REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

A Serial Independence Test by Kullback-Leibler via Quantile Symbolization

Authors: Emad Ashtari Nezhad
– Department of Statistics, University of Birjand, Birjand, Iran
emad_an@birjand.ac.ir; emaad7@gmail.com

Yadollah Waghei ✉
– Department of Statistics, University of Birjand, Birjand, Iran
ywaghei@birjand.ac.ir

G. R. Mohtashami Borzadaran
– Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran
grmohtashami@um.ac.ir; gmb1334@yahoo.com

Hamid Reza Nili Sani
– Department of Statistics, University of Birjand, Birjand, Iran
hnilisani@birjand.ac.ir

Hadi Alizadeh Noughabi
– Department of Statistics, University of Birjand, Birjand, Iran
alizadehhadi@birjand.ac.ir

Abstract:

- In this article, a new consistent, and powerful test is constructed for serial independence by using quantile symbolization. The test statistic is formulated based on the Kullback-Leibler divergence, and its asymptotic distribution under the null hypothesis is proven. Moreover, the consistency of the proposed test is discussed. A simulation study compares the proposed tests efficiency (size-corrected power) with alternative tests (such as the BDS test). Ultimately, an application to the real-world data example is presented.

Keywords:

- *Independence test, Quantile symbolization, Size-corrected power, Kullback-Leibler divergence, Time series.*

AMS Subject Classification:

- 37M10, 62G10, 94A15

✉ Corresponding author

1. INTRODUCTION

The concept of dependence is the ability to predict any random variable by observing the other variables. In time series analysis, non-zero values of the autocorrelation coefficient are considered as dependence. The independence test of time series, consisting of the autocorrelation function, was initially introduced in Von Neumann (1941) [39]. Durbin and Watson (1950) [18] introduced a hypothesis test for the autocorrelation of errors in a regression model. In statistical literature, various schemes with different properties have been used for testing independence. For instance, in classic tests, the test is done using either the correlation or the numerical similarities among the ranks of observations (see Mateus and Caeiro (2013) [30]). The run test is one of the most commonly used classic tests. This test was introduced by Wald and Wolfowitz (1940) [40] and is a method for comparing two populations with n_1 and n_2 sizes.

Divergence measures are criteria for studying the disparities in distribution functions or probability density functions. Besides, these measures can be considered the different criteria between the joint distribution function and the multiplication of marginal distribution functions(Bagnato et al., 2014 [4]). Divergence measures can also be used to construct a test of independence. For instance, a test based on the Kullback-Leibler measure was introduced by Robinson (1991) [34]. The interest in using Kullback-Leibler divergence has been growing in the lost three decades for seeking the dependence presented in time series. For example, Dionisio et al. (2004) [17] implemented a non-parametric test for serial independence based on mutual information. In Dionisio et al. (2004) [17], a Kullback-Leibler distance between the estimated joint distribution and the estimated marginal distributions was exploited. An independent test through measuring the integrated absolute difference was considered in Robinson (1991) [34]. Skaug and Tjøstheim (1993) [37] used the divergence measure of Cramer-von Mises as the basis of the test. Similarly, Ghoudi et al. (2001) [22] proposed test statistics based on the Kolmogorov-Smirnov distance. Furthermore, the BDS test(Broock et al., 1996 [9]) is a significant example of test methodologies using the correlation dimension measure. Symbolic dynamics have been considered significant for scientists in recent years. Tests via symbolic dynamics avoid the estimation of joint and marginal distributions. This allows us to construct a straightforward, consistent, computationally feasible, and powerful test for independence. Moreover, since the tests do not make any restrictive assumptions on the probabilistic distribution, they can provide more general applicability than other tests. Matilla-García and Marín (2008) [31] introduced a test by permutation symbolization via likelihood ratio, known as the $G(m)$ test or permutation entropy-based test. This test was used in Sensoy et al. (2015) [36] to forecast and compare the Islamic market and conventional equity market. Cánovas et al. (2013) [11], by comparing the tests proposed by Amigó et al. (2007) [2], Cánovas and Guillamón (2009) [10], and Matilla-García and Marín (2008) [31], showed that the BDS test possesses superiority over the others. This superiority is observed due to utilizing the numerical differences among observations instead of using their ranks. However, the low convergence rate of the BDS test has led to not preserving the nominal level of significance for the samples with small sizes. This issue has been confirmed by Matilla-García et al. (2014) [32], who showed that the BDS test does not preserve the nominal level even for the sample of 250, although it has more power than their test. Considering the overlap between the m-dimensional vectors generated from the observed time series, the test has been strongly criticized by Elsinger (2010) [19]. Accordingly, Elsinger (2010) [19] showed through the simulation study that the asymptotic distribution of the $G(m)$ test statistic deviated from the chi-square distribution remarkably. For the solution overlap problem, Ashtari Nezhad et al. (2018) [3] deduced that the asymptotic distribution of the test statistic $G(m)$ was weighted chi-square. They explored alternative methods, including overlap-control and bootstrap techniques, applicable to the overlap issue. Their conclusion was that the modified permutation entropy-based test not only exhibited higher accuracy but also substantially enhanced the test's power.

The asymptotic distribution of the weighted chi-square sum depends on the covariance matrix. The estimation of this matrix is difficult. Moreover, the independence test using permutation entropy is not as powerful as other tests(such as the BDS test). Another symbolization method can be proposed to solve the overlap problem without estimating the covariance matrix and increase the test power. López et al. (2010) [29] discussed the median symbolization and stated that when data is larger/smaller than the median, it is coded into 1/0. To increase the accuracy and power of the test, we extend the median symbolization by the quantile symbolization. The idea of quantile symbolization that is used to define the density of the marginals is very close to the idea of equifrequency classes (see Bagnato and Punzo (2010) [5]). Cánovas et al. (2023) [12] employed permutations combined with an additional symbol based on quantiles. This approach captures some of the quantitative characteristics of a time series. Utilizing this new codification system, they conducted independence tests for time series employing the classical chi-square test. Given the substantial number of symbols and the utilization of overlapping elimination to enhance test accuracy, it is essential to note that conducting this test necessitates a large sample size. In this article, we present an independent test with high power and accuracy simultaneously. The rest of the article is structured as follows. In Section 2, the methodology of constructing a test for serial independence is rewired, and a new test based on Kullback-Leibler divergence is presented. Section 3 discusses the asymptotic distribution and consistency of the proposed test statistic. In Section 4, the special case of the proposed test, named the quantile symbolization, is introduced and analyzed. After introducing competing tests in Section 5, a simulation study is conducted to compare the proposed quantile symbolization test with some competing tests in Section 6. Finally, a real-world data example is considered for illustrative purposes in Section 7. The paper is ended with some conclusions and remarks.

2. A KULLBACK-LEIBLER TEST BASED ON SYMBOLIZATION

In this section, a test based on Kullback-Leibler information is developed for serial independence.

2.1. The preliminaries

The basic idea behind serial dependence is that both the past and future of a time series contain information about unobserved variables that can be used to define the present time. For a stochastic process $\{Z_t, t \in \mathbb{Z}\}$, it is of interest to test the following hypothesis:

$$H_0 : \{Z_t, t \in \mathbb{Z}\}\text{is a sequence of independent and identically distributed random variables.}$$

It is clear that under hypothesis H_0, the stochastic process $\{Z_t, t \in \mathbb{Z}\}$ is stationary (see Cryer and Chan (2008)[13, page 17]). Therefore, the null hypothesis is rejected if either the variables are dependent or non-stationary. In most time series studies, we have only an observation corresponding to one of $\{Z_t, t = 1, 2, \ldots, n\}$. Hence, the random vector $\mathbf{Z}_t(m)$ defined as

$$\mathbf{Z}_t(m) = (Z_t, Z_{t+1}, \ldots, Z_{t+m-1}), \quad t = 1, 2, \ldots, n - m + 1$$

may be considered as the basis of the hypothesis testing. In most of the previous studies, the hypotheses have been designed using such vectors and considering the features of the independent and identically distributed random variables. For instance, bearing in mind that the number of the permutations of each m-dimensional vector of the independent and identically distributed random variables is $\frac{1}{m!}$, Matilla-García and Marín (2008) [31] constructed the null hypothesis, H_0. In our article, the null hypothesis is equivalent to the following relation:

$$P(Z_t \in B_1, Z_{t+1} \in B_2, \ldots, Z_{t+m-1} \in B_m) = P(Z_t \in B_1)P(Z_{t+1} \in B_2)\ldots P(Z_{t+m-1} \in B_m),$$

for $t = 1, 2, \ldots, n - m + 1$ and all $B_1, B_2, \ldots, B_m \in \mathcal{B}$ in where $\mathcal{B}$ denotes the Borel σ-algebra on $\mathbb{R}$.

The most important issue of this procedure is to estimate $P(Z_t \in B_1, Z_{t+1} \in B_2, \ldots, Z_{t+m-1} \in B_m)$. The more accurate prediction of this probability can obtain the more accurate test. For example, if the Z_t's that are larger (smaller) than the median (Me) are coded to 1 (2), the initial phase is to apply the following change of variable to Z_t:

$$\delta_t = \begin{cases} 1 & Z_t < Me \\ 2 & Z_t \geq Me \end{cases}.$$

Thereafter, sample space for $(\delta_t, \delta_{t+1}, \delta_{t+3-1})$ is $G = \{c_1 = (1, 1, 1), c_2 = (1, 2, 1), c_3 = (1, 2, 2), c_4 = (1, 1, 2), c_5 = (2, 1, 1), c_6 = (2, 2, 1), c_7 = (2, 1, 2), c_8 = (2, 2, 2)\}$. For independent and identically distributed continuous random variables under this symbolization, we have

$$P((\delta_t, \delta_{t+1}, \delta_{t+3-1}) = c_i) = (P(Z_t \leq Me))^3 = (\frac{1}{2})^3, \quad t = 1, 2, 3, \ldots, n - 3 + 1, i = 1, 2, \ldots, 8.$$

Therefore, comparing the probability of c_i's to $(\frac{1}{2})^3$, it is possible to discern whether the null hypothesis should be rejected or not. It should be noted that the test may raise some problems in practice. For example, Z_t may include ties or even be discrete. If the samples are drawn from continuous distributions, tied observations (i.e., observations of equal value) theoretically occur with a probability of zero. However, this assumption is not realistic because, in practice, it is possible to obtain two exactly identical values in the sample. Therefore, ties will sometimes occur, and how they are handled can impact the results of the specific test being employed. For instance, in the case of a continuous variable Z_t, we have

$P(Z_t = Me) = 0$. For this purpose, in our test, the value of δ_t is 2 when $Z_t = Me$. In general, let R be the number of symbols, i.e. $G = \{c_1, c_2, \dots, c_R\}$ as a set of symbols and the symbolization $f : \mathbb{R}^m \longrightarrow G$ as a function by which a special symbol of the set G is appointed to any borders like $\mathbf{Z}_t(m)$. For any symbol c_i, the random variable $W_{c_i,t}$ is defined according to the following transformation:

$$W_{c_i,t} = \begin{cases} 1 & f(\mathbf{Z}_t(m)) = c_i; \\ 0 & f(\mathbf{Z}_t(m)) \neq c_i, \end{cases} \quad i = 1, 2, \dots, R, \quad t = 1, 2, \dots, K,$$

where $K = n - m + 1$. Accordingly, when Z_t's are identically distributed, $W_{c_i,t}$ follows the Bernoulli distribution with success probability of p_{c_i}, so that $\sum_{i=1}^{R} p_{c_i} = 1$. On the other hand, the estimation of p_{c_i} results in $\widehat{p}_{c_i} = \frac{1}{K}\sum_{t=1}^{K} W_{c_i,t}$.

