Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Nonsense Dependence Index: A Rank-Based Diagnostic for Identifying Potentially Spurious Associations

STHITADHI DAS

sthdas999@gmail.com

Brainware University https://orcid.org/0009-0001-2484-9530

Research Article

Keywords: Kendall's τ, Bergsma–Dassios τ∗, Nonsense Dependence Index, Spuri- ous dependence, Trend-induced association, Dependence diagnostics, Nonparametric association measures, Rank-based statistics

Posted Date: June 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10159420/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

The *Nonsense Dependence Index*: A Rank-Based Diagnostic for Identifying Potentially Spurious Associations

Sthitadhi Das
Department of Mathematics
Brainware University, India

Abstract

The detection of statistical dependence constitutes a fundamental problem in modern data analysis. Classical measures such as Pearson's correlation coefficient, Kendall's τ, Spearman's *rho*, and more recent dependence measures including Bergsma–Dassios' τ^* are designed to identify departures from independence. However, in the presence of common trends, temporal effects, latent factors, or aggregation artifacts, these measures may indicate substantial dependence even when the observed association lacks substantive meaning. Such phenomena are commonly referred to as nonsense correlations and, more generally, nonsense dependence. In this article, we introduce a simple diagnostic quantity, termed the *Nonsense Dependence Index (NDI)*, $\text{NDI} = |\tau| - \sqrt{\tau^*}$, where τ denotes Kendall's rank correlation coefficient and τ^* denotes Bergsma–Dassios' τ^* statistic. The proposed index contrasts first-order ordinal association with higher-order dependence structure and is designed to identify situations in which apparent dependence may be driven by external trends rather than an intrinsic stochastic relationship. For genuine monotone relationships, NDI is expected to remain close to zero, whereas elevated positive values provide evidence of potential nonsense dependence. The practical utility of the proposed index is investigated through a real-data study involving well-known examples of spurious association arising from common temporal trends. Comparisons with conventional dependence measures demonstrate that the proposed diagnostic can effectively distinguish meaningful dependence from potentially misleading associations. The methodology offers a computationally simple and interpretable tool for exploratory dependence analysis and may serve as a useful complement to existing nonparametric measures of association.

Keywords: Kendall's τ; Bergsma–Dassios τ^*; Nonsense Dependence Index; Spurious dependence; Trend-induced association; Dependence diagnostics; Nonparametric association measures; Rank-based statistics.
AMS Subject Classification (2020): 62H20, 62G10, 62G20, 62G08, 60F05.

1 Introduction

The measurement of statistical association has occupied a central position in statistics for more than a century. Classical measures such as Pearson's product–moment correlation coefficient, Spearman's rank correlation coefficient, and Kendall's rank correlation coefficient provide quantitative summaries of the strength and direction of relationships between variables. More recently, the development of general dependence measures, including distance covariance, the maximal information coefficient, and the Bergsma–Dassios τ^* statistic, has expanded the scope of dependence analysis beyond linear and monotone associations [1, 2, 3, 7, 8].

Despite these advances, a long-standing difficulty in statistical practice concerns the phenomenon of *spurious* or *nonsense* association. The earliest systematic discussion of this issue is generally attributed to Yule [4], who demonstrated that large correlations may arise between unrelated variables solely because both exhibit common temporal trends. Using historical data, Yule showed that variables with no meaningful scientific relationship can nevertheless display substantial correlation coefficients, thereby creating the illusion of a genuine association. This phenomenon subsequently became known as *nonsense correlation* and has remained an important concern in statistics, econometrics, and the social sciences.

The problem received renewed attention in the context of nonstationary time series. Granger and Newbold [5] demonstrated that regressions involving unrelated integrated processes may produce apparently significant relationships, high coefficients of determination, and statistically significant parameter estimates despite the absence of any substantive connection between the variables. Phillips [6] later provided rigorous asymptotic analyses of such spurious regressions, revealing the severe distortions that can arise when conventional inferential procedures are applied to nonstationary data. These findings established that the phenomenon extends far beyond the classical setting considered by Yule and remains relevant in many contemporary applications.

Although the notion of nonsense correlation was originally formulated in terms of Pearson's correlation coefficient, the underlying issue is considerably more general. Modern dependence measures are designed to detect departures from independence and therefore may also indicate substantial dependence when two variables share common trends, latent factors, or other external structures. In such situations, the observed dependence may be statistically genuine in a technical sense but scientifically uninformative. Consequently, it is natural to extend the classical notion of nonsense correlation to the broader concept of *nonsense dependence*, referring to dependence patterns that arise primarily from external artifacts rather than from an intrinsic stochastic relationship between the variables under investigation.

Among contemporary rank-based dependence measures, the Bergsma–Dassios τ^* statistic has attracted considerable attention. Introduced by Bergsma and Dassios [8], τ^* possesses the important property that, under mild conditions, it vanishes if and only if the variables are independent. Subsequent theoretical and computational developments have further established τ^* as a powerful nonparametric measure for dependence assessment [9, 10]. Nevertheless, like other dependence measures, τ^* is fundamentally designed to detect departures from independence rather than to distinguish meaningful dependence from potentially spurious dependence induced by common trends or external factors.

Motivated by this observation, we propose a simple diagnostic quantity, termed the *Nonsense Dependence Index* (NDI), defined by

$$\mathrm{NDI}(X, Y) = |\tau(X, Y)| - \sqrt{\tau^*(X, Y)},$$

where τ denotes Kendall's rank correlation coefficient and τ^* denotes the Bergsma–Dassios τ^* statistic. The proposed index contrasts first-order ordinal association, as captured by Kendall's τ, with higher-order dependence information represented by τ^*. Rather than serving as a new dependence measure, the NDI is intended as a diagnostic tool for identifying situations in which observed association may warrant further scrutiny for potential nonsense dependence.

The objectives of this article are threefold. First, we formally introduce the NDI and investigate its basic mathematical properties. Second, we examine its behaviour under a variety of dependence structures, including independence, monotone dependence, nonlinear dependence, and trend-induced dependence. Third, through simulation experiments and real-data applications involving classical examples of spurious association, we assess the practical utility of the proposed diagnostic and compare its performance with several established measures of association and dependence.

The remainder of the paper is organized as follows. Section 2 introduces the proposed Nonsense Dependence Index and discusses its interpretation. Section 3 establishes its basic theoretical properties and asymptotic behaviour. Section 4 examines the behaviour of the index under representative dependence structures. Simulation results are presented in Section 5, while Section 6 contains several real-data applications. Concluding remarks and directions for future research are provided in Section 7.

2 The Nonsense Dependence Index

The concept of nonsense dependence arises from the observation that substantial measured association between two variables does not necessarily imply the existence of a meaningful stochastic relationship. Classical examples involving trending time series have demonstrated that variables may exhibit strong apparent association even when the observed dependence is primarily driven by common temporal patterns, latent factors, or other external mechanisms rather than intrinsic interaction. While modern dependence measures are highly effective in detecting departures from independence, they are generally not designed to distinguish scientifically meaningful dependence from dependence induced by such external artifacts.

In this section, we introduce a simple diagnostic quantity, termed the *Nonsense Dependence Index* (NDI), which contrasts two complementary rank-based summaries of association. The proposed index combines Kendall's rank correlation coefficient and the Bergsma–Dassios τ^* statistic in order to assess whether observed ordinal association is supported by the higher-order dependence structure present in the data. The resulting quantity is intended not as a replacement for existing dependence measures but rather as a supplementary diagnostic tool for identifying situations in which detected dependence may warrant further investigation.

2.1 Motivation

The motivation for the proposed index stems from the distinct characteristics of Kendall's τ and Bergsma–Dassios' τ^*. Kendall's τ [3] is one of the most widely used rank-based measures of association and provides a natural summary of monotone dependence through pairwise concordance and discordance. Owing to its simplicity and interpretability, Kendall's τ has become a standard tool in nonparametric statistics and applied data analysis.

