1_{KL}(Y_{n-t+1})$ ,  $\Delta_2 = DC2_{KL}(Y_t) - DC2_{KL}(Y_{n-t+1})$ , and  $\Delta = DC_{KL}(Y_t) - DC_{KL}(Y_{n-t+1})$  across all simulations and analyzed their distributions.

The results, summarized in Figure 1, reveal significant differences in the time-reversal properties of the divergence measures. Both  $DC1_{KL}$  and  $DC2_{KL}$  exhibit noticeable sensitivity to time reversal, with mean absolute differences of  $|\Delta_1| = 0.0068$  and  $|\Delta_2| = 0.0038$ , respectively. In contrast, the corresponding value for the symmetric KL divergence  $|\Delta| = 0.0019$  indicates substantially lower sensitivity. The density distributions, shown in Figure 1, are broad yet symmetric around zero for  $DC1_{KL}$  and  $DC2_{KL}$ , suggesting similar behavior in both directions. However, for  $DC_\phi$ , the distribution is sharply concentrated near zero, with a pronounced peak at zero itself highlighting its strong invariance under time reversal.

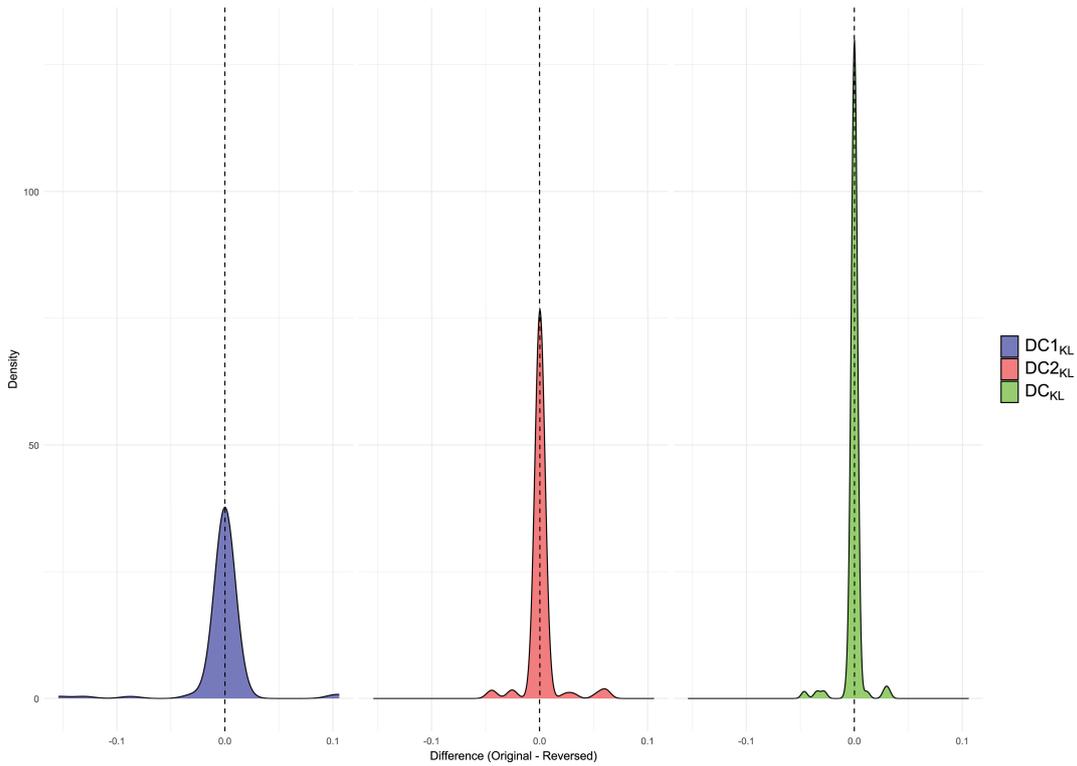


Figure 1. Time-reversal sensitivity of Kullback-Leibler

### 3.2. Asymptotic properties of test statistics

By employing the  $DC_\phi$ , we develop tests for independence that reject the null hypothesis when the calculated estimates surpass a specified high threshold. Suppose  $Q_{1-\alpha}$  as the  $1 - \alpha$  quantile of the  $DC_\phi$  distribution, establishing the critical region as  $DC_\phi > Q_{1-\alpha}$ . Given that the distribution of  $\{Y_t; t = 1, 2, \dots, n\}$  under the null hypothesis is unknown, we either utilize the asymptotic distribution or implement a bootstrap method to estimate  $Q_{1-\alpha}$ .

Let  $\mathbf{P} = (p_{c_1}, p_{c_2}, \dots, p_{c_R})^T$  and define

$$\hat{\mathbf{P}} = \left( \frac{1}{K} \sum_{t=1}^K W_{c_1,t}, \dots, \frac{1}{K} \sum_{t=1}^K W_{c_R,t} \right)^T .$$

It follows that

$$\sqrt{K}(\hat{\mathbf{P}} - \mathbf{P}) \xrightarrow{K \rightarrow \infty} N(0, \mathbf{\Sigma}),$$

where

$$\Sigma = \text{diag}(\mathbf{P}) - (2m + 1)\mathbf{P}\mathbf{P}^T + \text{diag}(\mathbf{P}) \sum_{l=1}^m \mathbf{Q}^{(l)} + \sum_{l=1}^m \mathbf{Q}^{(l)T} \text{diag}(\mathbf{P}). \quad (3.3)$$

In relation to equation (3.3), for  $1 \leq l \leq m$ , we define  $q_{i,j}^l = P(W_{c_i,t+l} = 1 \mid W_{c_j,t} = 1)$ . The matrix  $\mathbf{Q}^{(l)} = \{q_{i,j}^l\}_{i,j=1,2,\dots,R}$  represents the transition probabilities corresponding to a time lag of  $l$ . Under the null hypothesis  $H_0$ , the vector  $\mathbf{P}$  is expressed as  $\mathbf{P}^0 = \left(\frac{1}{R}, \dots, \frac{1}{R}\right)^T$ , and the matrix  $\Sigma$  is updated to  $\Sigma^0$  [16, Theorem 1]. Taking these assumptions into account, the asymptotic distribution of  $DC_\phi$  is outlined in the following theorem.

**Theorem 3.4.** *Suppose the random variables  $Z_1, Z_2, \dots, Z_r$  follow a standard normal distribution. Under  $H_0$ , we have*

$$\frac{2K}{\phi''(1)} DC_\phi \xrightarrow[n \rightarrow \infty]{d} \sum_{i=1}^r \lambda_i Z_i^2,$$

where  $\lambda_i$ , for  $i = 1, 2, \dots, r$ , are the eigenvalues of the matrix  $\mathbf{A}\Sigma^0$ , where  $\mathbf{A} = \text{diag}(\mathbf{P}^0)$ .

**Proof.** Let  $R = d^m$  denote the total count of symbols. We define  $\eta(x) = \frac{1}{2}(\phi(x) + x\phi(\frac{1}{x}))$  and  $\gamma(x_1, x_2, \dots, x_R) = \sum_{i=1}^R p_{c_i}^0 \eta\left(\frac{x_i}{p_{c_i}^0}\right)$ . By applying Taylor's expansion to  $\gamma$  at the point  $\mathbf{P} = (p_{c_1}, p_{c_2}, \dots, p_{c_R})^T$ , we can express

$$\begin{aligned} \gamma(\hat{p}_{c_1}, \hat{p}_{c_2}, \dots, \hat{p}_{c_R}) &= \gamma(p_{c_1}, p_{c_2}, \dots, p_{c_R}) \\ &+ \sum_{i=1}^{R-1} \frac{\partial \gamma(p_{c_1}, p_{c_2}, \dots, p_{c_R})}{\partial p_{c_i}} (\hat{p}_{c_i} - p_{c_i}) \\ &+ \frac{1}{2} \sum_{i,j=1}^{R-1} \frac{\partial^2 \gamma(p_{c_1}, p_{c_2}, \dots, p_{c_R})}{\partial p_{c_i} \partial p_{c_j}} (\hat{p}_{c_i} - p_{c_i})(\hat{p}_{c_j} - p_{c_j}) \\ &+ o(O_p(n^{-1})). \end{aligned} \quad (3.4)$$

Under the null hypothesis  $H_0$ , it follows that  $\mathbf{P} = \mathbf{P}^0$ , which implies that  $\gamma(p_{c_1}^0, p_{c_2}^0, \dots, p_{c_R}^0) = 0$ . Additionally, we find that

$$\left(\frac{\partial \gamma(p_{c_1}, p_{c_2}, \dots, p_{c_R})}{\partial p_{c_i}}\right)_{\mathbf{P}=\mathbf{P}^0} = 0,$$

and

$$\left(\frac{\partial^2 \gamma(p_{c_1}, p_{c_2}, \dots, p_{c_R})}{\partial p_{c_i} \partial p_{c_j}}\right)_{\mathbf{P}=\mathbf{P}^0} = \begin{cases} \frac{\phi''(1)}{p_{c_R}^0}, & i \neq j \\ \phi''(1) \left(\frac{1}{p_{c_i}^0} + \frac{1}{p_{c_R}^0}\right), & i = j \end{cases}. \quad (3.5)$$

Consequently, the initial two terms in equation (3.4) equal to zero, leading the third term to reduce to

$$\phi''(1) \left\{ \frac{1}{2} \sum_{i,j=1, i \neq j}^{R-1} \frac{(\hat{p}_{c_i} - p_{c_i}^0)(\hat{p}_{c_j} - p_{c_j}^0)}{p_{c_R}^0} + \sum_{i=1}^{R-1} \frac{(\hat{p}_{c_i} - p_{c_i}^0)^2}{p_{c_i}^0} \right\}. \quad (3.6)$$

From this, we can conclude that

$$\frac{2K}{\phi''(1)} DC_\phi = K \sum_{i=1}^R \frac{(\hat{p}_{c_i} - p_{c_i}^0)^2}{p_{c_i}^0} + \frac{2K}{\phi''(1)} o(O_p(n^{-1})). \quad (3.7)$$

Employing the asymptotic distribution

$$\sqrt{K}(\hat{\mathbf{P}} - \mathbf{P}^0) \xrightarrow[n \rightarrow \infty]{d} N(0, \Sigma^0),$$

and recognizing that  $\mathbf{A} = \text{diag}(\mathbf{P}^0)$ , we arrive at the conclusion [14, Corollary 2.1.]:

$$K(\widehat{\mathbf{P}} - \mathbf{P}^0)^T \mathbf{A}(\widehat{\mathbf{P}} - \mathbf{P}^0) = K \sum_{i=1}^R \frac{(\widehat{p}_{c_i} - p_{c_i}^0)^2}{p_{c_i}^0} \stackrel{d}{=} \sum_{i=1}^R \lambda_i Z_i^2, \quad \text{as } n \rightarrow \infty,$$

thereby completing the proof. □

A key characteristic of a test is its consistency, which signifies that its power asymptotically approaches 1. Previous independence tests, such as those of [23, 29, 34], have shown consistency for stationary and  $(m - 1)$ -dependent variables. This assumption is justified because dependence is only accounted for up to a certain order, with independence assumed beyond that point. Therefore, this leads to the following theorem under  $H_1$ .