Remark 2.1. The vectors $\mathbf{z}_t(m)$ represent our observations for estimating p_{c_i}. However, due to the overlap in $\mathbf{z}_t(m)$, these vectors are not independent of each other, even under the null hypothesis. Nevertheless, there are potential solutions to this issue. For instance, implementing control over the overlap is a viable option, albeit with certain limitations. This becomes particularly relevant when facing data scarcity as a common constraint, especially in economics, where the number of observations (sample size) is relatively limited compared to the number of symbols (refer to Ashtari Nezhad et al. (2018) [3], Section 5.1). Ashtari Nezhad et al. (2018) [3] demonstrated that the appropriate solution to address this issue involves utilizing asymptotic results for m-dependent processes. The approach for achieving this will be clarified in the subsequent section.

2.2. The proposed test statistic

The relative entropy was introduced by Kullback and Leibler (1951) [26] to draw a comparison between two probability measure. In special case, the difference between two probability measures $\mathbf{P}$ and $\mathbf{P}^0$ is:

$$D(\mathbf{P}, \mathbf{P}^0) = \sum_{i=1}^{R} p_i \log \frac{p_i}{p_i^0},$$

where $\sum_{i=1}^{R} p_i = \sum_{i=1}^{R} p_i^0 = 1$. Now, if the probability vector $\mathbf{P} = (p_{c_1}, p_{c_2}, ..., p_{c_R})^T$ includes the occurrence probabilities of any of the symbols $c_1, c_2, ..., c_R$, the hypothesis H_0 can be tested using $D(\widehat{\mathbf{P}}, \mathbf{P}^0)$, by which the estimation of $\mathbf{P}$ and its value under H_0 i.e. $\mathbf{P}^0 = (p_{c_1}^0, p_{c_2}^0, ..., p_{c_R}^0)^T$ could be compared. Accordingly, the statistic $M_n(\widehat{\mathbf{P}}, \mathbf{P}^0)$ can be presented as follows:

$$M_n(\widehat{\mathbf{P}}, \mathbf{P}^0) = 2KD(\widehat{\mathbf{P}}, \mathbf{P}^0).$$

It can be deduced that the more disparity between $\widehat{\mathbf{P}}$ and $\mathbf{P}^0$ leads to more amount of $M_n(\widehat{\mathbf{P}}, \mathbf{P}^0)$. Hence, the null hypothesis, H_0, is rejected for large amounts of $M_n(\widehat{\mathbf{P}}, \mathbf{P}^0)$.

3. THE TEST PROPERTIES

In this section, $M_n(\widehat{\mathbf{P}}, \mathbf{P}^0)$ is compared with other test statistics in the family of ϕ-divergence Pardo (2006) [33]. The asymptotic distribution of the test statistic is derived. The test statistic's consistency is proven by presenting the asymptotic distribution.

3.1. Efficiency

By choosing a new measure for calculating the disparity between the probabilities of occurrence of symbols in general and under the null hypothesis, H_0, we introduced a new test statistic. Accordingly, a wide class of tests via the ϕ-divergence under the symbolization was presented by the test statistic (see Pardo (2006) [33, page 113]):

$$M_n^{\phi}(\widehat{\mathbf{P}}, \mathbf{P}^0) = \frac{2K}{\phi''(1)} D_{\phi}(\widehat{\mathbf{P}}, \mathbf{P}^0),$$

where $D_\phi(\widehat{\mathbf{P}}, \mathbf{P}^0) = \sum_{i=1}^R p_{c_i}^0 \phi(\frac{\hat{p}_{c_i}}{p_{c_i}^0})$ and $\phi : [0, \infty) \longrightarrow (-\infty, \infty]$ is a convex function such that $\phi(1) = 0$ and $\phi''(1) > 0$. It is apparent that the more disparity between $\hat{p}_{c_i}$ and $p_{c_i}^0$ can lead to larger $M_n^\phi(\widehat{\mathbf{P}}, \mathbf{P}^0)$. Therefore, H_0 is rejected when $M_n^\phi(\widehat{\mathbf{P}}, \mathbf{P}^0)$ is large enough. For $\phi(x) = x\ln(x)-x+1$, $M_n^\phi(\widehat{\mathbf{P}}, \mathbf{P}^0)$ reduce to $M_n(\widehat{\mathbf{P}}, \mathbf{P}^0)$. In this section, we consider the test statistic M_n^ϕ for testing H_0, and its outcome based on $\{z_t, t = 1, 2, \ldots, n\}$ given by a_n. Let $F_{M_n^\phi}(a_n)$ be the distribution function of M_n^ϕ under the null hypothesis and

$$\triangle_R^+ = \Big\{\mathbf{P} = (p_{c_1}, p_{c_2}, \ldots, p_{c_R})^T : p_{c_i} > 0, i = 1, 2, \ldots, R, \sum_{i=1}^R p_{c_i} = 1\Big\}.$$

For $\mathbf{P} \neq \mathbf{P}^0$ and $\mathbf{P} \in \triangle_R^+$, the sequence $\{M_n^\phi, n \in \mathbb{N}\}$ has exact Bahadur slope(Bahadur, 1971 [7]) $c_\phi(\mathbf{P})$, if with probability one

$$\lim_{n\longrightarrow\infty} log\Big(1 - F_{M_n^\phi}(a_n)\Big) = -\frac{1}{2}c_\phi(\mathbf{P}).$$

Theorem 3.1. *Suppose ϕ is continuously differentiable and*

$$\phi(0) + \lim_{r\longrightarrow\infty} \frac{\phi(r)}{r} < \infty.$$

Then the exact Bahadur slope of M_n^ϕ is given by

$$c_\phi(\mathbf{P}) = \inf_{\nu\in B_\phi} 2D(\nu, \mathbf{P}^0),$$

where $\mathbf{P} \neq \mathbf{P}^0$ and

$$B_\phi = \Big\{\nu : \nu \in \triangle_R^+, D_\phi(\nu, \mathbf{P}^0) \geq D_\phi(\widehat{\mathbf{P}}, \mathbf{P}^0)\Big\}.$$

Proof: The proof of the theorem can be found in A.1. □

Remark 3.1. If $\phi(x) = x\ln(x) - x + 1$ then

$$c_{Kull}(\mathbf{P}) = 2D(\mathbf{P}, \mathbf{P}^0), \quad \mathbf{P} \in \triangle_R^+; \quad \mathbf{P} \neq \mathbf{P}^0.$$

Therefore,

$$c_\phi(\mathbf{P}) = \inf_{\nu\in B_\phi} 2D(\nu, \mathbf{P}^0) \leq 2D(\mathbf{P}, \mathbf{P}^0) = c_{Kull}(\mathbf{P}).$$

Remark 3.1 shows that the Kullback-Leibler test obtained for $\phi(x) = x\ln(x)-x+1$ has maximal Bahadur efficiency among all the ϕ-divergence test statistics.

3.2. The asymptotic distribution

Under the null hypothesis, the stochastic process $\{f(\mathbf{Z}_t(m)), t \in \mathbb{Z}\}$ is $(m-1)$-dependent (see Ashtari Nezhad et al. (2018) [3]). Using the central limit theorem for m-dependent variables, Elsinger (2010) [19, Theorem 1] has shown that

$$\sqrt{K}(\widehat{\mathbf{P}} - \mathbf{P}) \xrightarrow[K\longrightarrow\infty]{d} N(0, \mathbf{\Sigma}),$$

where $\widehat{\mathbf{P}} = (\frac{1}{K}\sum_{t=1}^K W_{c_1,t}, \frac{1}{K}\sum_{t=1}^K W_{c_2,t}, \ldots, \frac{1}{K}\sum_{t=1}^K W_{c_R,t})$ and if $q_{i,j}^l = P(W_{c_i,t+l} = 1|W_{c_j,t} = 1)$, $\mathbf{Q}^{(l)} = \{q_{i,j}^l\}_{i,j=1,2,\ldots,R}$ is the matrix of transform probability for the time distance of l, and when $1 \leq l \leq m$,

$$\mathbf{\Sigma} = diag(\mathbf{P}) - (2m+1)\mathbf{P}\mathbf{P}^T + diag(\mathbf{P})\sum_{l=1}^m \mathbf{Q}^{(l)} + \sum_{l=1}^m {\mathbf{Q}^{(l)}}^T diag(\mathbf{P}). \tag{3.1}$$

The following theorem gives the asymptotic distribution of the test statistic.

Theorem 3.2. *Suppose m is a fixed number, $\mathbf{A} = diag(\mathbf{P}^0)$ is diagonal matrix with $\mathbf{P}^0$ in the diagonal and $\lambda_i, i = 1, 2, ..., r$ are the eigenvalue of the matrix $\mathbf{A}\boldsymbol{\Sigma}^0$. If $\{Z_1, Z_2, ..., Z_n\}$ are independent and identically distributed, then $M_n(\widehat{\mathbf{P}}, \mathbf{P}^0)$ has an asymptotic distribution that can be represented by $\sum_{i=1}^{r} \lambda_i Y_i^2$, where the random variables $Y_1, Y_2, ..., Y_r$ have a standard normal distribution.*

Proof: The proof of the theorem can be found in A.2. □

Remark 3.2. (Soon, 1996 [38], Theorem 1.1). The summation of the dependent Bernoulli variables possesses the Binomial distribution when the dependence among the variables is insignificant, and the probability of occurrence for each variable is low. In our study, the former and the latter conditions might be satisfied, respectively, by controlling the overlaps among the variables and increasing the symbols. Under these conditions, the vector $K\widehat{\mathbf{P}}$ approximately possesses the multinomial distribution and

$$\boldsymbol{\Sigma} = diag(\mathbf{P}^0) - \mathbf{P}^0{\mathbf{P}^0}^T, \tag{3.2}$$

when the variables $\{Z_t; t = 1, 2, ..., n\}$ are independent.

Given equation (3.2) and based on the assumptions of Remark 3.2, the next theorem is deduced. The conditions of Remark 3.2 are held in most of the symbolizations, particularly in the quantile symbolization, which is introduced in the next section. Turning these conditions, the asymptotic distribution of the test statistic is obtained in the below theorem via Ferguson (1996) [20, Lemma 3, page 57].

Theorem 3.3. *If the variables $\{Z_t, t = 1, 2, ..., n\}$ are independent and identically distributed and R is the number of symbols, then under non-overlapping of $\mathbf{Z}_t(m)$s, $M_n(\widehat{\mathbf{P}}, \mathbf{P}^0)$ is asymptotically χ^2_{R-1} distributed, where χ^2_{R-1} denotes the chi square distribution with $R-1$ degree of freedom.*

Proof: The proof of the theorem can be found in A.3. □

Theorem 3.3 is proven in PROOFS Section.