In contrast, the Bergsma–Dassios τ^* statistic [8] is based on higher-order sign relationships among quadruples of observations and possesses the remarkable property that, under mild regularity conditions, it vanishes if and only if the variables are independent. Consequently, τ^* captures a broader range of dependence structures than classical rank correlation coefficients and has emerged as a powerful tool for nonparametric dependence assessment.

Although both measures are useful for detecting departures from independence, they quantify different aspects of association. Kendall's τ primarily reflects first-order ordinal agreement between observations, whereas τ^* incorporates additional information concerning higher-order dependence patterns. In many situations involving genuine monotone dependence, these two quantities are expected to exhibit broadly compatible behaviour. However, when apparent association is induced by external trends or other spurious mechanisms, the relationship between the two measures may become less coherent.

This observation motivates the construction of a diagnostic quantity that directly contrasts the information provided by Kendall's τ and τ^*. Specifically, we seek a measure that remains small when the ordinal association and higher-order dependence structure are in agreement, but becomes comparatively large when substantial rank association is not accompanied by a correspondingly strong higher-order dependence pattern. Such discrepancies may serve as indicators of potential nonsense dependence.

To achieve this objective, we define the Nonsense Dependence Index by combining the magnitude of Kendall's τ with a transformed version of τ^*. The square-root transformation is motivated by the heuristic relationship

$$\tau^* \approx \tau^2,$$

which is frequently observed under monotone dependence structures. Consequently, comparing $|\tau|$ with $\sqrt{\tau^*}$ provides a natural means of assessing the degree of agreement between first-order and higher-order measures of association.

The proposed diagnostic is therefore intended to complement, rather than replace, existing dependence measures. Whereas Kendall's τ and τ^* are designed to quantify and detect dependence, the NDI seeks to identify situations in which the observed dependence may be potentially misleading from a substantive perspective. In this sense, the NDI may be viewed as a diagnostic tool for exploratory data analysis, particularly in applications involving trending, nonstationary, or otherwise structurally complex data.

2.2 Definition

Let (X, Y) be a pair of random variables with joint distribution function

$$F_{XY}(x, y) = P(X \leq x,\ Y \leq y).$$

Denote by $\tau(X, Y)$ the population version of Kendall's rank correlation coefficient and by $\tau^*(X, Y)$ the population Bergsma–Dassios τ^* statistic. Throughout this article, we assume that both quantities are well defined.

The *Nonsense Dependence Index* (NDI) is defined as

$$\mathrm{NDI}(X, Y) = |\tau(X, Y)| - \sqrt{\tau^*(X, Y)}.$$

Correspondingly, for a random sample

$$\{(X_1, Y_1), \ldots, (X_n, Y_n)\},$$

the empirical Nonsense Dependence Index is given by

$$\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*},$$

where $\widehat{\tau}$ and $\widehat{\tau}^*$ denote the sample versions of Kendall's τ and Bergsma–Dassios τ^*, respectively.

The proposed index is constructed by contrasting two complementary summaries of dependence. The first component, $|\tau|$, measures the magnitude of ordinal association between the variables and is based on pairwise concordance and discordance. The second component, $\sqrt{\tau^*}$, reflects higher-order dependence information extracted from quadruples of observations. The difference between these two quantities therefore provides a measure of the agreement between first-order and higher-order dependence structures.

The NDI is intended as a diagnostic quantity rather than a dependence measure in the traditional sense. While both Kendall's τ and τ^* are designed to quantify departures from independence, the NDI seeks to identify situations in which the strength of observed rank association may not be fully supported by the underlying higher-order dependence structure.

2.3 Interpretation

The interpretation of the NDI is based on the relative magnitudes of $|\tau|$ and $\sqrt{\tau^*}$. Since Kendall's τ and τ^* capture different aspects of dependence, the quantity $\mathrm{NDI} = |\tau| - \sqrt{\tau^*}$ may be viewed as a measure of discrepancy between first-order ordinal association and higher-order dependence. Values of the NDI close to zero indicate substantial agreement between the two measures. Such behaviour is expected under independence, where both τ and τ^* vanish, as well as under many monotone dependence structures where the relationship between $|\tau|$ and $\sqrt{\tau^*}$ is approximately balanced. In these situations, the ordinal association detected by Kendall's τ is broadly consistent with the higher-order dependence structure captured by τ^*.

Positive values of the NDI occur when the magnitude of Kendall's τ exceeds the corresponding transformed τ^* value. This indicates that the observed rank association is comparatively stronger than the higher-order dependence reflected by τ^*. Such situations may arise in the presence of common trends, latent temporal effects, aggregation artifacts, or other mechanisms capable of generating apparent association without a correspondingly rich dependence structure. Consequently, relatively large positive values of the NDI may be regarded as indicators of potential nonsense dependence.

Negative values of the NDI indicate that the transformed τ^* value exceeds the magnitude of Kendall's τ. This behaviour suggests the presence of dependence structures that are not adequately summarized by monotone rank association alone. Examples include certain nonlinear and nonmonotone relationships for which τ^* remains sensitive despite relatively weak values of Kendall's τ.

It is important to emphasize that the NDI is not intended to provide a formal test for nonsense dependence. Rather, it should be viewed as a diagnostic tool whose values may guide further investigation of the underlying dependence structure. Large positive values suggest that the detected association deserves additional scrutiny, while values near zero indicate consistency between first-order and higher-order measures of dependence.

The practical usefulness of the NDI therefore lies in its ability to supplement existing dependence measures. When interpreted alongside Kendall's τ, τ^*, and other established measures of association, the proposed index provides an additional perspective on the nature of the observed relationship and may help identify situations in which apparent dependence is potentially misleading.

2.4 Novelty of the Proposed Index

The proposed Nonsense Dependence Index (NDI) introduces a novel perspective for the assessment of statistical dependence by explicitly contrasting two fundamentally different rank-based measures of association, namely Kendall's τ and the Bergsma–Dassios τ^* statistic. Unlike existing measures, which are primarily designed to quantify or detect departures from independence, the NDI is specifically intended to diagnose potential *nonsense dependence*, that is, apparent associations that may arise from common trends, latent factors, or other external artifacts rather than from a meaningful stochastic relationship.

A distinctive feature of the proposed index is that it combines first-order ordinal association, represented by $|\tau|$, with higher-order dependence information captured by τ^*. Consequently, the NDI provides information that is not directly available from either measure individually. Rather than introducing yet another dependence coefficient, the proposed methodology offers a diagnostic framework for evaluating the coherence between different aspects of dependence.

To the best of our knowledge, no existing measure in the literature has been specifically developed to identify potential nonsense dependence through a direct comparison of Kendall's τ and Bergsma–Dassios τ^*. The proposed index therefore contributes a simple, interpretable, and computationally convenient diagnostic tool that complements existing nonparametric measures of association and dependence.

3 Basic Properties of the Nonsense Dependence Index

3.1 Fundamental Properties

We begin by establishing several elementary properties of the proposed Nonsense Dependence Index. Throughout this section, we assume that τ denotes Kendall's rank correlation coefficient and τ^* denotes the Bergsma–Dassios τ^* statistic [8]. Recall that

$$\mathrm{NDI}(X,Y) = |\tau(X,Y)| - \sqrt{\tau^*(X,Y)}.$$

The following results provide the range of the index together with its behaviour under independence and perfect monotone dependence.