**Theorem 3.5.** *If  $\{Y_t, t \in \mathbb{Z}\}$  is  $(m - 1)$ -dependent, then*

$$\lim_{n \rightarrow \infty} P\left(\frac{2K}{\phi''(1)} DC_\phi > c\right) = 1,$$

where  $c \in \mathbb{R}^+$ .

**Proof.** Let  $R = d^m$  represent the total number of symbols. We define the function  $\eta(x) = \frac{1}{2}(\phi(x) + x\phi(\frac{1}{x}))$  and introduce the expression

$$\gamma(x_1, x_2, \dots, x_R) = \sum_{i=1}^R p_{c_i}^0 \eta\left(\frac{x_i}{p_{c_i}^0}\right).$$

The Taylor expansion of  $\gamma$  around  $\mathbf{P} = (p_{c_1}, p_{c_2}, \dots, p_{c_R})^T$  is given by:

$$\begin{aligned} \gamma(\widehat{p}_{c_1}, \widehat{p}_{c_2}, \dots, \widehat{p}_{c_R}) &= \gamma(p_{c_1}, p_{c_2}, \dots, p_{c_R}) \\ &+ \sum_{i=1}^{R-1} \frac{\partial \gamma(p_{c_1}, p_{c_2}, \dots, p_{c_R})}{\partial p_{c_i}} (\widehat{p}_{c_i} - p_{c_i}) + o(O_p(n^{-\frac{1}{2}})). \end{aligned} \tag{3.8}$$

Taking the partial derivative of  $\gamma(p_{c_1}, p_{c_2}, \dots, p_{c_R})$  in  $\mathbf{P}^*$ , we obtain

$$\left(\frac{\partial \gamma(p_{c_1}, p_{c_2}, \dots, p_{c_R})}{\partial p_{c_i}}\right)_{\mathbf{P}=\mathbf{P}^*} = \eta'\left(\frac{p_{c_i}^*}{p_{c_i}^0}\right) - \eta'\left(\frac{p_{c_R}^*}{p_{c_R}^0}\right).$$

Substituting  $\mathbf{P}^*$  into (3.8) under the alternative hypothesis  $H_1$ , we arrive at

$$\sum_{i=1}^{R-1} \left(\eta'\left(\frac{p_{c_i}^*}{p_{c_i}^0}\right) - \eta'\left(\frac{p_{c_R}^*}{p_{c_R}^0}\right)\right) (\widehat{p}_{c_i} - p_{c_i}^*) = \sum_{i=1}^R \eta'\left(\frac{p_{c_i}^*}{p_{c_i}^0}\right) (\widehat{p}_{c_i} - p_{c_i}^*).$$

Multiplying both sides of (3.8) by  $\sqrt{K}$ , we get

$$\sqrt{K} (DC_\phi - DC_\phi^*) = \sqrt{K} \sum_{i=1}^R \eta'\left(\frac{p_{c_i}^*}{p_{c_i}^0}\right) (\widehat{p}_{c_i} - p_{c_i}^*) + \sqrt{K} o(O_p(n^{-\frac{1}{2}})), \tag{3.9}$$

where  $DC_\phi^*$  is  $DC_\phi$  with  $p_{c_i}^*$  substituted for  $\widehat{p}_{c_i}$ . Applying Slutskys theorem, we deduce that equation (3.9) and the term

$$\sqrt{K} \sum_{i=1}^R \eta'\left(\frac{p_{c_i}^*}{p_{c_i}^0}\right) (\widehat{p}_{c_i} - p_{c_i}^*)$$

share the same asymptotic distribution. Referring to [16, Theorem 1, page 8], we define

the vector  $\mathbf{S} = \left(\eta'\left(\frac{p_{c_1}^*}{p_{c_1}^0}\right), \dots, \eta'\left(\frac{p_{c_R}^*}{p_{c_R}^0}\right)\right)^T$ . Thus, we have

$$\sqrt{K} \sum_{i=1}^R \eta'\left(\frac{p_{c_i}^*}{p_{c_i}^0}\right) (\widehat{p}_{c_i} - p_{c_i}^*) = \sqrt{K} \mathbf{S}^T (\widehat{\mathbf{P}} - \mathbf{P}^*) \xrightarrow{K \rightarrow \infty} N(0, \mathbf{S}^T \boldsymbol{\Sigma} \mathbf{S}), \tag{3.10}$$

where  $\sigma^2(\mathbf{P}^*)$  is expressed as  $\mathbf{S}^T \boldsymbol{\Sigma} \mathbf{S}$ , and  $\boldsymbol{\Sigma}$  is specified in (3.3) by replacing  $2(m-1)$  with  $m$  and substituting  $\mathbf{P}^*$  for  $\mathbf{P}$ . Since  $K = n - m + 1$ , there is no distinction between  $K \rightarrow \infty$  and  $n \rightarrow \infty$  in this context.

By applying Slutskys theorem, we obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\frac{2K}{\phi''(1)} DC_\phi > c\right) &= 1 - \lim_{n \rightarrow \infty} P\left(\frac{\sqrt{K} (DC_\phi - DC_\phi^*)}{\sigma(\mathbf{P}^*)} < \frac{\sqrt{K}}{\sigma(\mathbf{P}^*)} \left(\frac{c\phi''(1)}{2K} - DC_\phi^*\right)\right) \\ &= 1 - \Phi(-\infty) = 1, \end{aligned}$$

where  $\Phi$  is the standard normal distribution function, completing the proof.  $\square$

As described in Theorem 3.4, the distribution of the test statistic can be determined through the asymptotic method, with the null hypothesis being tested using its corresponding quantile. However, this procedure presents several notable challenges. Firstly, calculating the asymptotic distribution requires the evaluation of the  $\boldsymbol{\Sigma}$  matrix, which varies based on the value of  $m$ . This dependency complicates the process, as the matrix must be recalculated for each different value of  $m$ . Even after addressing this issue, the challenge of convergence speed arises: the test accuracy is diminished for small sample sizes, as it is primarily reliable for larger samples. Furthermore, obtaining the distribution of the test statistic requires the second derivative of the  $\phi$  function. This adds another layer of difficulty, as some  $\phi$  functions, such as total variation, may lack a second derivative. These factors limit the practicality of the asymptotic approach, which is why we propose the bootstrap method as an alternative for the subsequent tests.

### 3.3. Testing by bootstrap

The serial independence tests can be performed using the bootstrap approach with the test statistic  $DC_\phi$ . To compute the P-value, we follow these steps with  $B$  iterations:

1. First, compute  $DC_\phi$  for the original data set.
2. Generate a resample (permutation) of the original data set.
3. For each  $i$ -th resample, calculate  $DC_{\phi,i}$ .
4. Repeat steps 2 and 3 to create the bootstrap distribution of  $DC_\phi$ .
5. The P-value is then calculated using

$$\text{P-value} = \frac{1 + \sum_{i=1}^B I\{DC_{\phi,i} > DC_\phi\}}{B + 1},$$

where  $I\{\cdot\}$  is the indicator function.

The null hypothesis  $H_0$  is rejected at the significance level  $\alpha$  if the P-value is less than  $\alpha$ .

Having established the theoretical foundations of the test statistic and its asymptotic properties, we now evaluate its empirical performance through comprehensive simulations. These experiments serve two purposes: (1) to validate the theoretical results under finite samples and (2) to compare the proposed method with existing approaches under controlled conditions.

## 4. Evaluating proposed and alternative tests using simulation

In this section, we conduct a simulation study to assess the performance of the proposed test by comparing it with other established tests. The focus is on evaluating both the statistical power and the accuracy of the tests under various conditions. Furthermore, the comparison is extended to include the size-corrected powers of the tests, providing a more comprehensive view of their effectiveness. A specific simulation for certain functions of  $\phi$ , as outlined in Section 2, has also been performed to demonstrate the practical application of the proposed test.

The simulations were conducted in  $R$  (version 4.1.2) with 10,000 replications and  $B = 1,000$  bootstrap iterations, using sample sizes of 50, 350, 600, and 1,000. To determine the optimal values for  $d$  and  $m$ , we refer to the study in 4.2. These values have been derived from extensive simulations and serve as a guide to optimal performance in practice. To determine the empirical size, simulation data were generated based on the models  $M1$  through  $M6$ . The empirical size, indicated by  $\alpha_n$ , was calculated as the proportion of null hypothesis rejections in 10,000 iterations. Dependent data was also generated for models  $M7$  through  $M12$ , while trends were included in models  $M13$  to  $M15$ . In the following, the specifics of these models are outlined. The empirical power, represented as  $\beta_n^*$ , was determined by calculating the percentage of null hypothesis rejections in the 10,000 iterations. Although the bootstrap-based implementation increases the computational cost compared to parametric counterparts, its runtime scales approximately linearly with the sample size when parallelized. For example, testing  $n = 1,000$  observations (with  $m = 3$ ,  $d = 4$ ) takes about 2.1 seconds (versus 0.7 seconds for the BDS test) on a standard workstation. This additional cost is considered acceptable given the substantial gain in statistical power. For large datasets, the method remains feasible by reducing  $B$  or employing distributed computing strategies.

We consider 15 simulation models, systematically categorized into three groups based on their statistical properties: independent processes, dependent processes, and trending processes to thoroughly evaluate the performance of the proposed test across a wide range of scenarios. Each group is designed to investigate specific aspects of the test size control and power, ensuring a comprehensive assessment under various conditions.