3.3. Consistency

One of the most important characteristics of a test is consistency. The consistency of a test indicates that the test power is equal to 1 asymptotically. In other previous independence tests, such as Robinson (1991) [34], Hong and White (2005) [24], and Matilla-García and Marín (2008) [31], the consistency has been derived for stationary and $(m-1)$-dependent variables. This assumption does not seem to be unrealistic because the dependence until exact order is assumed in all models, and after those orders, the observations are independent. The rest of this section includes $(m-1)$-dependent variables and studying the consistency of the proposed test. Meanwhile, before presenting this issue, we note that the alternative hypothesis (variables are $(m-1)$-dependent) could be considered as follows:

$$H_1 : \mathbf{P} \neq \mathbf{P}^0.$$

Therefore, the value of $\mathbf{P}$ might be considered equal to $\mathbf{P}^*$ under the alternative hypothesis H_1. Bearing in mind this issue, the following theorem is presented.

Theorem 3.4. *Let $\{Z_t; t = 1, 2, ..., n\}$ to be $(m-1)$-dependent, then*

$$\sqrt{K}(D(\widehat{\mathbf{P}}, \mathbf{P}^0) - D(\mathbf{P}^*, \mathbf{P}^0)) \xrightarrow[n\longrightarrow\infty]{d} N(0, \sigma^2(\mathbf{P}^*)),$$

where $\sigma^2(\mathbf{P}^*) = \mathbf{S}^T\boldsymbol{\Sigma}\mathbf{S}$, $\mathbf{S} = \big(log(\frac{p^*_{c_1}}{p^0_{c_1}}), ..., log(\frac{p^*_{c_R}}{p^0_{c_R}})\big)^T$ and $\boldsymbol{\Sigma}$ is defined in (3.1) replacing $2(m-1)$ and $\mathbf{P}^*$ instead of m and $\mathbf{P}$, respectively.

Proof: The proof is given in A.4. □

Theorem 3.5. *Suppose that the stochastic process* $\{Z_t; t = 1, 2, ..., n\}$ *are* $(m-1)$*-dependent, then it can be shown as follows:*

$$\lim_{n \longrightarrow \infty} P(M_n(\widehat{P}, P^0) > c) = 1, \quad \forall c \in \mathbb{R}^+.$$

Proof: The proof of the theorem is given in A.5. □

4. QUANTILE SYMBOLIZATION

In previous sections, we have provided a general setting by means of symbolic maps. In this section, quantile symbolization is introduced as a special case of general symbolization.

4.1. Quantile symbolizing method

Suppose that $Q_{\frac{1}{d}}$ is the $\frac{1}{d}$-quantile of the distribution of the variables $\{Z_t, t = 1, 2, 3, .., n\}$. Then, B_i's may be a member of the following set:

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

Therefore, the null hypothesis, H_0, can be represented by

$$H_0 : P(Z_t \in B_1, Z_{t+1} \in B_2, ..., Z_{t+m-1} \in B_m) = (\frac{1}{d})^m \quad , B_i \in \vartheta, i = 1, 2, ..., m.$$

Now, we need to apply the transformation:

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

Afterward, using the δ_t, the symbolizing f forms the vectors $(\delta_t, \delta_{t+1}, \ldots, \delta_{t+m-1})^T$. Thus, the set $G = \{c_1, c_2, \ldots, c_R\}$ includes $R = d^m$ members, for which the probability of occurrence of the individual i is indicated by p_{c_i}. Now, applying the amounts of vectores $\mathbf{P}^0 = ((\frac{1}{d})^m, (\frac{1}{d})^m, ..., (\frac{1}{d})^m)$ and $\widehat{\mathbf{P}} = (\widehat{p}_{c_1}, \widehat{p}_{c_2}, ..., \widehat{p}_{c_R})$, the statistic $M_n(\widehat{\mathbf{P}}, \mathbf{P}^0)$ can be represented as

$$QS(m, d) = 2K \sum_{i=1}^{d^m} \widehat{p}_{c_i} log(d^m \widehat{p}_{c_i}).$$

Based on the result of Theorem 3.2, one can discern if the null hypothesis should be rejected or not.

4.2. Determination of d and m

Although the parameters d and m are free, when $n \longrightarrow \infty$, the sample size in real applications is fixed. So, the parameters need to be evaluated according to the sample size. The power of the proposed test can be improved by a suitable choice of d (see Bagnato et al. (2010) [6]). In the previous studies, such parameters were calculated via simulation studies. For instance, it is concluded by Agresti(1996) [1] for the chi-square test that the average amount of observations in each category needs to be at least 5. Applying this strategy, Matilla-García and Marín (2008) [31] evaluated the optimal amount of m. Turning to the BDS test, Kanzler (1999) [25] calculated the required parameters (m and ϵ) for different sample sizes through a simulation study. Accordingly, we conducted a simulation study to evaluate the optimal value for d and m. For this purpose, the performance of the test is illustrated under various sample sizes. There separate nominal values ($\alpha = 0.01$, 0.05, and 0.1) are considered for each sample size, and the level of significance of the test are calculated based on 2000 iterations for each scenario. If the empirical sizes for the proposed test is close to the nominal values, it indicates the accuracy of the test. It can be deduced that when the empirical sizes are more than the nominal values, it could be deduced that the test does not possess appropriate accuracy. This issue leads the power of the test to climb wrongly. Similar situations seem to be observed in some other alternative tests, particularly the BDS test Matilla-García et al. (2014) [32]. Contrastingly, if the empirical size is almost equal to the nominal level (either equal or lower than the nominal level), it exhibits that the empirical size is acceptable. Table 1 reveals the simulation result for various amounts of m and d ($d^m \leq n$) based on data generated from independent normal distribution. According to this table, it can be observed that for the sample size larger than 500, the amounts of d and m should satisfy the relation $5d^m \leq n$, which had been previously suggested in the Chi-square independence test and also the other tests based the Chi-square statistic (Matilla-García and Marín, 2008 [31]). Also, for the sample size of fewer than 500, the amounts of m and d are suggested according to Table 2.

5. OTHER ALTERNATIVE TESTS

This section includes alternative tests. The reason for using these tests is the different characteristics of these tests. The run test has high power for time series with linear dependence. The G(m) test for time series with non-linear dependence and moderate sample size has high power. The BDS test for time series with non-linear dependence and large sample size has high power. LjungBox test is a Portmanteau test that can complete the comparison of tests.

Run Test: The run test was initially introduced in Wald and Wolfowitz [40] for comparing two populations with sample sizes n_1 and n_2. However, this test has been used for testing the independence of a time series, in which the number of runs is introduced as a discrete random variable like U. To evaluate the number of runs, various different methods have been represented. For example, let $\{z_i, i = 1, ..., 7\}$ be observations of a time series. The following vector is reached by comparing the observations to the median such that we indicate the observations lower and larger than the median by respective amounts of A and B: (A, A, B, A, B, B, A). Accordingly, the number of runs (the number of made in situations from A/B to the other.) is equal to 5 ($u = 5$). Wald and Wolfowitz [40] showed that under the null hypothesis, if μ_u and σ_u are the mean and the standard deviation of U respectively, it can be observed that

$$\frac{U - \mu_u}{\sigma_u} \xrightarrow[n \longrightarrow \infty]{d} N(0, 1).$$

Thus, when $|\frac{U-\mu_u}{\sigma_u}| > z_{1-\alpha/2}$, the hypothesis H_0 is rejected. A comprehensive review can be found in Mateus and Caeiro (2013) [30].

BDS Test: Broock et al. (1996) [9] studied an independence test called BDS for time series, which is consisted in a correlation integral. The correlation integral used in BDS is reached from the following equation:

$$C_m(\epsilon) = P(||\mathbf{Z}_1 - \mathbf{Z}_2|| \leq \epsilon),$$

where the m-dimensional vectors $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are identically distributed.

In BDS test, the estimation of $C_m(\epsilon)$ is a U-statistic which is introduced by Dehling (2006) [14] and satisfies the equation