Theorem 1 (Bounds)**.** *For any pair of random variables* (X, Y) *for which* τ *and* τ^* *are well defined,*

$$-1 \leq \mathrm{NDI}(X,Y) \leq 1.$$

Proof. By definition,

$$\mathrm{NDI}(X,Y) = |\tau(X,Y)| - \sqrt{\tau^*(X,Y)}.$$

Since Kendall's tau always lies between -1 and 1, it follows that

$$0 \leq |\tau(X,Y)| \leq 1.$$

Similarly, Bergsma–Dassios τ^* takes values in the interval $[0, 1]$, so

$$0 \leq \sqrt{\tau^*(X,Y)} \leq 1.$$

Therefore, the NDI is the difference of two quantities belonging to $[0, 1]$. The largest possible value is obtained when $|\tau(X,Y)| = 1$ and $\sqrt{\tau^*(X,Y)} = 0$, giving $\mathrm{NDI}(X,Y) = 1$. Likewise, the smallest possible value is obtained when $|\tau(X,Y)| = 0$ and $\sqrt{\tau^*(X,Y)} = 1$, giving $\mathrm{NDI}(X,Y) = -1$. Hence, $-1 \leq \mathrm{NDI}(X,Y) \leq 1$. □

Theorem 2 (Independence)**.** *If* X *and* Y *are independent, then* $\mathrm{NDI}(X,Y) = 0$.

Proof. When X and Y are independent, Kendall's tau is equal to zero. Furthermore, Bergsma and Dassios (2014) showed that τ^* also vanishes under independence. Substituting these values into the definition of the NDI yields $\mathrm{NDI}(X,Y) = |0| - \sqrt{0} = 0$.

Therefore, the NDI equals zero whenever X and Y are independent. □

Theorem 3 (Perfect Monotone Dependence)**.** *Suppose that* $Y = f(X)$ *almost surely, where* f *is a strictly monotone function. Then* $\mathrm{NDI}(X,Y) = 0$.

Proof. If f is strictly increasing, every pair of observations is concordant and Kendall's tau equals 1. If f is strictly decreasing, every pair is discordant and Kendall's tau equals -1. In either case, $|\tau(X,Y)| = 1$. Moreover, under perfect monotone dependence, the ordering induced by X and Y is completely preserved, implying that τ^* also attains its maximum value, $\tau^*(X,Y) = 1$. Consequently, $\mathrm{NDI}(X,Y) = 1 - \sqrt{1} = 0$. Hence, the NDI vanishes under perfect monotone dependence. □

3.2 Additional Properties

In addition to the fundamental properties established in the previous subsection, the proposed Nonsense Dependence Index possesses several desirable structural characteristics inherited from Kendall's tau and the Bergsma–Dassios τ^* statistic. These properties ensure that the index remains interpretable under common transformations of the data and provide

further insight into its behaviour under various dependence structures, i.e. $\mathrm{NDI}(X, Y) = 0$ under both independence and perfect monotone dependence.

The above results demonstrate that the Nonsense Dependence Index inherits many of the attractive features of rank-based statistics. In particular, it is symmetric, invariant under strictly monotone transformations, unaffected by sign reversals, and bounded within a fixed interval. These properties contribute to the interpretability and robustness of the proposed diagnostic and facilitate meaningful comparisons across diverse applications.

4 Asymptotic Properties

In this section, we investigate the large-sample behaviour of the proposed Nonsense Dependence Index. Let $\mathrm{NDI} = |\tau| - \sqrt{\tau^*}$ denote the population index, where τ and τ^* are the population versions of Kendall's tau and Bergsma–Dassios τ^*, respectively. Given a random sample $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, the corresponding sample NDI is defined as

$$\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*},$$

where $\widehat{\tau}$ and $\widehat{\tau}^*$ denote the sample estimators of τ and τ^*.

Since both $\widehat{\tau}$ and $\widehat{\tau}^*$ are U-statistics, the asymptotic behaviour of $\widehat{\mathrm{NDI}}$ can be studied using classical U-statistic theory together with the continuous mapping theorem and the delta method. We begin by establishing the consistency of the proposed estimator.

4.1 Consistency of the Sample NDI

The first asymptotic result shows that the sample NDI consistently estimates its population counterpart.

Theorem 4 (Strong Consistency of the Sample NDI)**.** *Suppose that* $E\big|h_\tau(Z_1, Z_2)\big| < \infty$ *and* $E\big|h_{\tau^*}(Z_1, Z_2, Z_3, Z_4)\big| < \infty$*, where* h_τ *and* h_{τ^*} *denote the kernels corresponding to Kendall's tau and Bergsma–Dassios* τ^**, respectively, and* $Z_i = (X_i, Y_i)$*. Then,* $\widehat{\mathrm{NDI}} \xrightarrow{a.s.} \mathrm{NDI}$*, as* $n \to \infty$*.*

Proof. The sample Kendall's tau $\widehat{\tau}$ is a U-statistic of order two with bounded kernel

$$h_\tau(Z_i, Z_j) = \mathrm{sgn}\big((X_i - X_j)(Y_i - Y_j)\big).$$

By the strong law of large numbers for U-statistics, $\widehat{\tau} \xrightarrow{a.s.} \tau$. Similarly, the sample Bergsma–Dassios τ^* statistic $\widehat{\tau}^*$ is a U-statistic of order four. Since its kernel is bounded, the strong law for U-statistics yields $\widehat{\tau}^* \xrightarrow{a.s.} \tau^*$.

Consider the mapping

$$g(u, v) = |u| - \sqrt{v}, \qquad (u, v) \in [-1, 1] \times [0, 1].$$

The function g is continuous on its domain. Therefore, by the continuous mapping theorem, $g(\widehat{\tau}, \widehat{\tau}^*) \xrightarrow{a.s.} g(\tau, \tau^*)$.

Since $g(\widehat{\tau}, \widehat{\tau}^*) = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*} = \widehat{\mathrm{NDI}}$ and $g(\tau, \tau^*) = |\tau| - \sqrt{\tau^*} = \mathrm{NDI}$, it follows that $\widehat{\mathrm{NDI}} \xrightarrow{a.s.} \mathrm{NDI}$. Hence, the sample NDI is a strongly consistent estimator of the population NDI. □

Theorem 4 establishes that, as the sample size increases, the empirical NDI converges almost surely to its population counterpart. Consequently, the proposed diagnostic is statistically stable and its large-sample behaviour is inherited from the well-established consistency properties of Kendall's tau and Bergsma–Dassios τ^*.

The above result immediately implies consistency in probability.

Corollary 1. *Under the assumptions of Theorem 4,*

$$\widehat{\mathrm{NDI}} \xrightarrow{P} \mathrm{NDI}, \qquad n \to \infty.$$

Proof. The result follows directly from the fact that almost sure convergence implies convergence in probability. □

The consistency result is particularly important from a practical perspective. It guarantees that, for sufficiently large samples, the observed value of the NDI accurately reflects the underlying population discrepancy between first-order ordinal association and higher-order dependence structure. Therefore, any conclusions regarding potential nonsense dependence become increasingly reliable as the sample size grows.

4.2 Influence of Estimation Error in τ^*

The Nonsense Dependence Index is defined as $\mathrm{NDI} = |\tau| - \sqrt{\tau^*}$, with sample counterpart $\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*}$. Since the NDI depends nonlinearly on both Kendall's tau and Bergsma–Dassios τ^*, it is natural to investigate how estimation errors in these quantities affect the resulting estimator. Particular attention is required for the square-root transformation applied to τ^*, as it may amplify estimation variability when τ^* is close to zero.