The first group, comprising models  $M1$  to  $M6$ , focuses on independent and identically distributed (i.i.d.) processes with varied marginal distributions to assess the test size under correct specification. Model  $M1$  serves as the baseline, employing a standard Gaussian distribution with  $\varepsilon_t \sim N(0, 1)$  to verify the validity under normality assumptions. In contrast,  $M2$  introduces a heavy-tailed, skewed distribution using  $\varepsilon_t \sim \chi_3^2$ , with a variance of 6, to test robustness to asymmetry. Model  $M3$  further challenges the test with an infinite-variance scenario, adopting  $\varepsilon_t \sim t_2$ . For bounded distributions,  $M4$  utilizes a uniform distribution,  $\varepsilon_t \sim U(0, 1)$ , while  $M5$  explores extreme bimodality through a U-shaped Beta distribution,  $\varepsilon_t \sim \text{Beta}(0.5, 0.5)$ . Finally,  $M6$  examines a bounded Gaussian-like structure with a truncated normal distribution,  $\varepsilon_t \sim \text{TN}(0, 1, -1.75, 1.75)$ . These models collectively span a range of kurtosis (1.8 to 9) and skewness (0 to  $\sqrt{8}$ ), providing a robust evaluation of the behavior of the test under various i.i.d. conditions.

The second group, models  $M7$  to  $M12$ , investigates the power of the test against various forms of temporal dependence and non-linear dynamics. Model  $M7$ , a sign autoregressive process, is defined as  $Y_t = 0.3, \text{sign}(Y_{t-2}) + \varepsilon_t$ , capturing nonlinear dynamics with discontinuous transitions. Model  $M8$  employs the logistic map,  $Y_t = 4Y_{t-1}(1 - Y_{t-1})$ , a deterministic chaotic system that tests the detection of complex nonlinearities. To evaluate sensitivity to volatility clustering,  $M9$  adopts a nonlinear moving average structure,  $Y_t = 0.8\varepsilon_{t-2}^2 + \varepsilon_t$ . Model  $M10$  combines autoregressive and heteroskedastic elements through  $Y_t = 0.8|Y_{t-2}|^{0.5} + 0.6\varepsilon_t$ , while  $M11$  introduces conditional heteroskedasticity with an ARCH model,  $Y_t = \sqrt{1 + 0.8Y_{t-1}^2}, \varepsilon_t$ . Extending this,  $M12$  incorporates persistent volatility via a GARCH model,  $Y_t = \sqrt{1 + 0.6Y_{t-1}^2 + 0.3h_{t-1}^2}, \varepsilon_t$ . These models, with periods of one to two, cover a variety of dependence structures critical to evaluating the ability of the test to detect non-linear and dynamic patterns.

The third group, models  $M13$  to  $M15$ , addresses robustness to mean non-stationarity through trending processes. Model  $M13$  represents a random walk,  $Y_t = Y_{t-1} + \varepsilon_t$ , serving as a benchmark for unit root processes. Model  $M14$  introduces a deterministic linear trend,  $Y_t = 0.01 + 0.01t + \varepsilon_t$ , while  $M15$  incorporates a non-linear trend,  $Y_t = 0.01 + t^{0.1} + \varepsilon_t$ , to evaluate the performance of the test under complex trend dynamics. Together, these

models ensure that the robustness of the test is rigorously tested against stochastic and deterministic non-stationarities.

This carefully curated set of fifteen models provides a comprehensive framework for evaluating the proposed test. By spanning i.i.d. processes with diverse distributional properties, dependent processes with nonlinear and dynamic structures, and trending processes with varying forms of nonstationarity, the simulation study ensures a thorough assessment of the test size control and power across practically relevant scenarios.

The efficiency of each test is evaluated using the size-corrected power, given by the formula:

$$\beta^* = \overline{\beta_n^*} - \overline{\alpha_n},$$

where  $\overline{\beta_n^*}$  represents the average empirical power and  $\overline{\alpha_n}$  denotes the average empirical size. Please note that the tests were executed at the nominal level of  $\alpha = 0.05$ .

#### 4.1. Comparative tests

A comprehensive simulation has been carried out for the competing tests, and the details of these tests are outlined below:

**Run test:** The run test, as introduced by [37], is employed not only for comparing populations of sizes  $n_1$  and  $n_2$  but also serves as a method to assess the independence of time series data. In this context, the number of runs, denoted by  $U$ , acts as a discrete random variable. Under the null hypothesis, the mean  $\mu_u$  and standard deviation  $\sigma_u$  of  $U$  are considered. The findings of Wald and Wolfowitz [37] demonstrate that as the sample size  $n$  approaches infinity, the standardized variable converges in distribution to a standard normal distribution:  $\frac{U - \mu_u}{\sigma_u} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$ . Consequently, the null hypothesis  $H_0$  is rejected when the following condition holds:  $\left| \frac{U - \mu_u}{\sigma_u} \right| > z_{1-\alpha/2}$ .

**BDS test:** The BDS test, proposed by [7], is another significant method for examining time series independence, relying on the correlation integral defined as  $C_m(\epsilon) = P(\|\mathbf{Y}_1 - \mathbf{Y}_2\| \leq \epsilon)$ , where  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  are  $m$ -dimensional vectors. The estimate of  $C_m(\epsilon)$  is calculated using a  $U$ -statistic, which is articulated by [13] as

$$C_{m,n}(\epsilon) = \frac{2}{(n-m+1)(n-m)} \sum_{i=2}^{n-m+1} \sum_{j=1}^i I\{\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq \epsilon\}.$$

Under the null hypothesis  $H_0$ , Broock et al. [7] established that

$$\sqrt{n} \frac{C_{m,n}(\epsilon) - (C_1(\epsilon))^m}{\sigma_{n,m}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

Thus, the null hypothesis is rejected when the following inequality is satisfied:

$$\left| \sqrt{n} \frac{C_{m,n}(\epsilon) - (C_1(\epsilon))^m}{\sigma_{n,m}} \right| > z_{1-\alpha/2}.$$

**G(m) test:** The  $G(m)$  test, based on permutation entropy, operates under the null hypothesis  $H_0: P(Y_{t+i_1} \leq Y_{t+i_2} \leq \dots \leq Y_{t+i_m}) = \frac{1}{m!}, \forall t \in \mathbb{Z}$ . If  $\hat{h}(m) = -\sum_{i=1}^m \hat{p}_{\pi_i} \log(\hat{p}_{\pi_i})$  represents the entropy estimate, then the test statistic is given by

$$G(m) = -2K[\hat{h}(m) - \log(m!).]$$

Under the null hypothesis, the statistic  $G(m)$  follows a  $\chi_{m!-1}^2$  distribution, and the null hypothesis is rejected when  $G(m) > \chi_{m!-1, 1-\alpha}^2$  [29].

**Ljung-Box test:** The Ljung-Box test, proposed by [26], is another essential tool for testing the independence of time series. The test statistic is computed as follows:

$$LB(m) = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k},$$

where  $n$  denotes the sample size and  $\hat{\rho}_k^2$  represents the autocorrelation at lag  $k$ . Under the null hypothesis, the distribution of  $LB(m)$  is characterized by a  $\chi^2$  distribution with  $m$  degrees of freedom. Therefore, the null hypothesis is rejected when  $LB(m) > \chi_{m,1-\alpha}^2$ . In the simulation study, the parameters for the *BDS* test were fixed at  $m = 3$  and  $\epsilon = 2\sigma$  across all sample sizes. For the tests  $G(m)$  and  $LB(m)$ , the parameter  $m$  was set to 3 for small sample sizes ( $n = 50$ ) and adjusted to 4 for moderate and large sample sizes ( $n = 350, 600, 1000$ ).

## 4.2. Determining optimal $m$ and $d$

Although the parameters  $m$  and  $d$  can be selected freely, their performance under both the null hypothesis ( $H_0$ ) and the alternative hypothesis ( $H_1$ ) must be carefully evaluated to ensure that the empirical type I error rate remains close to its nominal level (typically  $\alpha = 0.05$ ) and that the chosen  $m$  and  $d$  values maximize the test's power. To this end, we conducted extensive simulations for all  $\phi$  functions discussed previously, covering sample sizes from 30 to 1500 and all feasible combinations of  $(m, d)$  satisfying the constraint  $d^m \leq n$ . The primary objectives were to verify that the nominal error rate is maintained and to identify the optimal  $(m, d)$  pair that maximizes test power for each sample size. The results for the total variation metric are presented in Figure 2 as an illustrative example, with similar behavior observed across all other  $\phi$  functions.

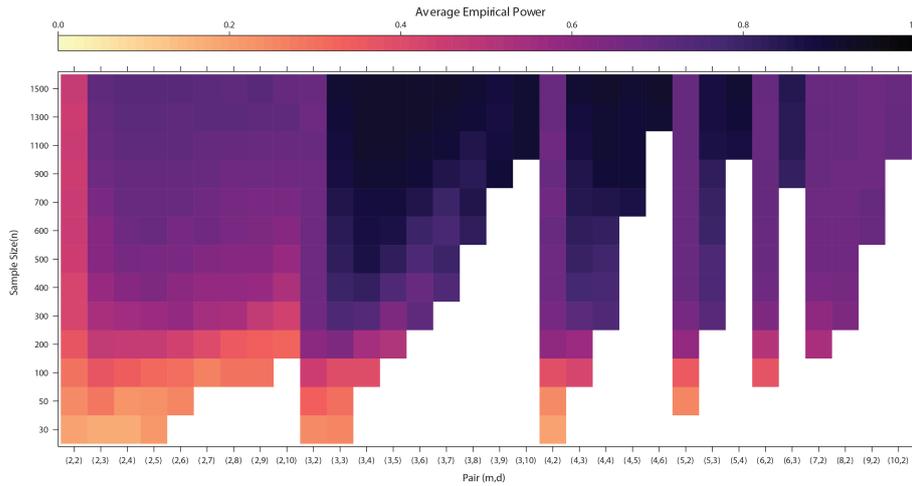
Figure 2 consists of two subfigures. Figure 2b illustrates the average empirical size, while Figure 2a visualizes the average empirical power for the total variation metric. For the empirical size (Figure 2b), the simulations confirmed that the nominal error rate (e.g. 0.05) is consistently maintained across all  $\phi$  functions and sample sizes, with minor deviations (e.g., within 0.045 to 0.055) for all  $(m, d)$  combinations. This robust control of the type I error rate, observed for total variation and all other  $\phi$  functions, ensures the reliability of the test.