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

$m=5$			$m=4$			$m=3$			$m=2$			n	d
0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01		
0.030	0.012	0.000	0.023	0.013	0.000	0.011	0.009	0.001	0.008	0.003	0.000	50	2
–	–	–	–	–	–	0.099	0.045	0.015	0.048	0.025	0.008	50	3
–	–	–	–	–	–	0.032	0.017	0.002	0.100	0.055	0.020	50	4
–	–	–	–	–	–	–	–	–	0.158	0.097	0.041	50	5
0.155	0.105	0.043	0.067	0.050	0.016	0.043	0.021	0.009	0.015	0.006	0.000	80	2
–	–	–	–	–	–	0.073	0.0456	0.009	0.014	0.005	0.001	80	3
–	–	–	–	–	–	0.142	0.067	0.012	0.019	0.007	0.001	80	4
–	–	–	–	–	–	–	–	–	0.025	0.01	0.002	80	5
0.122	0.076	0.033	0.075	0.047	0.019	0.039	0.022	0.006	0.013	0.008	0.002	120	2
–	–	–	0.225	0.128	0.041	0.052	0.031	0.011	0.015	0.005	0.000	120	3
–	–	–	–	–	–	0.131	0.076	0.017	0.013	0.005	0.001	120	4
–	–	–	–	–	–	–	–	–	0.019	0.004	0.000	120	5
0.123	0.076	0.027	0.072	0.043	0.015	0.041	0.019	0.007	0.007	0.004	0.000	200	2
0.377	0.245	0.064	0.185	0.120	0.050	0.048	0.028	0.008	0.011	0.007	0.000	200	3
–	–	–	–	–	–	0.100	0.061	0.019	0.008	0.003	0.001	200	4
–	–	–	–	–	–	0.220	0.126	0.039	0.012	0.006	0.000	200	5
0.101	0.064	0.027	0.068	0.044	0.017	0.032	0.021	0.005	0.017	0.008	0.001	250	2
0.49	0.359	0.157	0.162	0.099	0.034	0.046	0.026	0.006	0.006	0.004	0.000	250	3
–	–	–	0.461	0.304	0.096	0.069	0.041	0.009	0.007	0.003	0.001	250	4
–	–	–	–	–	–	0.207	0.128	0.029	0.012	0.003	0.000	250	5
0.099	0.071	0.035	0.066	0.040	0.014	0.032	0.015	0.006	0.008	0.004	0.000	300	2
0.487	0.354	0.174	0.141	0.093	0.032	0.048	0.025	0.008	0.009	0.004	0.001	300	3
–	–	–	0.483	0.353	0.155	0.054	0.028	0.011	0.011	0.005	0.000	300	4
–	–	–	–	–	–	0.177	0.107	0.028	0.004	0.002	0.000	300	5
0.092	0.055	0.021	0.068	0.043	0.013	0.026	0.013	0.003	0.011	0.005	0.000	400	2
0.488	0.376	0.182	0.109	0.070	0.031	0.043	0.021	0.004	0.011	0.004	0.000	400	3
–	–	–	0.503	0.374	0.165	0.056	0.029	0.008	0.009	0.006	0.001	400	4
–	–	–	0.320	0.173	0.025	0.123	0.073	0.017	0.010	0.003	0.000	400	5
0.088	0.052	0.022	0.062	0.037	0.012	0.044	0.027	0.008	0.009	0.005	0.000	500	2
0.426	0.309	0.146	0.088	0.052	0.019	0.040	0.018	0.005	0.009	0.004	0.000	500	3
–	–	–	0.428	0.306	0.137	0.044	0.023	0.007	0.007	0.003	0.000	500	4
–	–	–	0.689	0.517	0.219	0.089	0.054	0.013	0.006	0.003	0.000	500	5
0.091	0.059	0.023	0.068	0.036	0.012	0.034	0.019	0.008	0.012	0.007	0.001	700	2
0.331	0.232	0.113	0.082	0.047	0.017	0.036	0.021	0.006	0.011	0.007	0.002	700	3
–	–	–	0.299	0.208	0.088	0.039	0.022	0.007	0.005	0.002	0.000	700	4
–	–	–	0.859	0.745	0.469	0.073	0.040	0.007	0.007	0.002	0.000	700	5
0.089	0.059	0.027	0.060	0.039	0.014	0.033	0.019	0.005	0.016	0.007	0.002	900	2
0.234	0.175	0.079	0.073	0.041	0.016	0.039	0.018	0.006	0.011	0.004	0.000	900	3
–	–	–	0.232	0.147	0.056	0.047	0.024	0.004	0.011	0.004	0.000	900	4
–	–	–	0.831	0.732	0.491	0.061	0.033	0.009	0.008	0.004	0.000	900	5
0.085	0.058	0.028	0.062	0.035	0.014	0.038	0.022	0.004	0.013	0.005	0.000	1200	2
0.169	0.114	0.056	0.075	0.043	0.013	0.044	0.019	0.004	0.013	0.004	0.002	1200	3
0.974	0.936	0.801	0.155	0.100	0.036	0.037	0.022	0.006	0.007	0.003	0.000	1200	4
–	–	–	0.761	0.638	0.382	0.044	0.021	0.006	0.008	0.003	0.000	1200	5
–	–	–	–	–	–	0.078	0.042	0.012	0.008	0.002	0.000	1200	6
0.091	0.059	0.022	0.066	0.041	0.016	0.033	0.016	0.007	0.013	0.004	0.000	1500	2
0.167	0.106	0.039	0.076	0.044	0.013	0.045	0.026	0.004	0.009	0.003	0.000	1500	3
0.954	0.912	0.779	0.126	0.076	0.029	0.041	0.021	0.005	0.008	0.005	0.000	1500	4
–	–	–	0.609	0.488	0.284	0.047	0.024	0.004	0.009	0.004	0.000	1500	5
–	–	–	0.991	0.973	0.892	0.067	0.034	0.007	0.011	0.005	0.000	1500	6
–	–	–	–	–	–	0.140	0.085	0.022	0.008	0.002	0.000	1500	7
0.079	0.051	0.019	0.064	0.038	0.013	0.039	0.022	0.008	0.011	0.006	0.000	2000	2
0.135	0.082	0.031	0.059	0.033	0.008	0.034	0.017	0.003	0.004	0.004	0.000	2000	3
0.874	0.800	0.617	0.109	0.068	0.016	0.039	0.022	0.006	0.004	0.002	0.000	2000	4
0.968	0.909	0.656	0.434	0.331	0.165	0.045	0.022	0.005	0.008	0.003	0.00	2000	5
–	–	–	0.984	0.962	0.867	0.052	0.024	0.004	0.007	0.003	0.000	2000	6
–	–	–	–	–	–	0.099	0.053	0.012	0.008	0.002	0.000	2000	7

Table 1: Empirical size of QS test

Brocck et al. (1996) [9] asserted that under the hypothesis H_0, $C_m(\epsilon)$ is equal to $(C_1(\epsilon))^m$. Accordingly, if $\sigma_{n,m}$ is the standard deviation of $C_{m,n}(\epsilon) - (C_1(\epsilon))^m$, the test statistic of BDS test is evaluated by

$d=5$	$d=4$	$d=3$	$d=2$	n
$3d^m \leq n$	$2d^m \leq n$	$2d^m \leq n$	$3d^m \leq n$	$n \leq 50$
$4d^m \leq n$	$3d^m \leq n$	$3d^m \leq n$	$3d^m \leq n$	$50 < n \leq 500$

Table 2: Suitable values m and d

means of the following equation:

$$\sqrt{n}\frac{C_{m,n}(\epsilon)-(C_1(\epsilon))^m}{\sigma_{n,m}}.$$

Considering the U statistics proven in Denker and Keller (1983) [15], it was shown by Broock et al. (1996) [9] that the test statistic asymptotically possesses the standard normal distribution. Therefore, when

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

the null hypothesis, H_0, is rejected.

G(m) Test: Matilla-García and Marín (2008) [31] introduced the independence test consisting in the permutation entropy, in which if $\Gamma=\{\pi_1,\pi_2,...,\pi_{m!}\}$ is the set of all permutation symbols, the vectors $\mathbf{Z}_t(m)$ receive one of the permutations Γ under the relation "$\leq$". In other words, the vectors $\mathbf{Z}_m(t)$ achieve one of the symbols in the set Γ, according to the sequence (bigger or smaller) of its elements. Consequently, by ordering elements of $\mathbf{Z}_m(t)$, if the relation $Z_{t+i_1} \leq Z_{t+i_2} \leq \ldots \leq Z_{t+i_m}$ is held, the symbol of the vector $\mathbf{Z}_m(t)$ will be $(i_1, i_2, \ldots, i_m)$. Consequently, the hypothesis H_0 may be represented as follows:

$$P(Z_{t+i_1} \leq Z_{t+i_2} \leq \ldots \leq Z_{t+i_m}) = \frac{1}{m!}, \quad \forall(i_1,i_2,\ldots,i_m) \in \Gamma, \quad \forall t \in \mathbb{Z}.$$

Matilla-García and Marín (2008) [31] has defined the test statisct as:

$$G(m) = -2K\left[\widehat{h}(m) - log(m!)\right],$$

where $\widehat{h}(m) = -\sum_{i=1}^{m!}\widehat{p}_{\pi_i}log(\widehat{p}_{\pi_i})$ is the reputation entropy given in Bandt and Pompe (2002) [8]. Matilla-García and Marín (2008) [31] showed that the statistic $G(m)$ owns the chi-square distribution with $m!-1$ degrees of freedom under the null hypothesis. Thus, the null hypothesis, H_0, is rejected when the test statistic satisfies

$$G(m) > \chi^2_{m!-1,1-\alpha}.$$

Ljung-Box Test (LB): Ljung and Box (1978) [28] introduced the independence test based on m autocorrelations of the data. The Ljung-Box test statistic is:

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

Where n is the sample size, $\hat{\rho}_k^2$ denotes the sample autocorrelation at lag k, and m is the number of lags being tested. Under H_0, the statistic $LB(m)$ asymptotically follows a χ^2_m. Thus, the null hypothesis, H_0, is rejected when the test statistic satisfies

$$LB(m) > \chi^2_{m,1-\alpha}.$$

6. SIMULATION STUDY

Any decent comparison between different tests should include two topics which are the test accuracy (the ability to preserve the nominal level) and power simultaneously. Thus, the simulation study is conducted to compare the performance of the proposed test with other competing tests in terms of power and accuracy. Furthermore, the tests are compared by size-corrected powers. To collate finite sample performance of the tests, the sample size was chosen to be 50, 150, 250, and 500 in each scenario, representing small, moderate, fairly large, and large sample sizes, respectively. To illustrate the BDS test, the amounts of $m=3,6$ and $\epsilon=2\sigma$ were selected as the model parameters for all sample sizes. To

Data type	Model's name	Model
Independent data	$DGP1:$	$Z_t = \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N(0,1)$
	$DGP2:$	$Z_t = \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \chi^2(3)$
	$DGP3:$	$Z_t = \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} t(2)$
	$DGP4:$	$Z_t = \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} U(0,1)$
	$DGP5:$	$Z_t = \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} Beta(0.5,0.5)$
	$DGP6:$	$Z_t = \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} TN(0,1,-1.75,1.75)$
Dependent data	$DGP7\big(AR(2)\big):$	$Z_t = 0.3Z_{t-1} + 0.4Z_{t-2} + \varepsilon_t$
	$DGP8\big(ARMA(1,2)\big):$	$Z_t = 0.09Z_{t-1} + 0.1\varepsilon_{t-1} + 0.5\varepsilon_{t-2} + \varepsilon_t$
	$DGP9\big(Bilinear\big):$	$Z_t = 0.5\varepsilon_{t-2} + 0.4\varepsilon_{t-1}Z_{t-2} + \varepsilon_t$
	$DGP10\big(SignAR\big):$	$Z_t = 0.3sign(Z_{t-2}) + \varepsilon_t$
	$DGP11\big(Logistic\big):$	$Z_t = 4Z_{t-1}(1-Z_{t-1}), \quad Z_t \sim U(0,1)$
	$DGP12\big(Non-linearMA\big):$	$Z_t = 0.8\varepsilon_{t-2}^2 + \varepsilon_t$
	$DGP13:$	$Z_t = 0.8\lvert Z_{t-2}\rvert^{0.5} + 0.6\varepsilon_t$
	$DGP14\big(ARCH(1)\big):$	$Z_t = \sqrt{h_t}\varepsilon_t, \quad h_t = (1 + 0.8Z_{t-1}^2)$
	$DGP15\big(GARCH(1,1)\big):$	$Z_t = \sqrt{h_t}\varepsilon_t, \quad h_t = (1 + 0.6Z_{t-1}^2 + 0.3h_{t-1}^2)$
	$DGP16\big(\text{Random walk}\big):$	$Z_t = Z_{t-1} + \varepsilon_t$
	$DGP17:$	$Z_t = 0.01 + 0.01t + \varepsilon_t$
	$DGP18:$	$Z_t = 0.01 + t^{0.1} + \varepsilon_t$

Table 3: Simulation models

$G(m)$ and $LB(m)$, the parameter m was picked to be 3 for the small sample size and 4 for the moderate and large sample sizes. According to the Table 2, the amounts of $(3,3)$, $(4,2)$ and $(3,4)$ were selected as the parameters (m,d) for the small sample size. Also, the respective $(5,2)$ and $(4,3)$ were chosen as the parameters (m,d) for sample sizes 150 and 250.