Consider the function $g(u, v) = |u| - \sqrt{v}, \ (u, v) \in [-1, 1] \times (0, 1]$. Assume that $\tau \neq 0$ and $\tau^* > 0$. In a neighbourhood of (τ, τ^*), the function g is continuously differentiable with gradient

$$\nabla g(\tau, \tau^*) = \begin{pmatrix} \mathrm{sgn}(\tau) \\ -\dfrac{1}{2\sqrt{\tau^*}} \end{pmatrix}.$$

A first-order Taylor expansion therefore gives

$$\widehat{\mathrm{NDI}} - \mathrm{NDI} = \mathrm{sgn}(\tau)(\widehat{\tau} - \tau) - \frac{1}{2\sqrt{\tau^*}}(\widehat{\tau}^* - \tau^*) + o_p(n^{-1/2}).$$

This representation shows that the asymptotic behaviour of the NDI is determined by a weighted combination of the estimation errors in $\widehat{\tau}$ and $\widehat{\tau}^*$. The weight associated with the τ^* component, $\dfrac{1}{2\sqrt{\tau^*}}$, is particularly informative. When τ^* is small, this coefficient becomes large, indicating that even modest estimation errors in $\widehat{\tau}^*$ may substantially affect the resulting NDI estimate. In contrast, when τ^* is close to one, the contribution of the τ^* estimation error becomes comparatively less important. The following theorem formalizes this linearization.

Theorem 5. *Suppose that $\tau \neq 0$ and $\tau^* > 0$, and assume that*

$$\sqrt{n}\begin{pmatrix} \widehat{\tau} - \tau \\ \widehat{\tau}^* - \tau^* \end{pmatrix} = O_p(1).$$

Then, $\widehat{\mathrm{NDI}} - \mathrm{NDI} = \mathrm{sgn}(\tau)(\widehat{\tau} - \tau) - \dfrac{1}{2\sqrt{\tau^*}}(\widehat{\tau}^* - \tau^*) + o_p(n^{-1/2})$.

Proof. Because $\tau \neq 0$ and $\tau^* > 0$, the function $g(u, v) = |u| - \sqrt{v}$ is continuously differentiable in a neighbourhood of (τ, τ^*). Applying the multivariate delta method yields

$$g(\widehat{\tau}, \widehat{\tau}^*) = g(\tau, \tau^*) + \nabla g(\tau, \tau^*)^\top \begin{pmatrix} \widehat{\tau} - \tau \\ \widehat{\tau}^* - \tau^* \end{pmatrix} + o_p(n^{-1/2}).$$

Substituting the expression for the gradient immediately gives the stated expansion. □

Theorem 5 provides a stochastic linearization of the sample NDI and serves as the foundation for deriving its asymptotic distribution.

Theorem 6 (Boundary Asymptotics Under Independence)**.** *Suppose that X and Y are independent. Further assume that $\sqrt{n}\,\widehat{\tau} \xrightarrow{d} N(0, \sigma_\tau^2)$ and $n\,\widehat{\tau}^* \xrightarrow{d} W$, where W denotes the non-degenerate limiting random variable associated with the degenerate U-statistic $\widehat{\tau}^*$ under independence. Then*

$$\sqrt{n}\,\widehat{\mathrm{NDI}} = \sqrt{n}\left(|\widehat{\tau}| - \sqrt{\widehat{\tau}^*}\right) \xrightarrow{d} |Z|, \;\; \textit{where} \;\; Z \sim N(0, \sigma_\tau^2).$$

Proof. Under the null hypothesis of independence, $\tau = 0,\ \tau^* = 0$. Therefore, the population value of the Nonsense Dependence Index is also zero, and its sample counterpart reduces to

$$\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*}.$$

To determine its limiting distribution, multiply both sides by $\sqrt{n}$:

$$\sqrt{n}\,\widehat{\mathrm{NDI}} = \sqrt{n}\,|\widehat{\tau}| - \sqrt{n}\sqrt{\widehat{\tau}^*}.$$

The asymptotic behaviour of the first term follows from the classical theory for Kendall's tau. Specifically,

$$\sqrt{n}\,\widehat{\tau} \xrightarrow{d} Z, \qquad Z \sim N(0, \sigma_\tau^2),$$

and therefore, by the continuous mapping theorem, $\sqrt{n}\,|\widehat{\tau}| = \left|\sqrt{n}\,\widehat{\tau}\right| \xrightarrow{d} |Z|$.

Next, consider the Bergsma–Dassios component. Under independence, it is known that $n\widehat{\tau}^* = O_p(1)$, which implies $\widehat{\tau}^* = O_p(n^{-1})$. Taking square roots yields $\sqrt{\widehat{\tau}^*} = O_p(n^{-1/2})$, and consequently $\sqrt{n}\sqrt{\widehat{\tau}^*} = O_p(1)$.

Thus, the contribution of the τ^*-term remains bounded in probability, whereas the dominant first-order behaviour is governed by the Kendall tau component. Combining these observations with the convergence of $\sqrt{n}\,|\widehat{\tau}|$ and applying Slutsky's theorem, we obtain

$$\sqrt{n}\,\widehat{\mathrm{NDI}} \xrightarrow{d} |Z|.$$

Hence, under the null hypothesis of independence, the scaled Nonsense Dependence Index converges in distribution to the folded normal random variable $|Z|$, where $Z \sim N(0, \sigma_\tau^2)$. This establishes the stated result.

□

4.3 Asymptotic Distribution of the Sample NDI

We now establish the limiting distribution of the sample NDI. The result follows from the asymptotic theory of U-statistics together with the multivariate delta method.

Let

$$\sqrt{n}\begin{pmatrix}\widehat{\tau}-\tau\\ \widehat{\tau}^*-\tau^*\end{pmatrix}\xrightarrow{d} N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\Sigma\right),$$

where $\Sigma=\begin{pmatrix}\sigma_\tau^2 & \sigma_{\tau,\tau^*}\\ \sigma_{\tau,\tau^*} & \sigma_{\tau^*}^2\end{pmatrix}$ denotes the asymptotic covariance matrix of the joint estimator $(\widehat{\tau},\widehat{\tau}^*)$. The existence of this joint asymptotic normal distribution follows from standard multivariate U-statistic theory. We now obtain the limiting distribution of the NDI.

Theorem 7 (Asymptotic Normality of the Sample NDI)**.** *Suppose that $\tau\neq 0$ and $\tau^*>0$. Assume further that*

$$\sqrt{n}\begin{pmatrix}\widehat{\tau}-\tau\\ \widehat{\tau}^*-\tau^*\end{pmatrix}\xrightarrow{d} N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\Sigma\right).$$

Then,

$$\sqrt{n}\left(\widehat{\mathrm{NDI}}-\mathrm{NDI}\right)\xrightarrow{d} N(0,\sigma_{\mathrm{NDI}}^2),$$

where $\sigma_{\mathrm{NDI}}^2=\nabla g(\tau,\tau^*)^\top\Sigma\nabla g(\tau,\tau^*)$, *with* $\nabla g(\tau,\tau^*)=\begin{pmatrix}\mathrm{sgn}(\tau)\\ -\dfrac{1}{2\sqrt{\tau^*}}\end{pmatrix}$.

Proof. The result follows from the first-order asymptotic expansion established in Theorem 5. Specifically,

$$\widehat{\mathrm{NDI}}-\mathrm{NDI}=\nabla g(\tau,\tau^*)^\top\begin{pmatrix}\widehat{\tau}-\tau\\ \widehat{\tau}^*-\tau^*\end{pmatrix}+o_p(n^{-1/2}),$$

where $\nabla g(\tau,\tau^*)$ denotes the gradient of the transformation

$$g(\tau,\tau^*)=|\tau|-\sqrt{\tau^*}.$$

Multiplying both sides by $\sqrt{n}$, we obtain

$$\sqrt{n}\left(\widehat{\mathrm{NDI}}-\mathrm{NDI}\right)=\nabla g(\tau,\tau^*)^\top\sqrt{n}\begin{pmatrix}\widehat{\tau}-\tau\\ \widehat{\tau}^*-\tau^*\end{pmatrix}+o_p(1).$$

By assumption, the estimators $\widehat{\tau}$ and $\widehat{\tau}^*$ satisfy the joint asymptotic normality condition

$$\sqrt{n}\begin{pmatrix}\widehat{\tau}-\tau\\ \widehat{\tau}^*-\tau^*\end{pmatrix}\xrightarrow{d} N_2(0,\Sigma),$$

where Σ denotes the asymptotic covariance matrix of the vector $(\widehat{\tau},\widehat{\tau}^*)^\top$.