The average empirical power is represented in Figure 2a using a heatmap, where the horizontal axis represents the  $(m, d)$  pairs (e.g.  $(2, 2), (2, 3), \dots, (9, 2), (10, 2)$ ), the vertical axis represents the sample size ( $n$ , ranging from 30 to 1500), and the color intensity reflects the average empirical power, with darker shades indicating higher power (closer to 1) and lighter shades indicating lower power. Key observations for total variation, which are consistent with other  $\phi$  functions, include: (1) test power increases with larger sample sizes for all  $(m, d)$  pairs, as evidenced by progressively darker shading along the vertical axis for higher  $n$ , reflecting the increased information available to detect deviations from  $H_0$ ; (2) power is generally higher when  $m = 3$ , as shown by darker regions for  $(m, d)$  pairs where  $m = 3$  (e.g.,  $(3, 3), (3, 4)$ ), suggesting that  $m = 3$  balances model complexity and detection capability; and (3) power varies across  $(m, d)$  pairs, with distinct patterns emerging at different sample sizes, captured through the gradient of color intensities.

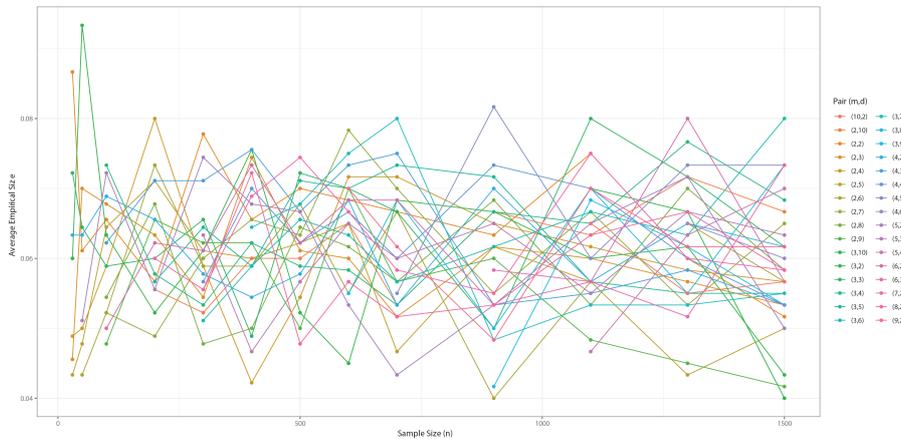
The optimal  $(m, d)$  pairs for Total Variation, which align with the findings of other  $\phi$  functions, were determined based on sample size: for  $n < 300$ , the pair  $(3, 3)$  is optimal, shown by the darkest regions for  $(3, 3)$  at lower sample sizes in Figure 2a; for  $400 \leq n \leq 700$ , the pair  $(3, 4)$  maximizes power, reflected by darker shading for  $(3, 4)$ ; for  $700 < n \leq 1100$ , the optimal pair is  $(3, 5)$ ; and for  $1100 < n \leq 1500$ , the pair  $(3, 6)$  is optimal. A general rule emerges: for sample sizes below 300,  $(3, 3)$  is recommended, and for every additional 400 samples,  $d$  increases by one unit (e.g.,  $(3, 4)$  for  $n \approx 400$ ,  $(3, 5)$  for  $n \approx 700$ ).

This trend, visually confirmed in Figure 2a for Total Variation, holds similarly for other  $\phi$  functions, with darker regions shifting to higher  $d$  values as the sample size increases.

In conclusion, Figure 2 serves as a critical visual tool summarizing the simulation results for Total Variation, with analogous patterns observed for all  $\phi$  functions. It confirms that the empirical type I error rate aligns closely with the nominal level across all combinations of  $(m, d)$  and that the test power increases with sample size, particularly when  $m = 3$ . The optimal pairs  $(m, d)$ , identified through the darkest regions in Figure 2a, provide practical guidance for parameter selection based on sample size, enhancing the reliability and efficiency of the statistical test in all  $\phi$  functions.



(a) Average empirical power.



(b) Average empirical size.

Figure 2. Total variation accuracy for different pairs of  $(m, d)$ .

### 4.3. Evaluating the effect of $\nu$ parameter in PD and RU divergences

In the preceding section, we determined the optimal  $(m, d)$  pair for various sample sizes, establishing a foundation for evaluating divergence criteria. Here, we extend this analysis to investigate the influence of the parameter  $\nu$  on two distinct divergence measures: Power divergence (PD) and Rukhin divergence (RU). These divergences generalize the concepts introduced earlier, offering flexibility in modeling by incorporating the tunable parameter  $\nu$ . Notably, as  $\nu$  approaches specific values (e.g., 0 or -1 for PD), the behavior of

Power divergence converges to well-known measures such as Kullback-Leibler divergence, providing a theoretical connection to established methods.

We conducted simulations using previously identified optimal pairs  $(m, d)$  to systematically assess the role of  $\nu$ . Our objective was to determine the value of  $\nu$  that maximizes the statistical power of each divergence criterion, quantified through the Mean Average Power (MAP). P is defined as the mean of the average empirical power  $\overline{\beta}_n$ , calculated across sample sizes ranging from 30 to 1500. This metric provides a robust measure of performance, as the nominal error remains stable across all tested  $\nu$  values for both PD and RU divergences. Consequently, the optimal  $\nu$  is the one that produces the highest MAP, reflecting the greatest ability to detect differences under the given divergence criterion.

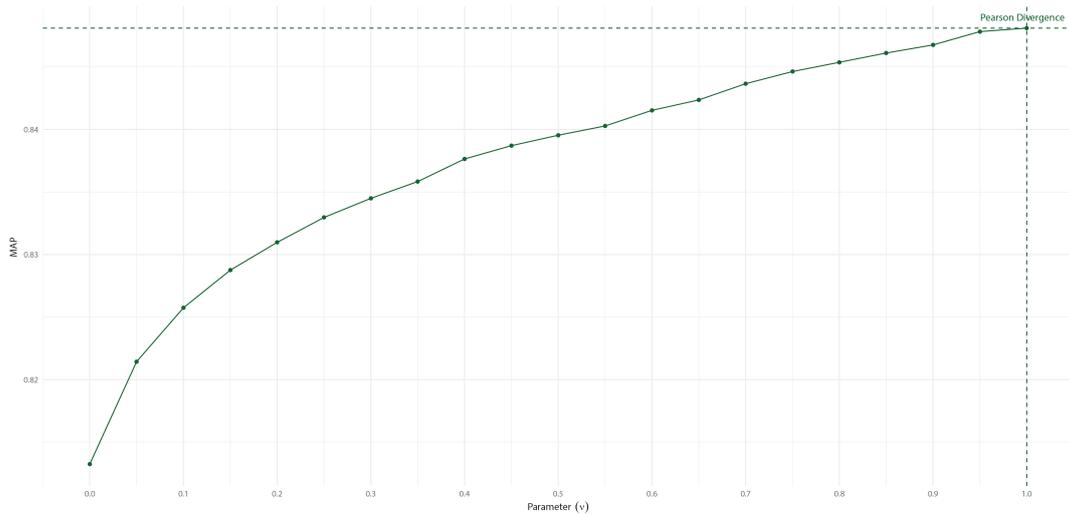
The results of these simulations are presented in Figure 3, which consists of two sub-figures that illustrate the relationship between  $\nu$  and MAP for each type of divergence. In Figure 3a, we show the MAP for Rukhin divergence as a function of  $\nu$ , with  $\nu$  ranging from 0 to 1 on the horizontal axis and MAP on the vertical axis. The plot reveals a clear trend: MAP increases steadily with  $\nu$  until it reaches a peak at  $\nu = 1$ . Beyond this point, further increases in  $\nu$  do not yield a higher power, establishing  $\nu = 1$  as the optimal value for the Rukhin divergence.

Similarly, Figure 3b illustrates the MAP for the power divergence over a selected range of  $\nu$  values (from -10 to 10) on the horizontal axis, with the MAP on the vertical axis. The trend for power divergence is more nuanced. MAP increases with  $\nu$  up to  $\nu = 2$ , where it reaches its maximum, and then decreases for higher values of  $\nu$ . This behavior indicates that  $\nu = 2$  is the optimal value for the power divergence, corresponding to the Pearson divergence, a well-known special case of PD. The peak at  $\nu = 2$  underscores the sensitivity of the power divergence to the choice of  $\nu$ , as values beyond this point lead to a reduction in statistical power. These findings, as visualized in Figure 3, highlight the critical role of  $\nu$  optimizing the performance of divergence-based methods. The selection of  $\nu = 1$  for the Rukhin divergence and  $\nu = 2$  for the Power divergence is not arbitrary but is supported by the empirical evidence presented in Figures 3a and 3b, respectively. These figures provide a clear visual representation of how MAP varies with  $\nu$ , allowing us to directly observe the trends that justify our parameter choices. Furthermore, the stability of the nominal error across all  $\nu$  values ensures that the observed differences in MAP are attributable to the power of the divergence criteria, strengthening the robustness of our results. The steady increase in MAP for the Rukhin divergence up to  $\nu = 1$  (Figure 3a) contrasts with the rise-and-fall pattern for the Power divergence around  $\nu = 2$  (Figure 3b), reflecting the distinct mathematical properties of these divergences. These visual insights are crucial for understanding the practical implications of our findings, particularly in applications where maximizing statistical power is paramount.

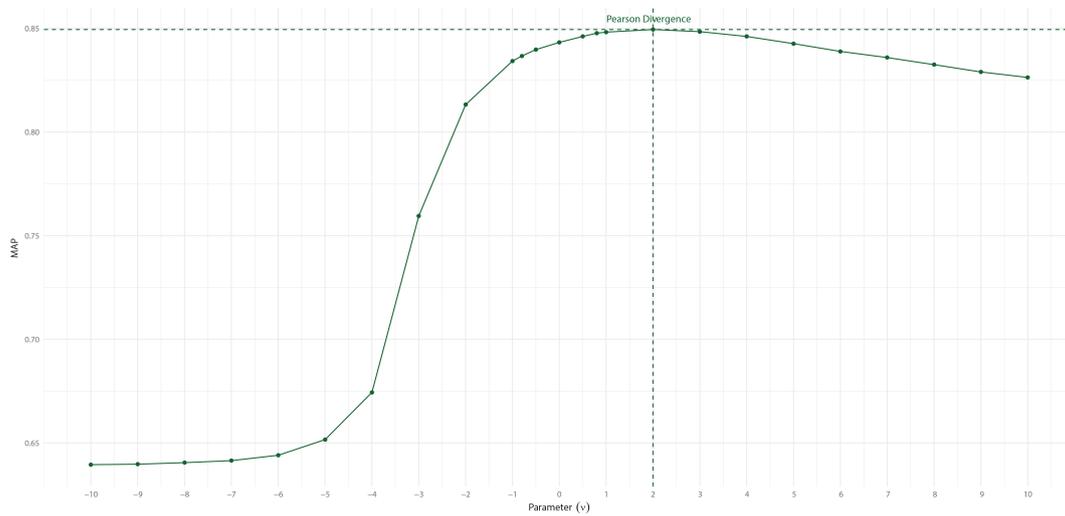
In summary, the simulations demonstrate that the optimal  $\nu$  values for the Rukhin and Power divergences are  $\nu = 1$  and  $\nu = 2$ , respectively, as these values maximize the mean average power (MAP) and thus the statistical power of the divergence criteria. Notably, at these optimal values, both divergences correspond to Pearson divergence, highlighting a unifying theoretical connection between Rukhin and Power divergences in their ability to maximize power. Figure 3 serves as the cornerstone of this analysis, providing a clear and compelling visualization of the relationship between  $\nu$  and MAP. These results not only validate the theoretical framework presented earlier, but also offer practical guidance for selecting  $\nu$  in divergence-based statistical methods, emphasizing the significance of Pearson's divergence in these contexts.