Table 3 represents the data generated via $DGP1$ to $DGP6$ models, which were independent and identically distributed from a standard normal, chi-square distribution with 3 degree of freedom, t distribution with 2 degree of freedom, uniform distribution defining over $(0,1)$, truncated normal distribution defining over $[-1.75, 1.75]$, respectively. The proportion of the number of rejections of the null hypothesis out of 2000 simulated data from $DGP1$ to $DGP6$ was considered as empirical size. Similarly, the dependent data indicated by $DGP7$ to $DGP15$, with the trends named $DGP16$ to $DGP18$, were generated. The percentage of rejecting the null hypothesis out of 2000 iterations was considered the test power. Tables 4, 5, 6 and 7 reveal the simulation results. Furthermore, an efficiency measure for each test is presented by size-corrected power

$$EFF_n = \overline{\beta_n} - 2\overline{\alpha_n},$$

For each test and n, $\overline{\beta_n}$ is the mean of empirical powers, and $\overline{\alpha_n}$ is the mean of empirical sizes.

Tables 4 illustrate that the BDS test does not preserve the nominal level based on the sample size 50. However, this deviation decreased with the increase in the sample sizes, achieving the nominal level for the sample size of 500. Also, the proposed test does not preserve the nominal level for the sample size 250, and the $QS(4,3)$ test deviated enormously from the nominal level. Accordingly, the BDS and $QS(4,3)$ tests do not possess the necessary accuracy at the significance level 0.05 for the sample sizes $n = 50, 150, 250$.

In $DGP7$-$DGP10$ and $DGP13$ that are mostly linear or very similar to linear models, it seems that simple tests, like runs and the Ljung-Box test, are better for making the diagnosis of the dependency. Given these results and omitting the BDS test that does not preserve the nominal level of significance, the runs test, $QS(4,2)$ and Ljung-Box demonstrate superiority for the small sample size. Turning this sample size, the test $QS(4,2)$ exhibits superior power for the models $DGP9$, $DGP10$, and $DGP13$, which are non-linear. The power of run tests has declined by growing in the sample size. Contrastingly, the tests $QS(4,2)$ and $QS(5,2)$ owned the most tremendous powers for the moderate and large sample sizes. Generally, it can be claimed that based on these results, $QS(4,2)$ possessed the strongest performance.

	Model	Run test	$BDS(3,2\sigma)$	$BDS(6,2\sigma)$	$QS(3,3)$	$QS(3,4)$	$QS(4,2)$	$G(3)$	$LB(3)$
Empirical size	$DGP1$	0.0660	0.2490	0.2780	0.0505	0.0240	0.0640	0.0255	0.0545
	$DGP2$	0.0575	0.1740	0.2080	0.0495	0.0250	0.0470	0.0225	0.0420
	$DGP3$	0.0620	0.1520	0.1885	0.0480	0.0195	0.0445	0.0255	0.0380
	$DGP4$	0.0640	0.3025	0.3315	0.0475	0.0230	0.0525	0.0340	0.0535
	$DGP5$	0.0595	0.2835	0.3075	0.0380	0.0220	0.0575	0.0250	0.0495
	$DGP6$	0.0705	0.2895	0.3205	0.0465	0.0235	0.0555	0.0265	0.0335
Empirical power	$DGP7$	0.5320	0.6260	0.6060	0.4975	0.3410	0.5345	0.0320	0.3570
	$DGP8$	0.2541	0.3790	0.3965	0.2005	0.1210	0.2545	0.0480	0.5225
	$DGP9$	0.1645	0.3650	0.4070	0.1695	0.0815	0.2390	0.0525	0.3455
	$DGP10$	0.2565	0.5595	0.5705	0.3600	0.1750	0.5570	0.1695	0.6310
	$DGP11$	0.1880	0.8865	0.8210	1.0000	1.0000	0.2750	1.000	0.0525
	$DGP12$	0.0680	0.2030	0.2740	0.0790	0.0420	0.0465	0.0370	0.0685
	$DGP13$	0.1555	0.4800	0.4875	0.1340	0.0610	0.2105	0.0680	0.3505
	$DGP14$	0.0610	0.5645	0.4875	0.0685	0.0355	0.0600	0.0305	0.1020
	$DGP15$	0.0580	0.5705	0.5770	0.0770	0.061	0.0565	0.0275	0.1310
	$DGP16$	0.9960	0.9930	0.9750	0.9965	0.9850	0.9885	0.4110	0.8910
	$DGP17$	0.0705	0.2725	0.2905	0.0470	0.0270	0.0805	0.0280	0.0525
	$DGP18$	0.0590	0.2540	0.2735	0.0440	0.0200	0.0605	0.0295	0.0490

Table 4: Empirical size and power for $n = 50$

	Model	Run test	$BDS(3,2\sigma)$	$BDS(6,2\sigma)$	$QS(3,3)$	$QS(3,4)$	$QS(4,2)$	$QS(5,2)$	$G(4)$	$LB(4)$
Empirical size	$DGP1$	0.0585	0.1335	0.1340	0.0235	0.0680	0.0440	0.0810	0.0585	0.0515
	$DGP2$	0.0655	0.0995	0.0900	0.0255	0.0690	0.0440	0.0775	0.0605	0.0485
	$DGP3$	0.0625	0.0805	0.0955	0.0255	0.0670	0.0430	0.0760	0.0575	0.0365
	$DGP4$	0.0555	0.1580	0.1400	0.0225	0.0615	0.0385	0.0760	0.0520	0.054
	$DGP5$	0.0585	0.1320	0.1340	0.0255	0.0740	0.0375	0.0665	0.0710	0.0465
	$DGP6$	0.0570	0.1355	0.1385	0.0275	0.0620	0.0380	0.0775	0.0550	0.0445
Empirical power	$DGP7$	0.9185	0.9590	0.9510	0.9515	0.9535	0.9740	0.9740	0.1700	0.6320
	$DGP8$	0.4815	0.5975	0.6410	0.6175	0.6275	0.7585	0.8220	0.3395	0.8990
	$DGP9$	0.1530	0.6765	0.7045	0.4815	0.5350	0.6680	0.7575	0.4920	0.6590
	$DGP10$	0.2390	0.5395	0.6180	0.9100	0.9080	0.9885	0.9900	0.7915	0.7290
	$DGP11$	0.1705	0.9810	0.9465	1.0000	1.0000	0.2950	0.4040	1.0000	0.0595
	$DGP12$	0.0815	0.2110	0.2920	0.1405	0.3115	0.0645	0.1030	0.1735	0.0860
	$DGP13$	0.1385	0.4000	0.4295	0.3485	0.3770	0.5970	0.6580	0.4350	0.6735
	$DGP14$	0.0530	0.9250	0.8330	0.0870	0.2300	0.0470	0.0855	0.0800	0.2365
	$DGP15$	0.0565	0.9640	0.9525	0.1565	0.3830	0.0360	0.0710	0.0690	0.3575
	$DGP16$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9965	1.0000
	$DGP17$	0.3040	0.1710	0.1850	0.1505	0.1950	0.2505	0.3170	0.0545	0.4105
	$DGP18$	0.0645	0.1275	0.1175	0.0275	0.0570	0.0445	0.0790	0.0510	0.0620

Table 5: Empirical size and power for $n = 150$

It can be deduced that the dependency of the models with complicated dependence ($DGP11$-$DGP12$ and $DGP14$-$DGP18$) could be discerned by increasing the parameter d in the $QS(m,d)$ test. Considering the models $DGP11$ and $DGP12$, the test $QS(3,3)$ for the small sample size, and $QS(3,4)$ for the moderate and large scenarios (excluding the test $QS(4,3)$ due to not preserving the nominal level of significance) exhibit superior power performance. As was expected, considering the previous studies like Matilla-García et al. (2014) [32], the BDS test possesses huge power for the models $ARCH$ and $GARCH$ ($DGP14$ and $DGP15$). The simulation results indicate that the performance of the $QS(3,4)$ test is comparable with that of the BDS test. Furthermore, all the tests have similar power for discerning the trend, particularly for the $DGP17$ model, except the $G(m)$ test.

Table 8 illustrates amounts of EFF_n for tests based on four separate sample sizes. Accordingly, it is revealed in this Table that the proposed test has the most size-corrected power in comparison with other tests.

	Model	Run test	$BDS(3,2\sigma)$	$BDS(6,2\sigma)$	$QS(3,3)$	$QS(3,4)$	$QS(4,2)$	$QS(4,3)$	$QS(5,2)$	$G(4)$	$LB(4)$
Empirical size	$DGP1$	0.0545	0.0970	0.0930	0.0160	0.0520	0.0415	0.1040	0.0700	0.0495	0.0570
	$DGP2$	0.0495	0.0740	0.0755	0.0230	0.0345	0.0435	0.0994	0.0755	0.0450	0.0425
	$DGP3$	0.0435	0.0680	0.0680	0.0205	0.0425	0.0425	0.0915	0.0685	0.0380	0.0500
	$DGP4$	0.0435	0.1045	0.1025	0.0245	0.0420	0.0355	0.0990	0.0640	0.0495	0.0505
	$DGP5$	0.0455	0.0950	0.0970	0.0315	0.0475	0.0375	0.1020	0.0705	0.0555	0.0545
	$DGP6$	0.0470	0.1050	0.0965	0.0245	0.0360	0.0435	0.0925	0.0685	0.0545	0.0515
Empirical power	$DGP7$	0.9890	0.9970	0.9965	0.9990	0.9990	0.9990	1.0000	0.9990	0.2535	1.0000
	$DGP8$	0.6435	0.8030	0.8555	0.9090	0.8665	0.9640	0.9650	0.9745	0.5725	1.0000
	$DGP9$	0.1420	0.8550	0.8865	0.8245	0.7725	0.9340	0.9570	0.9575	0.7615	0.9990
	$DGP10$	0.2485	0.5990	0.7055	0.9970	0.9975	1.0000	0.9990	0.9995	0.9430	1.0000
	$DGP11$	0.1705	0.9960	0.9775	1.0000	1.0000	0.3015	1.0000	0.4460	1.0000	0.0585
	$DGP12$	0.0590	0.3275	0.4180	0.2645	0.4325	0.0625	0.5195	0.0935	0.2640	0.1005
	$DGP13$	0.1320	0.4285	0.4920	0.6395	0.5575	0.8590	0.8290	0.8750	0.6725	0.9870
	$DGP14$	0.0430	0.9930	0.9635	0.1390	0.3140	0.0370	0.2655	0.0680	0.0905	0.2725
	$DGP15$	0.0485	0.9995	0.9965	0.2625	0.5060	0.0385	0.3990	0.0665	0.0730	0.4475
	$DGP16$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$DGP17$	0.9625	0.8200	0.9320	0.9735	0.9550	0.9860	0.9915	0.9935	0.0540	1.0000
	$DGP18$	0.0490	0.1145	0.1130	0.0265	0.0415	0.0330	0.0935	0.0655	0.0410	0.0720