The linear representation above shows that the scaled estimation error of the NDI is an asymptotically linear transformation of this bivariate normal limit. Therefore, by the multivariate delta method,

$$\nabla g(\tau,\tau^*)^\top \sqrt{n}\begin{pmatrix}\widehat{\tau}-\tau\\ \widehat{\tau}^*-\tau^*\end{pmatrix} \xrightarrow{d} N\left(0, \nabla g(\tau,\tau^*)^\top \Sigma \nabla g(\tau,\tau^*)\right).$$

Finally, since the remainder term $o_p(1)$ converges to zero in probability, Slutsky's theorem yields

$$\sqrt{n}\left(\widehat{\mathrm{NDI}} - \mathrm{NDI}\right) \xrightarrow{d} N\left(0, \nabla g(\tau,\tau^*)^\top \Sigma \nabla g(\tau,\tau^*)\right).$$

Thus, the Nonsense Dependence Index is asymptotically normal with asymptotic variance given by the quadratic form

$$\nabla g(\tau,\tau^*)^\top \Sigma \nabla g(\tau,\tau^*),$$

which reflects the joint variability of Kendall's tau and the Bergsma–Dassios coefficient. This completes the proof.

□

Expanding the asymptotic variance yields

$$\sigma^2_{\mathrm{NDI}} = \sigma^2_\tau + \frac{\sigma^2_{\tau^*}}{4\tau^*} - \frac{\mathrm{sgn}(\tau)}{\sqrt{\tau^*}}\,\sigma_{\tau,\tau^*}.$$

Theorem 7 establishes that the sample NDI is asymptotically normal and therefore admits standard large-sample inference procedures, including Wald-type confidence intervals and hypothesis tests. Moreover, the asymptotic variance explicitly quantifies the combined contributions of Kendall's tau, Bergsma–Dassios τ^*, and their asymptotic covariance to the uncertainty associated with estimation of the proposed index.

4.3.1 The Boundary Case of Independence

The asymptotic normality result established in Theorem 7 requires the regularity conditions $\tau \neq 0$ and $\tau^* > 0$. These assumptions exclude the important case of independence, for which $\tau = \tau^* = 0$. At this boundary point, the transformation $g(u,v) = |u| - \sqrt{v}$ is not differentiable, since both the absolute value and square-root functions fail to be differentiable at zero. Consequently, the classical multivariate delta method is not applicable, necessitating a separate asymptotic analysis.

The following theorem characterizes the limiting behaviour of the sample NDI under independence.

Theorem 8 (Asymptotic Distribution of the Sample NDI Under Independence)**.** *Suppose that X and Y are independent, so that $\tau = \tau^* = 0$. Assume that $\sqrt{n}\,\widehat{\tau} \xrightarrow{d} Z$, $Z \sim N(0,\sigma^2_\tau)$, and $n\,\widehat{\tau}^* \xrightarrow{d} W$, where W denotes the non-degenerate limiting random variable associated with the degenerate U-statistic $\widehat{\tau}^*$ under independence. Then*

$$\sqrt{n}\,\widehat{\mathrm{NDI}} \xrightarrow{d} |Z| - \sqrt{W}.$$

Proof. Under the null hypothesis of independence, both Kendall's tau and the Bergsma–Dassios coefficient vanish at the population level, that is, $\tau = \tau^* = 0$. Consequently, the population value of the Nonsense Dependence Index is also equal to zero:

$$\mathrm{NDI} = |\tau| - \sqrt{\tau^*} = 0.$$

To study the asymptotic behaviour of the estimator, consider the sample version $\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*}$. Multiplying both sides by $\sqrt{n}$ gives $\sqrt{n}\,\widehat{\mathrm{NDI}} = \sqrt{n}\,|\widehat{\tau}| - \sqrt{n\widehat{\tau}^*}$.

The limiting behaviour of the first term follows from the asymptotic normality of Kendall's tau. Specifically, by assumption, $\sqrt{n}\,\widehat{\tau} \xrightarrow{d} Z$, where Z is a centred Gaussian random variable. Since the absolute value function is continuous, the continuous mapping theorem implies that

$$\sqrt{n}\,|\widehat{\tau}| = \left|\sqrt{n}\,\widehat{\tau}\right| \xrightarrow{d} |Z|.$$

Next, consider the Bergsma–Dassios component. Under independence,

$$n\,\widehat{\tau}^* \xrightarrow{d} W,$$

where W is a nonnegative limiting random variable. Applying the continuous mapping theorem to the square-root function yields $\sqrt{n\widehat{\tau}^*} \xrightarrow{d} \sqrt{W}$. Combining these results, we obtain the joint convergence

$$\left(\sqrt{n}\,|\widehat{\tau}|,\ \sqrt{n\widehat{\tau}^*}\right) \xrightarrow{d} \left(|Z|, \sqrt{W}\right).$$

Finally, since subtraction is a continuous mapping, one more application of the continuous mapping theorem gives $\sqrt{n}\,\widehat{\mathrm{NDI}} = \sqrt{n}\,|\widehat{\tau}| - \sqrt{n\widehat{\tau}^*} \xrightarrow{d} |Z| - \sqrt{W}$. Therefore, under the null hypothesis of independence, the scaled Nonsense Dependence Index converges in distribution to the random variable $(|Z| - \sqrt{W})$. This establishes the desired asymptotic distribution and completes the proof.

□

Remark 1. *The limiting distribution in Theorem 8 is generally non-Gaussian. This contrasts with the regular asymptotic normality established in Theorem 7. The non-Gaussian limit arises because independence corresponds to the boundary point $(\tau, \tau^*) = (0, 0)$, where the transformation defining the NDI fails to be differentiable.*

Consequently, standard Wald-type inference based on asymptotic normality is not directly applicable under independence. In such situations, bootstrap or permutation-based procedures may provide more accurate finite-sample inference.

4.4 Confidence Intervals and Statistical Inference

The asymptotic results established in the previous subsections provide a theoretical basis for statistical inference concerning the Nonsense Dependence Index. In the regular case, where $\tau \neq 0$ and $\tau^* > 0$, Theorem 7 implies that the sample NDI is asymptotically normal. Consequently, standard large-sample confidence intervals and hypothesis testing procedures may be constructed. Specifically, Theorem 7 yields

$$\sqrt{n}\left(\widehat{\mathrm{NDI}} - \mathrm{NDI}\right) \xrightarrow{d} N(0, \sigma^2_{\mathrm{NDI}}),$$

where $\sigma^2_{\mathrm{NDI}} = \nabla g(\tau, \tau^*)^\top \Sigma \nabla g(\tau, \tau^*)$. Let $\widehat{\sigma}^2_{\mathrm{NDI}}$ denote a consistent estimator of the asymptotic variance. Then, $\dfrac{\sqrt{n}\left(\widehat{\mathrm{NDI}} - \mathrm{NDI}\right)}{\widehat{\sigma}_{\mathrm{NDI}}} \xrightarrow{d} N(0,1)$.

The above result immediately leads to asymptotic confidence intervals for the population NDI.