#### 4.4. Comprehensive evaluation of size-corrected power

Tables 10 through 13 display the simulation results across all sample sizes for a nominal Type I error of 0.05. To summarize the overall test performance, the  $\beta^*$  index was analyzed.



(a) Rukhin divergence.



(b) Power divergence.

**Figure 3.** Evaluation of the MAP index for  $RU$  and  $PD$  divergences.

In Table 2, the size-corrected power values for  $\beta^*$  are presented for the tests carried out at four different sample sizes. This table highlights that the proposed test,  $DC_\phi$ , outperforms the other tests in terms of size-corrected power.

In particular, the highest  $\beta^*$  values were observed for the tests  $CR$ ,  $TV$ ,  $BS$ , and  $TV$  in sample sizes of 50, 350, 600, and 1000, respectively. In particular, it shows strong performance in small samples (for example,  $n = 50$ , with  $\beta^*$  values like 0.37444 for  $CR$ ), and this advantage becomes even more pronounced as the sample size increases (e.g.,  $n = 350$ ,  $n = 600$ , and  $n = 1000$ ). In larger samples, particularly at  $n = 1000$ , the test achieves peak performance with high  $\beta^*$  values (e.g. 0.83000 for  $TV$ ). These results highlight the ability of the proposed test to maintain and improve its effectiveness in both small and large sample scenarios, making it a robust choice for various statistical applications.

The simulation results demonstrate the robustness of the proposed test in a wide range of scenarios. To further evaluate its practical utility, we applied it to diverse real-world datasets, including the Tehran Stock Price Index (TSPI), the SP 500 (GSPC), and the

**Table 2.** Analysis of size-corrected power across test types

Test	$n = 50$	$n = 350$	$n = 600$	$n = 1000$
<i>KL</i>	0.36722	0.64611	0.78389	0.81333
<i>PE</i>	0.36333	0.63389	0.75556	0.81056
<i>BS</i>	0.37111	0.65222	<b>0.79555</b>	0.81241
<i>TD</i>	0.37111	0.65222	0.79511	0.81139
<i>CR</i>	<b>0.37444</b>	0.64944	0.79513	0.81289
<i>MD</i>	0.36722	0.64611	0.78389	0.81159
<i>JD</i>	0.36722	0.64611	0.78389	0.81333
<i>HE</i>	0.36667	0.65222	0.78889	0.81500
<i>TV</i>	0.33389	<b>0.65944</b>	0.76556	<b>0.83000</b>
<i>JS</i>	0.37389	0.64954	0.79222	0.81500
<i>BDS</i>	0.33667	0.62944	0.75500	0.80000
<i>LB</i>	0.35278	0.53222	0.54889	0.55889
<i>G(m)</i>	0.26278	0.55444	0.60167	0.66333
<i>Run</i>	0.24111	0.28667	0.25667	0.26778

Lynx population series, where identifying dependencies can provide insight with economic or ecological significance.

## 5. Application

To illustrate the practical relevance and performance of the proposed test  $DC_\phi$ , this section applies it to three real-world time series data sets from financial and ecological domains. The analysis includes statistical summaries, model fitting, independence testing, and comparison with other existing tests. The results obtained are summarized and discussed to highlight the advantages of  $DC_\phi$ .

### 5.1. Application I: The Tehran stock price index (TSPI)

The predictability of stock prices is intrinsically linked to market efficiency, as efficient markets rapidly assimilate available information, reducing the predictability of stock prices. According to Fama et al. [17], a market is considered efficient if it promptly reacts to new information, thus limiting the possibilities of forecasting future stock prices.

The Tehran Stock Price Index (TSPI) data set spans 32 months, from March 19, 2014, with 639 observations, excluding holidays. The statistical summary (3) provides key insights into the behavior of TSPI. The mean index value is 9,788.223, indicating the average level over the period. The standard deviation of 1,202.647 reflects moderate volatility in the index. The minimum value of 8,016.8 and the maximum of 12,708 show the range of fluctuations, with a spread of 4,691.2 points. The kurtosis of -0.9004 suggests a platykurtic distribution, which implies fewer extreme values than a normal distribution. The positive skewness of 0.5061 indicates a slight rightward tilt, with more frequent values above the mean. These statistics highlight a relatively stable but slightly asymmetric index with moderate variability over the observed period.

**Table 3.** Statistical summary of Tehran stock price index

Mean	Standard Deviation	Minimum	Maximum	Kurtosis	Skewness
9788.223	1202.647	8016.8	12708	-0.9004	0.5061





**Figure 4.** Tehran stock price index: time series representation

## 5.2. Application II: S&P 500 index (GSPC)

The S&P 500 index, denoted as  $\hat{GSPC}$  in financial data repositories such as Yahoo Finance, is one of the most widely recognized benchmarks for the U.S. equity market. It reflects the performance of 500 leading publicly traded companies listed on the New York Stock Exchange (NYSE) and NASDAQ. Due to its broad sectoral coverage and market capitalization weighting, the S&P 500 serves as a comprehensive indicator of the overall dynamics and health of the US stock market and is frequently used in empirical financial research to investigate market behavior, volatility, and efficiency.

In this study, daily closing prices of the S&P 500 index were obtained from Yahoo Finance using the `quantmod` package in R, covering the period from January 1, 2022, to the date of analysis. The dataset consists of 841 observations (excluding non-trading days such as weekends and holidays), providing a consistent and reliable time series of market performance. These data serve as the foundation for statistical analysis, dependence testing, and time series modeling of market returns.

The statistical summary presented in Table 5 for the period from January 1, 2022, to the analysis date shows a mean daily closing price of 4724.764, with a high standard deviation of 723.475, indicating significant market volatility. The index ranged from a minimum of 3577.03 to a maximum of 6144.15, reflecting substantial price fluctuations likely driven by economic or geopolitical events. A kurtosis of 1.15716 suggests a distribution with thinner tails than a normal distribution, implying fewer extreme events, while a positive skewness of 0.43032 indicates a slight tendency toward higher-than-average prices, suggesting a potential bullish trend. These metrics from Table 5 highlight a dynamic market environment, valuable for risk assessment and time series modeling, though further analysis of returns or external factors could enhance understanding.

**Table 5.** Statistical summary of SP 500 index (GSPC)

Mean	Standard Deviation	Minimum	Maximum	Kurtosis	Skewness
4724.764	723.475	3577.03	6144.15	1.15716	0.43032

The log-return series of the S&P 500 index (GSPC), calculated using the formula  $R_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$ . The visualization in Figures 4, the time series plot of daily log-returns exhibits significant volatility, with extreme returns such as 0.0909 and -0.0616 indicates rapid market responses to external shocks, such as macroeconomic events or changes in investor sentiment. Table 3 reports a high standard deviation (723.475) and positive skewness (0.43032), which confirms a volatile market with a slight tendency toward higher values. The plot also reveals volatility clustering, where large returns, occur in succession, suggesting time-varying volatility that requires a sophisticated modeling approach.

Given the observed volatility clustering, asymmetry in returns (with larger negative returns like -0.0616), and potential heavy-tailed behavior indicated by extreme returns despite a kurtosis of 1.15716 in Table 3, we propose an ARMA-GARCH model with an asymmetric GARCH component (e.g., GJR-GARCH or EGARCH) and a Student's t-distribution for residuals. The ARMA component captures short-term autocorrelations in the mean, addressing temporal dependencies, while the GARCH component models the time-varying volatility and clustering effects. The asymmetric GARCH variant accounts for the stronger impact of negative shocks, and the t-distribution ensures robustness to extreme returns.

To evaluate whether the S&P 500 log-return series  $R_t$  behaves as an independent process, we conducted the  $DC_\phi$  test, employing a range of divergence measures including Kullback-Leibler ( $KL$ ), Pearson ( $PE$ ), Bhattacharyya ( $BS$ ), Total Divergence ( $TD$ ), Cramer ( $CR$ ), Matusita ( $MD$ ), Jeffreys ( $JD$ ), Hellinger ( $HE$ ), Total Variation ( $TV$ ), and Jensen-Shannon ( $JS$ ). The results, detailed in Table 6, consistently produce p-values below 0.05 in all measures, leading to the rejection of the null hypothesis  $H_0$  that  $R_t$  is independent. This indicates the presence of significant temporal dependencies within the return series, justifying the application of a sophisticated time series model to capture these dynamics.