Table 6: Empirical size and power for $n = 250$

	Model	Run test	$BDS(3,2\sigma)$	$BDS(6,2\sigma)$	$QS(3,3)$	$QS(3,4)$	$QS(4,2)$	$QS(4,3)$	$QS(5,2)$	$G(4)$	$LB(4)$
Empirical size	$DGP1$	0.0490	0.0660	0.0615	0.0610	0.0625	0.0245	0.0215	0.0300	0.0495	0.0500
	$DGP2$	0.0600	0.0555	0.0555	0.0560	0.0520	0.0215	0.0255	0.0400	0.0420	0.0385
	$DGP3$	0.0540	0.0515	0.0585	0.0610	0.0605	0.0240	0.0205	0.0335	0.0525	0.049
	$DGP4$	0.0450	0.0650	0.0700	0.0695	0.0705	0.0150	0.0160	0.0305	0.0475	0.0555
	$DGP5$	0.0535	0.0725	0.0765	0.0715	0.0690	0.0185	0.0245	0.0390	0.0415	0.0475
	$DGP6$	0.0605	0.0665	0.0715	0.0740	0.0780	0.0255	0.0195	0.0420	0.0440	0.0515
Empirical power	$DGP7$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5575	1.0000
	$DGP8$	0.8755	0.9755	0.9885	0.9900	0.9870	0.9995	0.9975	1.0000	0.9410	1.0000
	$DGP9$	0.1740	0.9850	0.9925	0.9945	0.9945	0.9975	0.9940	0.9990	0.9910	0.9900
	$DGP10$	0.2600	0.7650	0.8495	0.8910	0.9005	1.0000	1.0000	1.0000	0.9995	1.0000
	$DGP11$	0.1845	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.3005	1.0000	0.0595
	$DGP12$	0.0615	0.5545	0.7050	0.7310	0.7160	0.6985	0.8655	0.0915	0.5470	0.1140
	$DGP13$	0.1445	0.5530	0.6465	0.6715	0.6590	0.9850	0.9500	0.9970	0.9600	0.9931
	$DGP14$	0.0560	1.0000	0.9995	0.9995	0.9990	0.3250	0.6040	0.0425	0.1350	0.3090
	$DGP15$	0.0490	1.0000	1.0000	1.0000	1.0000	0.5550	0.8405	0.0375	0.0885	0.5440
	$DGP16$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$DGP17$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0500	1.0000
	$DGP18$	0.0690	0.0730	0.0665	0.0715	0.0785	0.0210	0.0230	0.0465	0.0515	0.1035

Table 7: Empirical size and power for $n = 500$

7. REAL DATA APPLICATION

The aim of economic units and investors in exchange is to prevent loss. Therefore, the most important issue for investors is forecasting stock price changes. The predictability of the stock price is closely related to the market's efficiency theorem. Efficiency is a basic concept in financial markets that is used for a market; the availability of information does not affect the predictability on which individual supply and demand have an insignificance impact. In other words, efficiency consists of the principle that information adjusts and affects the stock price so quickly. From Fama et al.'s perspective Fama et al. (1969)[21], a market may be called efficient if it is very good at adjusting soon to new information. Accordingly, it can be deduced that efficiency leads the market to the unpredictability of stock prices in the future. Efficiency can be categorized into three different categories, namely weak efficiency, average efficiency, and strong efficiency. Weak efficiency describes a situation in which the information belongs to previous periods, and their impacts have been reflected in the stock and therefore do not affect predictions anymore. Thus, the predictability rejects the hypothesis of weak efficiency (Sensoy et al.,

Test	$n = 50$	$n = 150$	$n = 250$	$n = 500$
Run test	0.112	0.186	0.279	0.299
$BDS(3, 2\sigma)$	0.029	0.383	0.563	0.700
$BDS(6, 2\sigma)$	−0.031	0.395	0.601	0.723
$QS(3, 3)$	0.213	0.439	0.623	0.731
$QS(3, 4)$	0.200	0.414	0.619	0.730
$QS(4, 2)$	0.173	0.395	0.520	0.755
$QS(4, 3)$	−	−	0.556	0.814
$QS(5, 2)$	−	0.370	0.489	0.555
$G(m)$	0.108	0.270	0.380	0.518
$LB(m)$	0.206	0.389	0.559	0.580

Table 8: Size-corrected power of the tests

2015 [36]). The index of Tehran stock price has been collected over the period of 32 months starting at March 19th, 2014. The index of Tehran data, given in the official website of the Tehran Stock[1], includes 639 data. Apparently, the data for the holidays have not been recorded. The R_t index is considered for the dependence of the stock price exchanges (Matilla-García and Marín, 2008 [31]), defined as

$$R_t = ln(\frac{P_t}{P_{t-1}}),$$

where P_t and P_{t-1} are price indexes for the current and the prior periods, respectively. Figure 1 illustrates this time series.

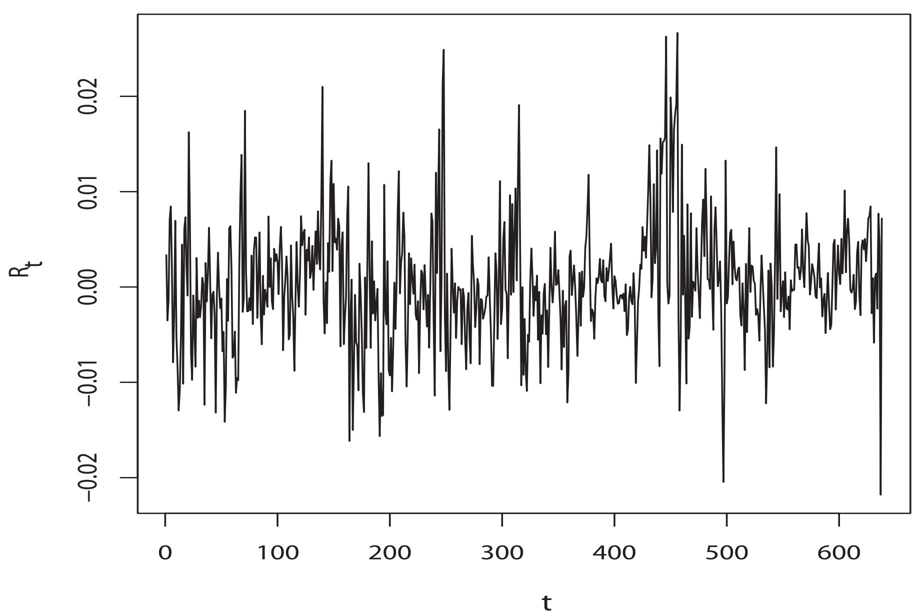

Figure 1: Index of tehran stock price changes

We have applied the $QS(3, 5)$ test for the variable R_t to test the independence. Additionally,

[1] www.irbourse.com

another index defined by

$$IEF(Z_t) = 1 - exp\{-D(\widehat{\mathbf{P}}, \mathbf{P}^0)\},$$

is used for dependence (inefficiency) of $\{Z_t, t = 1, 2, 3, .., n\}$. It can be shown that $D(\widehat{\mathbf{P}}, \mathbf{P}^0) \in \mathbb{R}^+$ and it goes to infinity by climbing in the dependency. Consequently, as the IEF becomes closer to 1, much more dependency (inefficiency) among observations is observed. Table 9 reveals the index $IEF(R_t)$ and the $QS(3, 5)$ test results. According to this table, the $QS(3, 5)$ test leads to rejecting the null hypothesis, H_0, against the alternative (P-value < 0.05). Rejecting the hypothesis H_0, we could fit a decent model for the variable R_t. As aforementioned, the residuals in different time series models ought to be independent and identically distributed. Therefore, dependence tests may be regarded as adequacy tests. Consequently, we applied the model $ARMA(0, 1) - GARCH(1, 1)$ to the data Ruppert and Matteson (2015) [35], employing the $QS(3, 5)$ test as the adequacy test. The fitted model results have been evaluated by means of the following equation:

$$R_t = 2.482 \times 10^{-5} + 0.4698\epsilon_{t-1} + \sqrt{\widehat{h}_t}\hat{\epsilon}_t,$$

$$\hat{h}_t = \sqrt{1.483 \times 10^{-6} + 0.2008\hat{h}_{t-1}\hat{\epsilon}^2_{t-1} + 0.7714\hat{h}_{t-1}}.$$

Moreover, the results of the $QS(3, 5)$ test for the standard residuals are illustrated in Table 9. Given the results for $\hat{\epsilon}_t$, since P-value > 0.05 we conclude that the the null hypothesis is not rejected, i.e. The residuals ($\hat{\epsilon}_t$) are independent and identically distributed.

Variable	$QS(3, 5)$	Degrees of freedom	P-value	IEF
R_t	283.637	124	2.41×10^{-14}	0.199314
$\hat{\epsilon}_t$	120.4158	124	0.5743	0.09019

Table 9: The $QS(3, 5)$ test results for R_t and standardized residuals

8. CONCLUSION

This article developed a new powerful test with high accuracy for time series analysis. First, the test construction was described using the introduced symbolization, and then the test statistic was driven via Kullback and Leibler (1951) [26]. Considering the properties of the Kullback-Leibler measure, a new method was obtained for the optimality of symbolization. The asymptotic distribution of the test statistic under the hypothesis H_0 (independent and identically distributed) was the sum of weighted chi-square variables. It was deduced that the test statistic converges asymptotically to chi-square distribution under the null hypothesis in certain circumstances. Furthermore, under the dependence hypothesis, the asymptotic distribution of the test statistic and then the consistency for the test were driven. We represented the test based on quantile symbolization. A simulation study was carried out to inspect the proposed test's power and accuracy (preserving the nominal level). The proposed test results were compared with four alternative tests. The following statements result from the simulation study. The quantile symbolization-based test possesses a decent accuracy in each sample size scenario. In other words, the proposed test can preserve the nominal level of significance, and it exhibits much better performance than the well-known BDS test. By comparison with four alternative tests, the proposed test exhibits superior power. The proposed test possessed higher power in all cases to discern the simple dependency. To diagnose the complex dependency, such as the $GARCH$ model, the superior power of the test suggested was proven in most scenarios compared to competing tests. It is revealed by Table 8 that the proposed test had the highest size-corrected power in comparison with the other tests. Ultimately, the stock price changes of the Tehran bourses were analyzed through the proposed test. It was observed that there is a dependency between the data. Afterward, having fitted the model $ARMA(0, 1) - GARCH(1, 1)$ to the data, the adequacy test (independent and identically distributed) was applied. So, it was shown through the quantile symbolization-based test that the residuals are independent and identically distributed.

REFERENCES

[1] Agresti, A. (1996). *An introduction to categorical data analysis*, Vol. 135, New York: Wiley.

[2] Amigó, J.M.; Zambrano, S. and Sanjuán, M.A. (2007). True and false forbidden patterns in deterministic and random dynamics, *EPL (Europhysics Letters)*, **79**, 50001.

[3] Ashtari Nezhad, E.; Waghei, Y.; Mohtashami Borzadaran, G.R.; Nilli Sani, H.R. and Alizadeh Noughabi, H. (2018). The Modified Permutation Entropy-Based Independence Test of Time series, *Communications in Statistics-Simulation and Computation*, **48**, 2877–2897.

[4] Bagnato, L.; De Capitani, L. and Punzo, A. (2014). Testing serial independence via density-based measures of divergence, *Methodology and Computing in Applied Probability*, **16**, 627–641.

[5] Bagnato, L. and Punzo, A. (2010). On the use of chi-square test to check serial independence, *Statistica & Applicazioni*, **8**, 57–74.

[6] Bagnato, L.; Punzo, A. and Nicolis, O. (2012). The autodependogram: a graphical device to investigate serial dependences, *Journal of Time Series Analysis*, **33**, 233–254.