Theorem 9 (Asymptotic Confidence Interval)**.** *Suppose that the assumptions of Theorem 7 hold. Then an asymptotic* $(1-\alpha)$ *confidence interval for* NDI *is given by*

$$\left(\widehat{\mathrm{NDI}} - z_{\alpha/2}\frac{\widehat{\sigma}_{\mathrm{NDI}}}{\sqrt{n}},\ \widehat{\mathrm{NDI}} + z_{\alpha/2}\frac{\widehat{\sigma}_{\mathrm{NDI}}}{\sqrt{n}}\right),$$

where $z_{\alpha/2}$ *denotes the upper* $\alpha/2$ *quantile of the standard normal distribution.*

Proof. By Theorem 7, it is evident that $\dfrac{\sqrt{n}\left(\widehat{\mathrm{NDI}} - \mathrm{NDI}\right)}{\sigma_{\mathrm{NDI}}} \xrightarrow{d} N(0,1)$. Since $\widehat{\sigma}_{\mathrm{NDI}} \xrightarrow{P} \sigma_{\mathrm{NDI}}$, Slutsky's theorem gives $\dfrac{\sqrt{n}\left(\widehat{\mathrm{NDI}} - \mathrm{NDI}\right)}{\widehat{\sigma}_{\mathrm{NDI}}} \xrightarrow{d} N(0,1)$. Hence, it can be further deduced that $P\left(-z_{\alpha/2} \le \dfrac{\sqrt{n}\left(\widehat{\mathrm{NDI}} - \mathrm{NDI}\right)}{\widehat{\sigma}_{\mathrm{NDI}}} \le z_{\alpha/2}\right) \longrightarrow 1-\alpha$. Rearranging the inequality yields the stated confidence interval.

□

We now consider hypothesis testing. Since large positive values of the NDI are indicative of potential nonsense dependence, a natural inferential problem is

$$H_0 : \mathrm{NDI} = 0 \ \text{ against } \ H_1 : \mathrm{NDI} > 0.$$

Under the regular asymptotic regime, the corresponding Wald-type test statistic is $T_n = \dfrac{\sqrt{n}\,\widehat{\mathrm{NDI}}}{\widehat{\sigma}_{\mathrm{NDI}}}$. Under H_0, $T_n \xrightarrow{d} N(0,1)$ and the null hypothesis is rejected at significance level α whenever $T_n > z_\alpha$. The following theorem summarizes this result.

Theorem 10 (Wald-Type Test for NDI)**.** *Assume the conditions of Theorem 7. Then the test that rejects* $H_0 : \mathrm{NDI} = 0$ *against* $H_1 : \mathrm{NDI} > 0$ *whenever* $T_n > z_\alpha$ *has asymptotic significance level* α*.*

Proof. Under H_0, NDI $= 0$. Hence, $T_n = \dfrac{\sqrt{n}\left(\widehat{\mathrm{NDI}} - \mathrm{NDI}\right)}{\widehat{\sigma}_{\mathrm{NDI}}}$. By Theorem 7 and Slutsky's theorem, $T_n \xrightarrow{d} N(0,1)$. Therefore, $P(T_n > z_\alpha) \longrightarrow \alpha$, which establishes the asymptotic level of the test. □

The preceding inference procedures are valid only in the regular setting $\tau \neq 0,\ \tau^* > 0$. When the variables are independent, Theorem 8 shows that the limiting distribution of the sample NDI is generally non-Gaussian. Consequently, Wald-type intervals and tests may exhibit poor finite-sample performance near the independence boundary.

For such situations, bootstrap and permutation methods provide attractive alternatives. Let $\widehat{\mathrm{NDI}}^{*(1)}, \ldots, \widehat{\mathrm{NDI}}^{*(B)}$ denote bootstrap replicates of the sample NDI obtained from B resamples. Percentile-based confidence intervals may then be constructed from the empirical quantiles of the bootstrap distribution. Similarly, permutation procedures based on randomly permuting one variable while keeping the other fixed yield nonparametric significance tests that preserve the null hypothesis of independence.

The use of bootstrap and permutation methods is particularly appealing because both Kendall's tau and Bergsma–Dassios τ^* are rank-based statistics, making the resulting inference procedures robust and broadly applicable across a wide range of dependence structures. Consequently, the proposed inferential framework allows practitioners not only to estimate the magnitude of the NDI but also to formally assess the statistical evidence supporting the presence of potential nonsense dependence.

5 Behaviour Under Representative Dependence Structures

To better understand the practical interpretation of the proposed Nonsense Dependence Index, it is instructive to examine its behaviour under several representative dependence structures. In this section, we investigate the theoretical behaviour of the NDI under independence and various forms of dependence. Particular attention is paid to the relationship between Kendall's tau and Bergsma–Dassios τ^*, since the NDI is constructed from the discrepancy between these two quantities.

The examples considered in this section serve two purposes. First, they provide insight into the expected magnitude and sign of the NDI under well-understood dependence structures. Second, they establish a conceptual framework for the simulation studies presented in the following section.

5.1 Independence

Independence represents the fundamental benchmark against which all dependence measures are assessed. Since both Kendall's tau and Bergsma–Dassios τ^* are designed to vanish under independence, the proposed NDI is also expected to attain its baseline value in this setting.

Suppose that X and Y are independent random variables. Then

$$F_{XY}(x, y) = F_X(x)F_Y(y),$$

for all $(x, y) \in \mathbb{R}^2$.

By the defining property of Kendall's tau, $\tau(X, Y) = 0$. Similarly, Bergsma and Dassios (2014) showed that $\tau^*(X, Y) = 0$ under independence. Substituting these values into the definition of the NDI yields $\boxed{\mathrm{NDI}(X, Y) = |\tau(X, Y)| - \sqrt{\tau^*(X, Y)} = 0.}$

Consequently, independence corresponds to the reference point $\boxed{\text{NDI} = 0.}$ This result establishes an important baseline for interpreting the proposed index. Any substantial deviation from zero must therefore arise from some form of dependence structure present in the data.

From a finite-sample perspective, however, the empirical quantities $\widehat{\tau}$ and $\widehat{\tau}^*$ are subject to sampling variability. Consequently, $\widehat{\text{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*}$ will generally fluctuate around zero even when the variables are independent. By the consistency results established in Section 4,

$$\widehat{\text{NDI}} \xrightarrow{P} 0, \qquad n \to \infty.$$

Thus, under independence, the NDI possesses a stable limiting behaviour and converges to its theoretical benchmark value. To illustrate this phenomenon, consider the simple model

$$X \sim N(0,1), \qquad Y \sim N(0,1),$$

with X and Y independent. For this model, $\tau = 0,\ \tau^* = 0,\ \text{NDI} = 0.$

An analogous conclusion holds for any independent continuous distributions. Therefore, unlike many classical dependence measures whose interpretation relies on the magnitude of a single coefficient, the proposed NDI uses the independence case primarily as a reference level against which more complex dependence structures may be compared.

5.2 Linear Dependence

We next consider the behaviour of the NDI under genuine linear association. Linear dependence constitutes one of the most commonly encountered forms of statistical relationship and provides a useful benchmark for evaluating the performance of the proposed index.

Suppose that (X, Y) follows a bivariate distribution exhibiting positive linear dependence. A classical example is the bivariate normal model

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

where $\rho \in (-1, 1)$ denotes the Pearson correlation coefficient. For the bivariate normal distribution, Kendall's tau is given by $\boxed{\tau = \frac{2}{\pi}\sin^{-1}(\rho).}$ Consequently, the magnitude of Kendall's tau increases monotonically with the strength of the linear association. Since linear dependence represents a genuine departure from independence, Bergsma–Dassios τ^* is also strictly positive whenever $\rho \neq 0$. Moreover, as the strength of dependence increases, both τ and τ^* approach their maximal values.

For perfect linear dependence, $Y = a + bX$, all observation pairs are either concordant ($b > 0$) or discordant ($b < 0$). Consequently, $|\tau| = 1$ and $\tau^* = 1$, yielding $\text{NDI} = |\tau| - \sqrt{\tau^*} = 1 - \sqrt{1} = 0$. For intermediate values of ρ, both Kendall's tau and τ^* increase simultaneously as the degree of linear association strengthens. Consequently, the difference $|\tau| - \sqrt{\tau^*}$ is expected to remain relatively small.

This behaviour is consistent with the intended interpretation of the NDI. Since linear dependence represents a genuine and well-structured form of association, there is no reason to expect a substantial discrepancy between ordinal association and higher-order dependence.