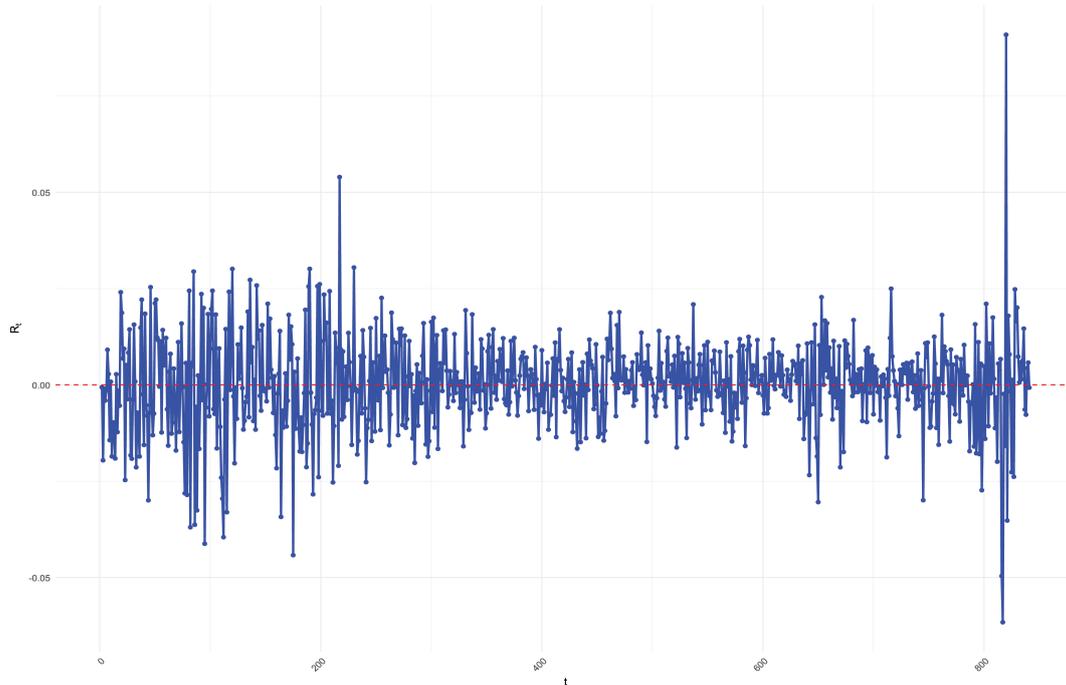
Based on these results, we proceeded to fit an ARMA (0,1)-GJR-GARCH (1,1) model with Student's t distribution for residuals to effectively model the observed dependencies and volatility patterns. The fitted model is expressed as follows:

$$R_t = 5.5461 \times 10^{-4} + 0.014968\epsilon_{t-1} + \sqrt{\hat{h}_t}\hat{\epsilon}_t,$$

where the conditional variance is given by

$$\hat{h}_t = 2.0150 \times 10^{-6} + 0.037055\epsilon_{t-1}^2 + 0.9994 \left( \epsilon_{t-1}^2 \mathbb{I}_{\{\epsilon_{t-1} < 0\}} \right) + 0.9077\hat{h}_{t-1},$$

and  $\hat{\epsilon}_t \sim t(9.674)$ . The Student's t-distribution with 9.674 degrees of freedom accommodates the heavy-tailed behavior, making the model robust for forecasting and risk analysis of the S&P 500 log-returns.



**Figure 5.** Returns of the SP 500 index (GSPC) over time

To investigate the independence of the S&P 500 log-return series  $R_t$  and the standardized residuals  $\hat{\epsilon}_t$  from the fitted ARMA(0,1)-GJR-GARCH(1,1) model, we conducted the  $DC_\phi$  test using multiple divergence measures: Kullback-Leibler ( $KL$ ), Pearson ( $PE$ ), Bhattacharyya ( $BS$ ), Total Divergence ( $TD$ ), Cramer ( $CR$ ), Matusita ( $MD$ ), Jeffreys ( $JD$ ), Hellinger ( $HE$ ), Total Variation ( $TV$ ), and Jensen-Shannon ( $JS$ ). As presented in Table 6, the test results for  $R_t$  show  $DC_\phi$  values ranging from 0.0075 ( $HE$ ) to 0.1851 ( $TV$ ), with all p-values equal to 0.000, leading to the rejection of the null hypothesis  $H_0$  of independence across all measures. This confirms significant temporal dependencies in  $R_t$ , justifying the use of the ARMA(0,1)-GJR-GARCH(1,1) model. In contrast, for the standardized residuals  $\hat{\epsilon}_t$ , the  $DC_\phi$  values are lower (0.0087 to 0.1515), with p-values ranging from 0.089 to 0.143, all exceeding 0.05, indicating acceptance of  $H_0$ . This suggests that the model successfully captures the dependencies in  $R_t$ , producing residuals that behave as an independent process, thus validating the model’s adequacy for modeling the S&P 500 log-returns.

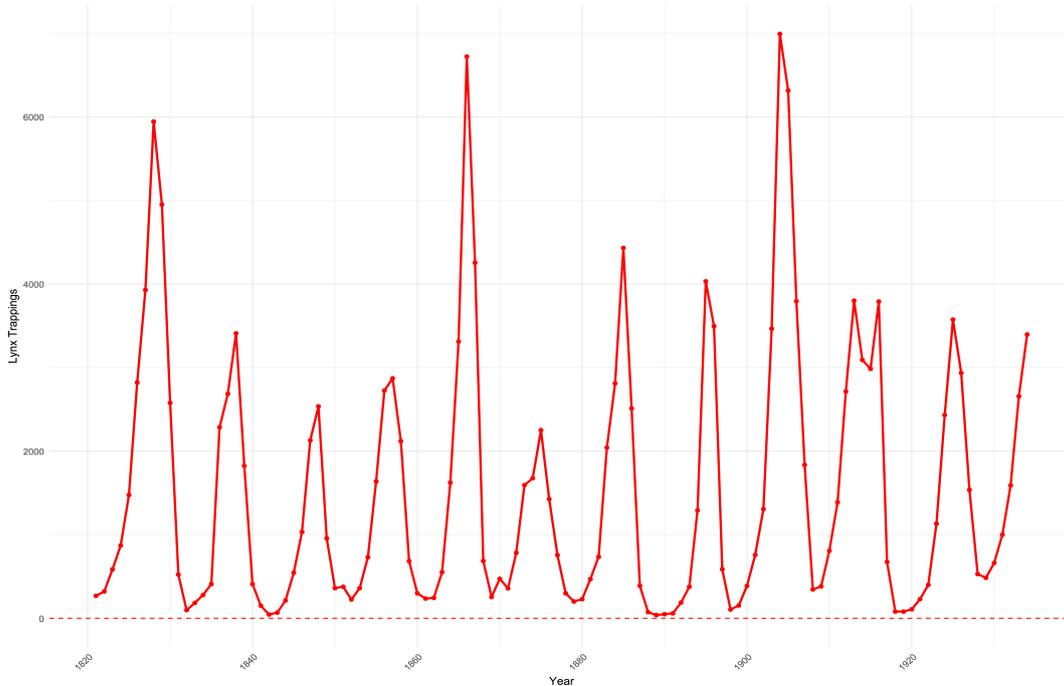
**Table 6.**  $DC_\phi$  Test results for S&P 500 log-returns and residuals

Variable	Test Result	KL	PE	BS	TD	CR	MD	JD	HE	TV	JS
$R_t$	$DC_\phi$	0.0301	0.0312	0.0148	0.295	0.0297	0.0301	0.0601	0.0075	0.1851	0.0149
	P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Decision	<i>REJ</i>									
$\hat{\epsilon}_t$	$DC_\phi$	0.0175	0.0177	0.0087	0.0174	0.0175	0.0175	0.034	0.0438	0.1515	0.0087
	P-value	0.112	0.109	0.089	0.097	0.114	0.119	0.113	0.143	0.125	0.136
	Decision	<i>ACC</i>									

### 5.3. Application III: Lynx population cycles

The `lynx` dataset, available in the base R package, contains annual records of the number of Canadian lynx trapped in the Mackenzie River region of Canada. This dataset,

sourced from historical trapping records, comprises 114 observations spanning from 1821 to 1934. The data show a pronounced seasonal trend with an approximate 10-year cycle, reflecting the cyclical population dynamics of the lynx, likely influenced by ecological factors such as prey availability. This characteristic periodicity makes the `lynx` data set a classic example for time series analysis, particularly for modeling seasonal and cyclical patterns (see Figure 6).



**Figure 6.** Lynx population cycles (1821–1934)

Table 7 presents the descriptive statistics of the *lynx* time series. As shown, the mean value of the series is 1538.018 with a relatively high standard deviation of 1585.844, indicating considerable fluctuations over time. The distribution is positively skewed (1.3497) and leptokurtic (1.4627), suggesting the presence of occasional extreme values and deviations from normality. These features highlight the importance of using robust statistical methods when analyzing such ecological data.

**Table 7.** Statistical Summary of Lynx time series

Mean	Standard Deviation	Minimum	Maximum	Kurtosis	Skewness
1538.018	1585.844	39	6991	1.4627	1.3497

To assess the temporal independence of the `lynx` dataset, we applied divergence-based tests, as summarized in Table 8. For all divergence measures (KL, PE, BS, TD, CR, MD, JD, HE, TV, JS), the P-values were consistently 0.000, leading to the rejection of the null hypothesis of independence for the data ( $lynx_t$ ). This outcome confirms a significant temporal dependence in the series, consistent with the observed cyclical patterns.

**Table 8.**  $DC_\phi$  Test results for Lynx data and residuals

Variable	Test	Result	KL	PE	BS	TD	CR	MD	JD	HE	TV	JS
$lynx_t$	$DC_\phi$		0.3099	0.3817	0.1406	0.2813	0.2893	0.3099	0.6199	0.0755	0.6402	0.1473
	P-value		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Decision		REJ									
$\hat{\epsilon}_t$	$DC_\phi$		0.1247	0.1568	0.0562	0.1125	0.1158	0.1247	0.2494	0.0303	0.3433	0.0590
	P-value		0.169	0.118	0.202	0.202	0.194	0.169	0.169	0.178	0.382	0.187
	Decision		ACC									

Given the pronounced seasonal trend observed in the data (see Figure 6), we fitted a seasonal ARIMA model to capture these cyclical dynamics. The model, characterized by coefficients  $ar1 = 1.3421$ ,  $ar2 = -0.6738$ ,  $ma1 = -0.2027$ ,  $ma2 = -0.2564$ , and a mean of 1544.4039.

Subsequently, we evaluated the residuals ( $\hat{\epsilon}_t$ ) of the fitted model using the same divergence measures. The P values, ranging from 0.118 to 0.382, all exceeded the 0.05 threshold, indicating that there is no reason to reject the null hypothesis for the residuals. This indicates that the seasonal ARIMA model effectively captured the temporal dependencies in the `lynx` dataset, producing residuals that do not exhibit significant dependence. These findings underscore the adequacy of the model in accounting for the cyclical and seasonal patterns inherent in the data.

#### 5.4. Comparative analysis and model validation

Before comparing the tests, it is important to note some domain-specific characteristics observed in the data sets. The financial series exhibited typical stylized facts such as volatility clustering and leverage effects, while the Lynx data set showed a pronounced 10-year cyclicity (see Figure 6). Furthermore, all data sets revealed potential deviations from normality, as suggested by the statistics of skewness and kurtosis (Table 3). These features highlight the necessity of employing robust and sensitive dependence tests capable of detecting nonlinear and non-Gaussian structures commonly present in real-world time series.