[7] Bahadur, R.R. (1971). *Some Limit Theorems in Statistics*, Vol. 4., siam.

[8] Bandt, C. and Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series, *Physical Review Letters*, **88**, 174102.

[9] Broock, W.A.; Scheinkman, J.A.; Dechert, W.D and LeBaron, B. (1996). A test for independence based on the correlation dimension, *Econometric Reviews*, **15**, 197–235.

[10] Cánovas, J.S. and Guillamón, A. (2009). Permutations and time series analysis, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **19**, 043103.

[11] Cánovas, J.S., Guillamón, A. and Vera, S. (2013). Testing for independence: Permutation based tests vs. BDS test, *The European Physical Journal Special Topics*, **222**, 275–284.

[12] Cánovas, J.S.; Guillamón, A. and Ruiz-Abellón, M.C. (2023). A new coding system to test independence of time series, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **33**, 013122.

[13] Cryer, J.D. and Chan, K. (2008). *Time Series Analysis With Applications in R*, New York: Springer.

[14] Dehling, H. (2006). Limit theorems for dependent U-statistics, *In Dependence in probability and statistics*, Springer New York, 65-86.

[15] Denker, M. and Keller, G. (1983). On U-statistics and v. misestatistics for weakly dependent processes, *Zeitschrift fr Wahrscheinlichkeitstheorie und verwandte Gebiete*, **64**, 505–522.

[16] Dik, J.J. and de Gunst, M.C.M. (1985). The distribution of general quadratic forms in normal variables, *Statistica Neerlandica*, **39**, 14–26.

[17] Dionisio, A.; Menezes, R. and Mendes, D.A. (2004). Mutual information: a measure of dependency for nonlinear time series, *Physica A: Statistical Mechanics and its Applications*, **344**, 326–329.

[18] Durbin, J. and Watson, G.S. (1950). Testing for serial correlation in least squares regression, *I. Biometrika*, **37**, 409–428.

[19] Elsinger, H. (2010). Independence tests based on symbolic dynamics (No. 165), *Oesterreichische Nationalbank*.

[20] Ferguson, T.S. (1996). *A course in large sample theory (Vol. 49)*, London: Chapman and Hall.

[21] Fama, E.F.; Fisher, L.; Jensen, M.C. and Roll, R. (1969). The adjustment of stock prices to new information, *International economic review*, **10**, 1–21.

[22] Ghoudi, K.; Kulperger, R.J. and Remillard, B. (2001). A nonparametric test of serial independence for time series and residuals, *Journal of Multivariate Analysis*, **79**, 191–218.

[23] Gut, A. (2006). *Probability: A Graduate Course*, Springer Science and Business Media.

[24] Hong, Y. and White, H. (2005). Asymptotic distribution theory for nonparametric entropy measures of serial dependence, *Econometrica*, **73**, 837–901.

[25] Kanzler, L. (1999). Very fast and correctly sized estimation of the BDS statistic, *Christ Church and Department of Economics*, University of Oxford.

[26] Kullback, S. and Leibler, R.A. (1951). On information and sufficiency, *The Annals of Mathematical Statistics*, **22**, 79–86.

[27] Lehmann, E.L. (1999). *Elements of large-sample theory*, Springer Science and Business Media.

[28] Ljung, G.M. and Box, G.E. (1978). On a measure of lack of fit in time series models, *Biometrika*, **65**, 297–303.

[29] López, F.; Matilla-García, M.; Mur, J. and Marín, M.R. (2010). A nonparametric spatial independence test using symbolic entropy, *Regional Science and Urban Economics*, **40**, 106–115.

[30] Mateus, A. and Caeiro, F. (2013). Comparing several tests of randomness based on the difference of observations, *In AIP Conf. Proc.*, **1558**, 809–812.

[31] Matilla-García, M. and Marín, M.R. (2008). A non-parametric independence test using permutation entropy, *Journal of Econometrics*, **144**, 139–155.

[32] Matilla-García, M.; Marín, M.R.; Dore, M.I. and Ojeda, R.B. (2014). Nonparametric correlation integral–based tests for linear and nonlinear stochastic processes, *Decisions in Economics and Finance*, **37**, 181–193.

[33] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*, Chapman and Hall/CRC, Taylor & Francis Group.

[34] Robinson, P.M. (1991). Consistent nonparametric entropy-based testing, *The Review of Economic Studies*, **58**, 437–453.

[35] Ruppert, D. and Matteson, D.S. (2015). *Statistics and Data Analysis for Financial Engineering with R examples*, New York: Springer.

[36] Sensoy, A.; Aras, G. and Hacihasanoglu, E. (2015). Predictability dynamics of Islamic and conventional equity markets, *The North American Journal of Economics and Finance*, **31**, 222–248.

[37] Skaug, H.J. and Tjøstheim, D. (1993). A nonparametric test of serial independence based on the empirical distribution function, *Biometrika*, **80**, 591–602.

[38] Soon, S.Y.T. (1996). Binomial approximation for dependent indicators, *Statistica Sinica*, **6**, 703–714.

[39] Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance, *The Annals of Mathematical Statistics*, **12**, 367–395.

[40] Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population, *The Annals of Mathematical Statistics*, **11**, 147–16.

A. PROOFS

A.1. Proof of Theorem 3.1.

By substituting M_n^{ϕ} instead of $T_n^{\phi}(\hat{\mathbf{P}}, \mathbf{P}^0)$ in the proof of Theorem 4.1 in Pardo (2006) [33, page 171], the proof is completed.

A.2. Proof of Theorem 3.2.

Suppose R is equal to number of symbols, $\phi(x) = xlog(x) - x + 1$ and $\psi(x_1, x_2, \ldots, x_R) = \sum_{i=1}^{R} p^0_{c_i} \phi(\frac{x_i}{p^0_{c_i}})$. Using Taylor's expansion for the ψ in $(\widehat{p}_{c_1}, \widehat{p}_{c_2}, \ldots, \widehat{p}_{c_R})$ and around $\mathbf{P} = (p_{c_1}, p_{c_2}, \ldots, p_{c_R})^T$, we have

$$\begin{aligned}\psi(\widehat{p}_{c_1}, \widehat{p}_{c_2}, \ldots, \widehat{p}_{c_R}) = {} & \psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R}) \\ & + \sum_{i=1}^{R-1} \frac{\partial \psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R})}{\partial p_{c_i}} (\widehat{p}_{c_i} - p_{c_i}) \\ & + \frac{1}{2} \sum_{i,j=1}^{R-1} \frac{\partial \psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R})}{\partial p_{c_i} \partial p_{c_j}} (\widehat{p}_{c_i} - p_{c_i})(\widehat{p}_{c_j} - p_{c_j}) \\ & + o(O_p(n^{-1})), \end{aligned} \tag{1.1}$$

where c is a constant, and also $\frac{O(n)}{n} \xrightarrow[n \longrightarrow \infty]{} c$, $\frac{o(n)}{n} \xrightarrow[n \longrightarrow \infty]{} 0$, in which $\xrightarrow{\mathcal{P}}$ denotes convergence in Probability. Now, under the hypothesis H_0, the value of $\mathbf{P} = (p_{c_1}, p_{c_2}, \ldots, p_{c_R})^T$ is equal to $\mathbf{P}^0 = (p^0_{c_1}, p^0_{c_2}, \ldots, p^0_{c_R})^T$, and therefore

$$\psi(p^0_{c_1}, p^0_{c_2}, \ldots, p^0_{c_R}) = 0.$$

Moreover, we have:

$$\Big(\frac{\partial \psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R})}{\partial p_{c_i}}\Big)_{P=P^0} = \Big(\phi'(\frac{p_{c_i}}{p^0_{c_i}}) - \phi'(\frac{p_{c_R}}{p^0_{c_R}})\Big)_{P=P^0} = 0, \tag{1.2}$$

$$\Big(\frac{\partial^2 \psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R})}{\partial p_{c_i} \partial p_{c_j}}\Big)_{P=P^0} = \frac{1}{p^0_{c_R}}, \quad i \neq j, \tag{1.3}$$

$$\Big(\frac{\partial^2 \psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R})}{\partial p_{c_i} \partial p_{c_j}}\Big)_{P=P^0} = \frac{1}{p^0_{c_i}} + \frac{1}{p^0_{c_R}} \quad i = j. \tag{1.4}$$

Accordingly, the first and the second part of (1.1) are equal to zero. Given the equations (1.3) and (1.4), the third part of (1.1) is calculated as follows:

$$\begin{aligned} & \frac{1}{2}\Big[\sum_{i,j=1, i\neq j}^{R-1} \frac{\Big(\widehat{p}_{c_i} - p^0_{c_i}\Big)\Big(\widehat{p}_{c_j} - p^0_{c_j}\Big)}{p^0_{c_R}} + \sum_{i=1}^{R-1} \frac{\Big(\widehat{p}_{c_i} - p^0_{c_i}\Big)^2}{p^0_{c_R}} + \sum_{i=1}^{R-1} \frac{\Big(\widehat{p}_{c_i} - p^0_{c_i}\Big)^2}{p^0_{c_i}}\Big] \\ & = \frac{1}{2} \sum_{i=1}^{R} \frac{\Big(\widehat{p}_{c_i} - p^0_{c_i}\Big)^2}{p^0_{c_i}}. \end{aligned} \tag{1.5}$$

Therefore, using the equation (1.5), it may be deduced that

$$M_n(\widehat{\mathbf{P}},\mathbf{P}^0) = K\sum_{i=1}^{R}\frac{\left(\widehat{p}_{c_i}-p^0_{c_i}\right)^2}{p^0_{c_i}} + 2Ko\Big(O_p(n^{-1})\Big). \tag{1.6}$$

Using Slutsky's theorem (Gut, 2006 [23]), the right side of the equation (1.6) possesses the same distribution as $K\sum_{i=1}^{R}\frac{\left(\widehat{p}_{c_i}-p^0_{c_i}\right)^2}{p^0_{c_i}}$. Now considering $\mathbf{P}^0 = (p^0_{c_1}, p^0_{c_2} ..., p^0_{c_R})^T$, the following asymptotic distribution is reached by means of Elsinger (2010) [19, Theorem 1, page 8]:

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

Thus, bearing in mind Dik and Gunst (1985) [16, Corollary 2.1], if $\mathbf{A} = diag(\mathbf{P}^0)$, we have

$$K(\widehat{\mathbf{P}}-\mathbf{P}^0)^T A(\widehat{\mathbf{P}}-\mathbf{P}^0) = K\sum_{i=1}^{R}\frac{\left(\widehat{p}_{c_i}-p^0_{c_i}\right)^2}{p^0_{c_i}} \overset{d}{\equiv} \sum_{i=1}^{r}\lambda_i Z_i^2,$$

and therefore, the proof is completed.

A.3. Proof of Theorem 3.3.