Therefore, the NDI should generally remain close to zero throughout a broad range of linear dependence models.

From a practical perspective, this property is desirable because it prevents the proposed index from falsely identifying ordinary linear relationships as examples of nonsense dependence.

5.3 Monotone Nonlinear Dependence

We now consider monotone nonlinear relationships. Such relationships frequently arise in economics, biology, environmental science, and engineering applications where the response changes systematically with the explanatory variable but not necessarily at a constant rate.

Let $Y = f(X)$, where f is a strictly monotone nonlinear function. Typical examples include $Y = e^X$, $Y = \log(X)$, $Y = X^3$ and $Y = \sqrt{X}$. Although these relationships are nonlinear, they preserve the ordering of observations and therefore maintain a strong ordinal structure.

Suppose first that f is strictly increasing. Then for any two observations,

$$X_i < X_j \quad \Longrightarrow \quad f(X_i) < f(X_j).$$

Hence every pair of observations is concordant, implying $\tau = 1$. Likewise, because the dependence is perfectly monotone, $\tau^* = 1$. Consequently, $\text{NDI} = 1 - \sqrt{1} = 0$. If f is strictly decreasing, every pair of observations is discordant, yielding $\tau = -1$ and therefore $|\tau| = 1$. Since the dependence remains perfectly monotone, $\tau^* = 1$, and again $\text{NDI} = 0$. Thus, perfect monotone nonlinear relationships are treated in exactly the same manner as perfect linear relationships by the NDI.

In practice, however, data are often contaminated by random noise. A more realistic model is $Y = f(X) + \varepsilon$, where $E(\varepsilon) = 0$ and ε is independent of X. Under such models, the relationship remains monotone on average but is no longer deterministic. Consequently, $0 < |\tau| < 1$ and $0 < \tau^* < 1$. Because both measures respond to the same underlying monotone structure, they are expected to increase together as the signal-to-noise ratio increases. Therefore, the resulting NDI values should generally remain modest in magnitude.

This observation is particularly important for the interpretation of the proposed index. Many classical nonlinear relationships exhibit strong dependence while remaining monotone. Since both Kendall's tau and τ^* capture this dependence effectively, little discrepancy arises between the two measures, and the NDI remains close to its benchmark value.

Accordingly, monotone nonlinear dependence should not be interpreted as nonsense dependence. Rather, it represents a genuine form of association for which the proposed index is expected to exhibit stable and comparatively small values.

5.4 Nonmonotone Dependence

We now investigate the behaviour of the proposed Nonsense Dependence Index under nonmonotone dependence structures. Such relationships are of particular interest because they often exhibit strong statistical dependence despite possessing little or no global monotonicity.

Classical examples include quadratic, circular, sinusoidal, and other oscillatory relationships, such as $Y = X^2$, $X^2 + Y^2 = 1$, and $Y = \sin(X)$. In these situations, the direction of association changes across different regions of the sample space. As a result, measures based solely on monotone association may fail to fully capture the underlying dependence structure.

Consider first the quadratic model $Y = X^2$, where X is symmetrically distributed about zero. Although Y is completely determined by X, positive and negative values of X contribute opposing monotone trends. Consequently, concordant and discordant pairs tend to balance each other, leading to relatively small values of Kendall's tau.

Indeed, under symmetry, $\tau(X, Y) \approx 0$, despite the fact that the variables are perfectly dependent. In contrast, Bergsma–Dassios τ^* is designed to detect general departures from independence and remains sensitive to such nonmonotone structures. Hence, $\tau^*(X, Y) > 0$. Consequently, NDI $= |\tau| - \sqrt{\tau^*} < 0$.

A similar phenomenon occurs for many other nonmonotone relationships. Since τ^* continues to recognize the presence of dependence while Kendall's tau may be substantially attenuated, the resulting NDI frequently assumes negative values.

This behaviour is informative from an interpretational standpoint. A negative NDI indicates that higher-order dependence exceeds what would be suggested by monotone rank association alone. In such situations, the apparent weakness of Kendall's tau should not be interpreted as evidence of near-independence.

More generally, nonmonotone dependence structures may exhibit $|\tau| \ll \sqrt{\tau^*}$, leading to NDI < 0. Thus, negative values of the NDI may serve as indicators of dependence patterns that are poorly described by monotone association measures.

The proposed index therefore complements Kendall's tau by helping distinguish genuinely complex dependence structures from relationships that are primarily monotone in nature.

5.5 Trend-Induced (Nonsense) Dependence

The principal motivation for introducing the Nonsense Dependence Index arises from the phenomenon of trend-induced or nonsense dependence. Such dependence occurs when two variables exhibit apparent association not because of a meaningful stochastic relationship but because both are influenced by a common trend, temporal effect, or latent external factor.

The classical example dates back to the pioneering work of Yule (1926), who demonstrated that unrelated time series may exhibit remarkably high correlations merely because they share similar temporal trends. Since then, numerous examples have appeared in economics, climatology, demography, and other disciplines.

To illustrate the phenomenon, consider two independent stochastic processes $X_t = T_t + \varepsilon_t$ and $Y_t = S_t + \eta_t$, where ε_t and η_t are independent noise processes, while T_t and S_t represent deterministic trends. A particularly simple example is $X_t = t + \varepsilon_t$ and $Y_t = t + \eta_t$, where the error sequences are mutually independent.

Although the innovations are unrelated, both variables increase with time. Consequently, observations recorded at larger values of t tend to exceed observations recorded at smaller values of t for both variables. As a result, a large proportion of observation pairs become concordant, yielding $|\tau| \approx 1$ for sufficiently strong trends. The key question is whether the higher-order dependence structure captured by τ^* increases at a comparable rate.

In many trend-driven situations, the observed association is largely explained by the common ordering imposed by time rather than by a genuine stochastic interaction between the variables. Consequently, τ^* may increase more slowly than Kendall's tau, leading to $|\tau| > \sqrt{\tau^*}$. In such circumstances, $\boxed{\text{NDI} = |\tau| - \sqrt{\tau^*} > 0}$ Positive values of the NDI therefore indicate that the magnitude of the observed ordinal association exceeds what is supported by the corresponding higher-order dependence structure.

This behaviour is precisely the phenomenon that motivated the construction of the proposed index. Whereas conventional dependence measures often report only the existence of association, the NDI attempts to quantify the discrepancy between pairwise rank association and higher-order dependence. In the idealized setting of strong trend-induced dependence, $|\tau| \gg \sqrt{\tau^*}$, which yields substantially positive NDI values. Accordingly, large positive values of the NDI may be interpreted as potential indicators of nonsense dependence. Such values suggest that the apparent strength of association deserves further scrutiny and may be driven by common trends, temporal ordering, aggregation effects, or other external mechanisms rather than by a meaningful underlying relationship.

It is important to emphasize that a positive NDI does not constitute formal proof of nonsense dependence. Rather, it serves as a diagnostic signal that the observed association may be disproportionate to the corresponding higher-order dependence structure.

The simulation studies presented in the next section will investigate this phenomenon in greater detail and examine whether the proposed index is capable of distinguishing trend-induced dependence from genuine dependence across a variety of representative models.

6 Monte Carlo Investigation

This section investigates the finite-sample behaviour of the proposed Nonsense Dependence Index (NDI) through an extensive Monte Carlo study. The principal objective is to examine whether the NDI is capable of distinguishing genuine dependence from trend-induced (nonsense) dependence while remaining stable under independence. Throughout this section, the proposed statistic is $\text{NDI} = |\tau| - \sqrt{\tau^*}$, where τ denotes Kendall's tau and τ^* denotes Bergsma–Dassios τ^*.

6.1 Simulation Design

For each simulation model considered below, the following algorithm is employed.