To evaluate the performance of the proposed test, we compare it against four popular serial dependence tests: BDS, Runs, Ljung-Box, and  $G(m)$ . Table 9 presents the P-values obtained from these tests for both the original datasets and the residuals of fitted models.

A well-performing test is expected to reject the null hypothesis of independence (i.e. detect dependence) in the original data where serial dependence exists and fail to reject it in model residuals, assuming the model has adequately captured the serial structure.

As shown in Table 9, competing methods exhibit clear deficiencies. The BDS test continues to report highly significant p-values even in residuals (e.g.  $1.20 \times 10^{-30}$ ), suggesting oversensitivity and the inability to adjust for model effects. The Runs test fails to detect dependence in some original datasets (e.g., GSPC), while erroneously indicating dependence in residuals. Ljung-Box and  $G(m)$  tests show inconsistent patterns, with low sensitivity in some original data sets and false positives in residuals.

These results demonstrate that the proposed method outperforms existing approaches in both detecting true serial dependence and validating model adequacy.

**Table 9.** P-values of different serial dependence tests applied to three datasets.

Dataset	Test	Original Data	Residuals
GSPC	BDS	$9.64 \times 10^{-10}$	0.0298
	Runs	0.6861	0.0000
	Ljung-Box	0.5972	0.8696
	G(m)	0.5218	0.6923
TSPI	BDS	0.0000	$1.20 \times 10^{-30}$
	Runs	0.6325	0.00004
	Ljung-Box	0.0000	0.0000
	G(m)	$8.27 \times 10^{-44}$	$1.66 \times 10^{-14}$
Lynx	BDS	$3.08 \times 10^{-07}$	$3.95 \times 10^{-03}$
	Runs	0.0000	0.0000
	Ljung-Box	0.0000	0.2740
	G(m)	$2.38 \times 10^{-18}$	0.3159

## 6. Conclusion

In this paper, we introduce a new class of tests for assessing independence in time series, built on the foundation of Phi-divergence and structured around quantile-based symbolization. The asymptotic distribution of the test statistic was derived to establish its theoretical basis, and a bootstrap version was proposed to address the limitations of the asymptotic approach. We also established the consistency of the test, validating its effectiveness.

Using simulations, we explored this test class by focusing on a specific Phi function and comparing its performance with that of other available tests. Through these simulations, optimal parameter values for the test statistic were identified. In addition, two types of divergence were examined, Rukhin and power divergence, both of which depend on an unknown parameter. Simulation studies investigated the effect of this parameter, ultimately showing that Pearson's divergence is optimal for these criteria.

The simulation results indicated that this test class generally achieved higher size-corrected power than competing methods, with specific Phi-divergence cases achieving the best results in Cressie and Read[12], total variation (TV), Balakrishnan and Sanghvi[5], TV divergences for sample sizes 50, 350, 600, and 1000, respectively. The choice of divergence measure significantly affects test performance. Pearson (optimal in both the Rukhin and the power divergence classes) and total variation divergences consistently outperform others in our simulations due to their mathematical properties: Pearson's quadratic form improves sensitivity to symbol frequency deviations, while the linearity of the total variation ensures robustness to non-normality and outliers. This suggests that the proposed method benefits from divergences that balance sensitivity and stability, particularly under discretization via quantile symbolization. However, the optimal measure may vary with dependency structures, inviting further study of measure selection criteria. Overall, The proposed test demonstrates robust performance across various scenarios, including short time series, where it often outperforms competing methods such as the BDS and Ljung-Box tests. This advantage is particularly evident in simulations with sample sizes as small as 50. Additionally, the test is distribution-free, meaning that it does not assume normality or any specific distribution for the data, as validated by its application to diverse distributions (e.g., chi-square, uniform, and truncated normal) in our simulations. However, its computational complexity increases with larger dimensions (e.g., larger  $m$  or  $d$ ), which may pose challenges for very large datasets. Despite this, the bootstrap

approach mitigates small-sample issues, ensuring reliability even in non-ideal conditions. Future work could explore adaptive parameter selection to further optimize performance for specific data characteristics.

Finally, we demonstrate the application of the proposed tests using stock price change data from the Tehran Stock Exchange. These tests successfully detected the dependence of the data and, after fitting a suitable model, they confirmed the models adequacy and validated the independence and identical distribution of residuals, underscoring the practical applicability of this method.

Beyond the current framework, several promising extensions warrant further investigation. First, adapting the test to handle multivariate and high-dimensional time series, potentially through tensor-based symbolization, would extend its applicability to domains such as climate science and genomics, particularly in spatial and spatiotemporal contexts. Second, automated parameter optimization using machine learning techniques (e.g. meta-learning to select the parameter  $\nu$  in PD/RU divergences) could replace manual tuning without compromising statistical power. Third, integrating the test within machine learning pipelines as a module to detect dependency structures, especially in autoregressive or reinforcement learning models, could significantly improve model diagnostics. Finally, addressing computational scalability through GPU acceleration or distributed computing would enable efficient processing of large-scale datasets (e.g.,  $n > 10^5$ ), making the method suitable for big data applications.

## Acknowledgements

I sincerely thank the editor, reviewers and the editorial team of the journal for their valuable and insightful feedback. Their constructive comments have greatly improved the quality and clarity of this paper.

**Author contributions.** Emad Ashtari Nezhad conceptualized the study, developed the methodology, conducted simulations, analyzed the data, and wrote the manuscript.

**Conflict of interest statement.** The author declares no conflict of interest.

**Funding.** No funding was received for conducting this study.

**Data availability.** The data and software codes will be provided upon request to the author.

## References

- [1] J.M. Amigó, S. Zambrano, and M.A. Sanjuán, *True and false forbidden patterns in deterministic and random dynamics*, EPL, **79**, 50001, 2007.
- [2] S. Anjali, N.R. Kaur, and B. Ghosh, *GSSX method: A golden-section-based symbolic representation for time series analysis*, Adv. Data Sci., **17**, 2025.
- [3] E. Ashtari Nezhad, Y. Waghei, G.R. Mohtashami Borzadaran, H.R. Nilli Sani, and H. Alizadeh Noughabi, *The modified permutation entropy-based independence test of time series*, Commun. Stat. Simul. Comput., **48**(10), 28772897, 2019.
- [4] E. Ashtari Nezhad, Y. Waghei, G.R. Mohtashami Borzadaran, H.R. Nilli Sani, and H. Alizadeh Noughabi, *A Serial Independence Test by Kullback-Leibler via Quantile Symbolization*, Accepted May 2024, REVSTAT Stat. J., 2024.
- [5] V. Balakrishnan and L.D. Sanghvi, *Distance between populations on the basis of attribute*, Biometrics, **24**, 859865, 1968.
- [6] F. Betken, M. Li, and P. Zhao, *Ordinal patterns for change point detection in time series*, Comput. Stat. Data Anal., **129**, 2025.

- [7] W.A. Broock, J. A. Scheinkman, W.D. Dechert, and B. LeBaron, *A test for independence based on the correlation dimension*, *Econom. Rev.*, **15**, 197235, 1996.
- [8] J.S. Cánovas and A. Guillamón, *Permutations and time series analysis*, *Chaos*, **19**, 043103, 2009.
- [9] J.S. Cánovas, A. Guillamón, and S. Vera, *Testing for independence: Permutation-based tests vs. BDS test*, *Eur. Phys. J. Spec. Top.*, **222**, 275284, 2013.
- [10] A. Combettes, *Symbolic representations for time series: Advances in pattern detection*, *Comput. Stat.*, **39**(2), 2024.
- [11] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, 2006.
- [12] N. Cressie and T.R.C. Read, *Multinomial goodness-of-fit tests*, *J. R. Stat. Soc. Ser. B*, **46**, 440464, 1984.
- [13] H. Dehling, *Limit theorems for dependent U-statistics*, In *Dependence in Probability and Statistics*, Springer, 6586, 2006.
- [14] J.J. Dik and M.C.M. de Gunst, *The distribution of general quadratic forms in normal variables*, *Stat. Neerl.*, **39**, 1426, 1985.
- [15] A. Dionisio, R. Menezes, and D. A. Mendes, *Mutual information: A measure of dependency for nonlinear time series*, *Physica A*, **344**, 326329, 2004.
- [16] H.J. Elsinger, *Independence tests based on symbolic dynamics*, Oesterreichische Nationalbank, No. 165, 2010.
- [17] E.F. Fama, L. Fisher, M.C. Jensen, and R. Roll, *The adjustment of stock prices to new information*, *Int. Econ. Rev.*, **10**, 121, 1969.
- [18] K. Ghoudi, R.J. Kulperger, and B. Remillard, *A nonparametric test of serial independence for time series and residuals*, *J. Multivar. Anal.*, **79**, 191218, 2001.
- [19] A. Gut, *Probability: A Graduate Course*, Springer, 2006.
- [20] J.D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.
- [21] A. Hassani, M.J. Mollah, and J.S. Lee, *White noise misapplications in time series modeling: Implications for model diagnostics*, *Econom. Rev.*, **32**, 2025.
- [22] J. He, X. Liu, and S. Zheng, *Non-parametric Symbolic Approximate Representation (NSAR) for time series classification in smart manufacturing*, *Ind. Eng. J.*, **45**(2), 2024.
- [23] Y. Hong and H. White, *Asymptotic distribution theory for nonparametric entropy measures of serial dependence*, *Econometrica*, **73**, 837901, 2005.
- [24] R. Hounyo and Y. Lin, *Wild bootstrap inference with multiway clustering and serially correlated time effects*, *Stat. Inference J.*, **39**(1), 2024.
- [25] J. Jiang, L. Gao, and Z. Shao, *Distance covariance for object-valued time series in metric spaces*, *J. Time Ser. Anal.*, **44**(3), 2023.
- [26] G.M. Ljung and G.E.P. Box, *On a measure of lack of fit in time series models*, *Biometrika*, **65**, 297303, 1978.
- [27] Z. Liu, X. Zhang, and P. Chen, *Kernel-based joint independence tests for multivariate time series*, *J. Multivar. Anal.*, **153**, 2023.
- [28] F. López, M. Matilla-García, J. Mur, and M.R. Marín, *A non-parametric spatial independence test using symbolic entropy*, *Reg. Sci. Urban Econ.*, **40**, 106115, 2010.
- [29] M. Matilla-García and M.R. Marín, *A non-parametric independence test using permutation entropy*, *J. Econom.*, **144**, 139155, 2008.
- [30] M. Mohammadi, D. Li, and J. Kim, *Model-free prediction approach for time series using nonparametric methods*, *J. Forecast.*, **44**, 2024.
- [31] L. Pardo, *Statistical Inference Based on Divergence Measures*, Chapman and Hall/CRC, Taylor & Francis Group, 2006.
- [32] Y. Polyanskiy and Y. Wu, *Information Theory: From Coding to Learning*, Cambridge University Press, 2023.
- [33] J.S. Racine and E. Maasoumi, *A versatile and robust metric entropy test of time-reversibility, and other hypotheses*, *J. Econom.*, **138**, 547567, 2007.