As mentioned in Theorem 3.2, $M_n(\widehat{\mathbf{P}},\mathbf{P}^0)$ and $K(\widehat{\mathbf{P}}-\mathbf{P}^0)^T diag(\mathbf{P}^0)(\widehat{\mathbf{P}}-\mathbf{P}^0)$ possess the same asymptotic distribution. Accordingly, for the conditions of Ferguson (1996) [20, Lemma 3, page 57], we have

$$\begin{aligned} X &= \sqrt{K}diag((\mathbf{P}^0)^{-\frac{1}{2}})(\widehat{\mathbf{P}}-\mathbf{P}^0),\\ \mathbf{\Sigma} &= diag((\mathbf{P}^0)^{-\frac{1}{2}})\big(diag(\mathbf{P}^0)-\mathbf{P}^0(\mathbf{P}^0)^T\big)diag((\mathbf{P}^0)^{-\frac{1}{2}}). \end{aligned}$$

However,

$$\begin{aligned} \mathbf{\Sigma\Sigma} = I &- diag((\mathbf{P}^0)^{-\frac{1}{2}})\mathbf{P}^0(\mathbf{P}^0)^T diag((\mathbf{P}^0)^{-\frac{1}{2}}) - diag((\mathbf{P}^0)^{-\frac{1}{2}})\mathbf{P}^0(\mathbf{P}^0)^T diag((\mathbf{P}^0)^{-\frac{1}{2}})\\ &+ diag((\mathbf{P}^0)^{-\frac{1}{2}})\mathbf{P}^0(\mathbf{P}^0)^T diag((\mathbf{P}^0)^{-\frac{1}{2}})diag((\mathbf{P}^0)^{-\frac{1}{2}})\mathbf{P}^0(\mathbf{P}^0)^T diag((\mathbf{P}^0)^{-\frac{1}{2}}). \end{aligned}$$

On the other hand, considering the fact that

$$(\mathbf{P}^0)^T diag((\mathbf{P}^0)^{-\frac{1}{2}})diag((\mathbf{P}^0)^{-\frac{1}{2}})\mathbf{P}^0 = 1,$$

we have the following equation:

$$\mathbf{\Sigma\Sigma} = I - diag((\mathbf{P}^0)^{-\frac{1}{2}})\mathbf{P}^0(\mathbf{P}^0)^T diag((\mathbf{P}^0)^{-\frac{1}{2}}) = \mathbf{\Sigma}. \tag{1.7}$$

Therefore, $\mathbf{\Sigma}$ is an idempotent matrix and also

$$\begin{aligned} rank(\mathbf{\Sigma}) &= trace\Big(diag((\mathbf{P}^0)^{-\frac{1}{2}})\big(diag(\mathbf{P}^0)-\mathbf{P}^0(\mathbf{P}^0)^T\big)diag((\mathbf{P}^0)^{-\frac{1}{2}})\Big)\\ &= trace\Big(diag((\mathbf{P}^0)^{-1})\big(diag(\mathbf{P}^0)-\mathbf{P}^0(\mathbf{P}^0)^T\big)\Big)\\ &= R(1-\frac{1}{R}) = R-1, \end{aligned}$$

Where R is the number of symbols, $rank(.)$ and $trace(.)$ are the rank and trace of a matrix. Now, considering Ferguson (1996) [20, Lemma 3, page 57], the proof is completed.

A.4. Proof of Theorem 3.4.

Suppose R be the number of symbols and $\phi(x) = xlog(x) - x + 1$. Taylor's expansion of $\psi(x_1, x_2, \ldots, x_R) = \sum_{i=1}^{R} p^0_{c_i}\phi(\frac{x_i}{p^0_{c_i}})$ in $(\widehat{p}_{c_1}, \widehat{p}_{c_2}, \ldots, \widehat{p}_{c_R})$ and around $\mathbf{P} = (p_{c_1}, p_{c_2}, ..., p_{c_R})$ leads to

$$\begin{aligned} \psi(\widehat{p}_{c_1}, \widehat{p}_{c_2}, \ldots, \widehat{p}_{c_R}) &= \psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R}) \\ &+ \sum_{i=1}^{R-1} \frac{\partial\psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R})}{\partial p_{c_i}}(\widehat{p}_{c_i} - p_{c_i}) \\ &+ o(O_p(n^{-\frac{1}{2}})). \end{aligned} \tag{1.8}$$

Now, by partial derivative of $\psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R})$ at the point $\mathbf{P}^*$ we have

$$\begin{aligned} \Big(\frac{\partial\psi(p_{c_1}, p_{c_2}, \ldots, p_{c_R})}{\partial p_{c_i}}\Big)_{P=P^*} &= \Big(\frac{\partial\big(\sum_{i=1}^{R-1} p^0_{c_i}\phi(\frac{p_{c_i}}{p^0_{c_i}}) + p^0_{c_R}\phi(\frac{p_{c_R}}{p^0_{c_R}})\big)}{\partial p_{c_i}}\Big)_{\mathbf{P}=\mathbf{P}^*} \\ &= \phi'(\frac{p^*_{c_i}}{p^0_{c_i}}) - \phi'(\frac{p^*_{c_R}}{p^0_{c_R}}). \end{aligned}$$

Consequently, by replacing $\mathbf{P}$ with $\mathbf{P}^*$ in (1.8), it can be observed that under the alternative hypothesis (H_1) we have

$$\begin{aligned} \sum_{i=1}^{R-1}\big(\phi'(\frac{p^*_{c_i}}{p^0_{c_i}}) - \phi'(\frac{p^*_{c_R}}{p^0_{c_R}})\big)(\widehat{p}_{c_i} - p_{c_i}) &= \sum_{i=1}^{R-1}\phi'(\frac{p^*_{c_i}}{p^0_{c_i}})(\widehat{p}_{c_i} - p^*_{c_i}) - \sum_{i=1}^{R-1}\phi'(\frac{p^*_{c_R}}{p^0_{c_R}})(\widehat{p}_{c_i} - p^*_{c_i}) \\ &= \sum_{i=1}^{R-1}\phi'(\frac{p^*_{c_i}}{p^0_{c_i}})(\widehat{p}_{c_i} - p^*_{c_i}) - \phi'(\frac{p^*_{c_R}}{p^0_{c_R}})(\sum_{i=1}^{R-1}\widehat{p}_{c_i} - \sum_{i=1}^{R-1}p^*_{c_i}) \\ &= \sum_{i=1}^{R-1}\phi'(\frac{p^*_{c_i}}{p^0_{c_i}})(\widehat{p}_{c_i} - p^*_{c_i}) - \phi'(\frac{p^*_{c_R}}{p^0_{c_R}})(1 - \widehat{p}_{c_R} - (1 - p^*_{c_R})) \\ &= \sum_{i=1}^{R}\phi'(\frac{p^*_{c_i}}{p^0_{c_i}})(\widehat{p}_{c_i} - p^*_{c_i}). \end{aligned}$$

Considering equation (1.8), by multiplying $\sqrt{K}$ we have

$$\sqrt{K}\big(D(\widehat{\mathbf{P}}, \mathbf{P}^0) - D(\mathbf{P}^*, \mathbf{P}^0)\big) = \sqrt{K}\sum_{i=1}^{R}\phi'(\frac{p^*_{c_i}}{p^0_{c_i}})(\widehat{p}_{c_i} - p^*_{c_i}) + \sqrt{K}\Big(o(O_p(n^{-\frac{1}{2}}))\Big). \tag{1.9}$$

On the other hand, using Slutsky's theorem(Gut, 2006 [23]) it can be deduced that the equation (1.9) and $\sqrt{K}\sum_{i=1}^{R}\phi'(\frac{p^*_{c_i}}{p^0_{c_i}})(\widehat{p}_{c_i} - p^*_{c_i})$ possess the same asymptotic distribution. Given Elsinger (2010) [19, Theorem 1, page 8] , if we set

$$\mathbf{S} = \big(\phi'(\frac{p^*_{c_1}}{p^0_{c_1}}), \ldots, \phi'(\frac{p^*_{c_R}}{p^0_{c_R}})\big)^T = \big(log(\frac{p^*_{c_1}}{p^0_{c_1}}), \ldots, log(\frac{p^*_{c_R}}{p^0_{c_R}})\big)^T,$$

it can be shown that

$$\sqrt{K}\sum_{i=1}^{R}\phi'(\frac{p^*_{c_i}}{p^0_{c_i}})(\widehat{p}_{c_i} - p^*_{c_i}) = \sqrt{K}\mathbf{S}^T(\widehat{\mathbf{P}} - \mathbf{P}^*) \xrightarrow[K\longrightarrow\infty]{d} N(0, \mathbf{S}^T\mathbf{\Sigma}\mathbf{S}),$$

and therefore, the proof is completed. It is worth mentioning that since $K = n - m + 1$, there is no difference between $K \longrightarrow \infty$ and $n \longrightarrow \infty$ here.

A.5. Proof of Theorem 3.5.

If $\sigma^2(\mathbf{P}^*)$ is the variance given in Theorem 3.4,

$$\begin{aligned} P(M_n(\widehat{\mathbf{P}},\mathbf{P}^0) > c)) &= P(2KD(\widehat{\mathbf{P}},\mathbf{P}^0) > c) \\ &= P\Big(\frac{\sqrt{K}\big(D(\widehat{\mathbf{P}},\mathbf{P}^0) - D(\mathbf{P}^*,\mathbf{P}^0)\big)}{\sigma(\mathbf{P}^*)} > \frac{\sqrt{K}}{\sigma(\mathbf{P}^*)}\big(\frac{c}{2K} - D(\mathbf{P}^*,\mathbf{P}^0)\big)\Big), \end{aligned}$$

for any positive K. Given the theorem assumption, we have

$$\mathbf{P} = \mathbf{P}^* \neq \mathbf{P}^0,$$

and

$$D(\mathbf{P}^*,\mathbf{P}^0) > 0.$$

Accordingly, it is observed that

$$\frac{\sqrt{K}}{\sigma(\mathbf{P}^*)}\big(\frac{c}{2K} - D(\mathbf{P}^*,\mathbf{P}^0)\big) \xrightarrow[n\longrightarrow\infty]{d} -\infty.$$

On the other hand, considering Theorem 3.4, it could be claimed that $\sqrt{K}\big(D(\widehat{\mathbf{P}},\mathbf{P}^0)-D(\mathbf{P}^*,\mathbf{P}^0)\big)$ converges to normal distribution.

Now, using Slutsky's theorem (see Lehmann (1999) [27], page 70.), we have

$$\begin{aligned} \lim_{n\longrightarrow\infty} P(M_n(\widehat{\mathbf{P}},\mathbf{P}^0) > c)) &= 1 - \lim_{n\longrightarrow\infty} P\Big(\frac{\sqrt{K}\big(D(\widehat{\mathbf{P}},\mathbf{P}^0) - D(\mathbf{P}^*,\mathbf{P}^0)\big)}{\sigma(\mathbf{P}^*)} \\ &< \frac{\sqrt{K}}{\sigma(\mathbf{P}^*)}\big(\frac{c}{2K} - D(\mathbf{P}^*,\mathbf{P}^0)\big)\Big) \\ &= 1 - \Phi(-\infty) = 1, \end{aligned}$$

where Φ is the standard normal distribution function, and so the proof is completed.