- [34] P.M. Robinson, *Consistent nonparametric entropy-based testing*, Rev. Econ. Stud., **58**, 437453, 1991.
- [35] A.L. Rukhin, *Optimal estimator for the mixture parameter by the method of moments and information affinity*, In Trans., 12th Prague Conference on Information Theory, 214219, 1994.
- [36] H.J. Skaug and D. Tjøstheim, *A nonparametric test of serial independence based on the empirical distribution function*, Biometrika, **80**, 591602, 1993.
- [37] A. Wald and J. Wolfowitz, *On a test whether two samples are from the same population*, Ann. Math. Stat., **11**, 147162, 1940.
- [38] J. Wang, H. Xie, and L. Zhang, *Foundation model with series-symbol data generation for sparse time series analysis*, J. Comput. Stat., **39**, 2025.
- [39] J. Wang, L. Li, and Y. Chen, *Detecting state correlations in heterogeneous time series using advanced methods*, J. Time Ser. Econom., **16**, 2024.
- [40] L. Weiß and M. Schnurr, *Generalized ordinal patterns for discrete-valued time series*, Chaos, **34**, 053109, 2024.
- [41] J. Yu, S. Li, and H. Zhou, *Semiparametric latent ANOVA model for event-related potentials analysis in time series*, J. Neurosci. Methods, **223**, 2024.
- [42] Y. Zhou and P. Müller, *Testing independence between random objects in metric spaces using profile association measures*, Ann. Stat., **53**, 2025.

## APPENDIX

### Empirical size and power

Table 10. Empirical size and power across various tests for  $n = 50$

	Model	<i>KL</i>	<i>PE</i>	<i>BS</i>	<i>TD</i>	<i>CR</i>	<i>MD</i>	<i>JD</i>	<i>HE</i>	<i>TV</i>	<i>JS</i>	<i>BDS</i>	<i>LB</i>	<i>G(m)</i>	<i>Run</i>
Empirical size	<i>M1</i>	0.06	0.08	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.27	0.07	0.01	0.05
	<i>M2</i>	0.08	0.08	0.07	0.07	0.07	0.08	0.08	0.08	0.10	0.07	0.18	0.03	0.02	0.07
	<i>M3</i>	0.03	0.03	0.01	0.01	0.01	0.03	0.03	0.03	0.04	0.01	0.22	0.05	0.01	0.05
	<i>M4</i>	0.13	0.14	0.13	0.13	0.13	0.13	0.13	0.14	0.16	0.14	0.33	0.05	0.04	0.09
	<i>M5</i>	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.07	0.26	0.04	0.02	0.09
	<i>M6</i>	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.08	0.10	0.30	0.07	0.03	0.03
Empirical power	<i>M7</i>	0.34	0.36	0.34	0.34	0.35	0.34	0.34	0.34	0.27	0.35	0.50	0.77	0.07	0.13
	<i>M8</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.82	0.04	1.00	0.20
	<i>M9</i>	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.11	0.16	0.19	0.08	0.01	0.11
	<i>M10</i>	0.24	0.23	0.23	0.23	0.23	0.24	0.24	0.24	0.18	0.24	0.52	0.30	0.02	0.12
	<i>M11</i>	0.04	0.04	0.06	0.06	0.07	0.04	0.04	0.05	0.10	0.07	0.46	0.17	0.01	0.05
	<i>M12</i>	0.15	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.14	0.15	0.59	0.19	0.02	0.06
	<i>M13</i>	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.93	0.99	1.00	1.00	0.42	0.99
	<i>M14</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	<i>M15</i>	0.09	0.10	0.08	0.08	0.08	0.09	0.09	0.09	0.08	0.08	0.29	0.09	0.01	0.08

**Table 11.** Empirical size and power across various tests for  $n = 350$

	Model	<i>KL</i>	<i>PE</i>	<i>BS</i>	<i>TD</i>	<i>CR</i>	<i>MD</i>	<i>JD</i>	<i>HE</i>	<i>TV</i>	<i>JS</i>	<i>BDS</i>	<i>LB</i>	<i>G(m)</i>	<i>Run</i>
Empirical size	<i>M1</i>	0.15	0.14	0.14	0.14	0.15	0.15	0.15	0.15	0.13	0.15	0.09	0.05	0.06	0.09
	<i>M2</i>	0.05	0.05	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.06	0.07	0.03	0.02	0.03
	<i>M3</i>	0.08	0.08	0.09	0.09	0.09	0.08	0.08	0.08	0.10	0.08	0.05	0.01	0.03	0.02
	<i>M4</i>	0.09	0.10	0.08	0.08	0.08	0.09	0.09	0.08	0.09	0.08	0.13	0.03	0.07	0.05
	<i>M5</i>	0.09	0.08	0.07	0.07	0.07	0.09	0.09	0.07	0.06	0.07	0.15	0.03	0.02	0.01
	<i>M6</i>	0.09	0.10	0.08	0.08	0.08	0.09	0.09	0.09	0.08	0.09	0.06	0.05	0.06	0.04
Empirical power	<i>M7</i>	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.65	1.00	0.97	0.27
	<i>M8</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.06	1.00	0.17
	<i>M9</i>	0.62	0.62	0.63	0.63	0.63	0.62	0.62	0.63	0.61	0.63	0.36	0.07	0.37	0.11
	<i>M10</i>	0.77	0.76	0.78	0.78	0.78	0.77	0.77	0.78	0.80	0.78	0.43	1.00	0.86	0.15
	<i>M11</i>	0.48	0.44	0.46	0.46	0.46	0.48	0.48	0.47	0.49	0.46	1.00	0.31	0.06	0.08
	<i>M12</i>	0.69	0.64	0.70	0.70	0.69	0.69	0.69	0.69	0.69	0.69	1.00	0.53	0.07	0.10
	<i>M13</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	<i>M14</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	<i>M15</i>	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.11	0.08	0.05	0.12	0.05	0.06

**Table 12.** Empirical size and power across various tests for  $n = 600$

	Model	<i>KL</i>	<i>PE</i>	<i>BS</i>	<i>TD</i>	<i>CR</i>	<i>MD</i>	<i>JD</i>	<i>HE</i>	<i>TV</i>	<i>JS</i>	<i>BDS</i>	<i>LB</i>	<i>G(m)</i>	<i>Run</i>
Empirical size	<i>M1</i>	0.11	0.13	0.11	0.11	0.11	0.11	0.11	0.11	0.12	0.11	0.06	0.05	0.08	0.07
	<i>M2</i>	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.06	0.05	0.06	0.09
	<i>M3</i>	0.07	0.06	0.05	0.05	0.05	0.07	0.07	0.07	0.07	0.06	0.05	0.04	0.05	0.09
	<i>M4</i>	0.07	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.09	0.07	0.03	0.03	0.07	0.06
	<i>M5</i>	0.06	0.08	0.08	0.08	0.08	0.06	0.06	0.06	0.14	0.07	0.07	0.06	0.08	0.07
	<i>M6</i>	0.04	0.07	0.02	0.02	0.02	0.04	0.04	0.03	0.06	0.03	0.08	0.05	0.05	0.04
Empirical power	<i>M7</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.86	1.00	1.00	0.29
	<i>M8</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.05	1.00	0.23
	<i>M9</i>	0.93	0.90	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.73	0.13	0.64	0.09
	<i>M10</i>	0.97	0.97	0.98	0.98	0.98	0.97	0.97	0.97	0.98	0.98	0.66	1.00	0.99	0.08
	<i>M11</i>	0.72	0.66	0.73	0.73	0.73	0.72	0.72	0.72	0.71	0.73	1.00	0.38	0.16	0.07
	<i>M12</i>	0.91	0.87	0.92	0.92	0.92	0.91	0.91	0.91	0.94	0.91	1.00	0.62	0.13	0.06
	<i>M13</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	<i>M14</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	<i>M15</i>	0.17	0.15	0.21	0.21	0.21	0.17	0.17	0.20	0.17	0.21	0.07	0.18	0.08	0.12

**Table 13.** Empirical size and power across various tests for  $n = 1000$

	Model	<i>KL</i>	<i>PE</i>	<i>BS</i>	<i>TD</i>	<i>CR</i>	<i>MD</i>	<i>JD</i>	<i>HE</i>	<i>TV</i>	<i>JS</i>	<i>BDS</i>	<i>LB</i>	<i>G(m)</i>	<i>Run</i>
Empirical size	<i>M1</i>	0.09	0.10	0.09	0.09	0.09	0.09	0.09	0.09	0.10	0.09	0.11	0.08	0.03	0.08
	<i>M2</i>	0.07	0.10	0.08	0.08	0.08	0.07	0.07	0.07	0.06	0.08	0.05	0.03	0.05	0.03
	<i>M3</i>	0.10	0.10	0.11	0.11	0.11	0.10	0.10	0.10	0.07	0.10	0.02	0.03	0.03	0.03
	<i>M4</i>	0.07	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.07	0.01	0.04	0.05	0.06
	<i>M5</i>	0.09	0.06	0.09	0.09	0.09	0.09	0.09	0.09	0.08	0.09	0.05	0.04	0.02	0.07
	<i>M6</i>	0.12	0.11	0.10	0.10	0.10	0.12	0.12	0.11	0.05	0.10	0.04	0.02	0.06	0.05
Empirical power	<i>M7</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	1.00	1.00	0.28
	<i>M8</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.08	1.00	0.17
	<i>M9</i>	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.16	0.89	0.09
	<i>M10</i>	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.73	1.00	1.00	0.12
	<i>M11</i>	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	1.00	0.35	0.29	0.08
	<i>M12</i>	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.65	0.14	0.06
	<i>M13</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	<i>M14</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	<i>M15</i>	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.16	0.14	0.09	0.15	0.01	0.09