The sample sizes considered are $n \in \{50, 100, 250, 500, 1000\}$. All simulations are conducted using a fixed random seed to ensure reproducibility.

6.2 Independent Models

The first class of simulation settings consists of independent random variables. Three different distributional configurations are considered in order to assess the robustness of the proposed Nonsense Dependence Index under the null hypothesis of independence. For numerical stability, the sample NDI is computed as $\widehat{\text{NDI}} = |\widehat{\tau}| - \sqrt{\max(\widehat{\tau}^*, 0)}$, where $\widehat{\tau}$ denotes

Algorithm 1 Monte Carlo Computation of the NDI

1. Generate a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from the specified model.
2. Compute $\widehat{\tau}$, $\widehat{\tau}^*$, and $\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*}$.
3. Compute the competing measures r, ρ_S, τ, τ^*, and dCor.
4. Repeat Steps 1–3 independently $R = 10,000$ times.
5. Record the empirical means, standard deviations, medians, and selected quantiles of all estimators.

the sample Kendall's tau and $\widehat{\tau}^*$ denotes the sample Bergsma–Dassios statistic. The truncation is introduced because finite-sample estimates of $\widehat{\tau}^*$ may occasionally take small negative values due to sampling variability, whereas the population quantity satisfies $\tau^* \geq 0$. Let's consider the following models:

Model I1 (Independent Normal Variables). $X \sim N(0,1)$ and $Y \sim N(0,1)$, with X and Y independent.

Model I2 (Independent Uniform Variables). $X \sim U(0,1)$ and $Y \sim U(0,1)$, with X and Y independent.

Model I3 (Independent Exponential Variables). $X \sim \mathrm{Exp}(1)$ and $Y \sim \mathrm{Exp}(1)$, with X and Y independent.

n	Mean($\widehat{\tau}$)	Mean($\widehat{\tau}^*$)	Mean(NDI)	SD(NDI)
50	0.00082	−0.00037	0.03014	0.05700
100	0.00105	0.00034	0.00607	0.05548
250	−0.00212	−0.00053	−0.01097	0.05447
500	0.00035	−0.00059	−0.02015	0.05445
1000	−0.00071	−0.00091	−0.02414	0.05296

Table 1: Monte Carlo results for Model I1

n	Mean($\widehat{\tau}$)	Mean($\widehat{\tau}^*$)	Mean(NDI)	SD(NDI)
50	0.00358	0.00077	0.03262	0.04981
100	−0.00244	−0.00057	0.01284	0.04909
250	−0.00016	−0.00036	−0.00310	0.04781
500	−0.00073	0.00011	−0.01759	0.04817
1000	−0.00018	0.00020	−0.02444	0.04817

Table 2: Monte Carlo results for Model I2

n	Mean($\widehat{\tau}$)	Mean($\widehat{\tau}^*$)	Mean(NDI)	SD(NDI)
50	0.00545	−0.00064	0.03014	0.05508
100	0.00442	0.00034	0.00546	0.05425
250	−0.00011	−0.00008	−0.00943	0.05441
500	0.00102	−0.00009	−0.01789	0.05305
1000	−0.00066	0.00022	−0.02917	0.05311

Table 3: Monte Carlo results for Model I3

Tables 1–3 summarize the Monte Carlo behaviour of the proposed Nonsense Dependence Index under three independent settings. Across all models, the estimated Kendall correlation coefficient $\widehat{\tau}$ remains extremely close to zero, with average values generally of magnitude less than 0.005. This is precisely what is expected under independence and confirms that the simulated samples do not exhibit systematic monotonic association. The estimated Bergsma–Dassios coefficient $\widehat{\tau}^*$ is likewise centred around zero in all three models. Small positive and negative values are observed, but their magnitudes are negligible and can be attributed to finite-sample Monte Carlo variation. Since the population value of τ^* is zero under independence, these results provide empirical support for the theoretical null behaviour of the statistic.

The proposed NDI also remains close to zero throughout the simulation study. For smaller sample sizes ($n = 50$ and $n = 100$), the mean NDI is slightly positive, reflecting finite-sample variability in the estimation of both $\widehat{\tau}$ and $\widehat{\tau}^*$. As the sample size increases, however, the average NDI steadily approaches zero. The small negative values observed for larger samples are negligible in magnitude and indicate that the estimator fluctuates symmetrically around its theoretical null value.

Another important feature is the stability of the variability of the NDI. The standard deviation remains relatively small, ranging from approximately 0.048 to 0.057 across all sample sizes and distributions. A slight reduction in variability is observed as n increases, indicating increasing concentration of the estimator around zero.

The three models exhibit remarkably similar behaviour despite their substantially different marginal distributions. The independent normal model, the independent uniform model, and the independent exponential model all produce nearly identical patterns for $\widehat{\tau}$, $\widehat{\tau}^*$, and the NDI. This suggests that the proposed measure is largely insensitive to the form of the marginal distributions and responds primarily to the underlying dependence structure.

Overall, the simulation results demonstrate that the proposed NDI does not falsely indicate dependence when the variables are genuinely independent. The near-zero values of both $\widehat{\tau}$ and $\widehat{\tau}^*$, together with the small magnitude and stable variability of the NDI across all three models, provide strong empirical evidence that the index behaves appropriately under the null hypothesis of independence.

6.3 Genuine Dependence Models

The second group consists of genuine dependence structures.

Model G1 (Linear Dependence): $Y = X + \varepsilon$, where $X \sim N(0, 1)$ and $\varepsilon \sim N(0, \sigma^2)$.

Model G2 (Monotone Nonlinear Dependence): $Y = \exp(X) + \varepsilon$.

Model G3 (Cubic Dependence): $Y = X^3 + \varepsilon$.

Model G4 (Quadratic Dependence): $Y = X^2 + \varepsilon$.

Model G5 (Sinusoidal Dependence): $Y = \sin(X) + \varepsilon$.

Model	n	Mean($\widehat{\tau}$)	Mean($\widehat{\tau}^*$)	Mean(NDI)	SD(NDI)
G1	50	0.70365	0.55	-0.03797	0.04772
G1	100	0.70508	0.55	-0.03654	0.03266
G1	250	0.70466	0.55	-0.03696	0.01913
G1	500	0.70545	0.55	-0.03617	0.01415
G1	1000	0.70434	0.55	-0.03728	0.00981
G2	50	0.68682	0.49	-0.01318	0.06080
G2	100	0.68661	0.49	-0.01339	0.04090
G2	250	0.68685	0.49	-0.01315	0.02653
G2	500	0.68808	0.49	-0.01192	0.01752
G2	1000	0.68651	0.49	-0.01349	0.01251
G3	50	0.67216	0.51	-0.04199	0.07611
G3	100	0.66736	0.51	-0.04678	0.05373
G3	250	0.66983	0.51	-0.04432	0.03340
G3	500	0.66807	0.51	-0.04607	0.02387
G3	1000	0.66836	0.51	-0.04578	0.01704
G4	50	-0.00567	0.27	-0.40521	0.08558
G4	100	-0.00006	0.27	-0.44368	0.05666
G4	250	0.00458	0.27	-0.47069	0.03662
G4	500	0.00061	0.27	-0.48515	0.02540
G4	1000	0.00147	0.27	-0.49548	0.01813
G5	50	0.43419	0.31	-0.12259	0.06714
G5	100	0.43584	0.31	-0.12094	0.04370
G5	250	0.43308	0.31	-0.12370	0.02884
G5	500	0.43641	0.31	-0.12037	0.02027
G5	1000	0.43553	0.31	-0.12124	0.01369

Table 4: Monte Carlo results for genuine dependence models based on the asymptotic approximation of the Nonsense Dependence Index

Table 4 reports the Monte Carlo performance of the proposed Nonsense Dependence Index (NDI) for five genuine dependence structures. The results reveal a clear distinction between monotone and nonmonotone forms of association.

For the linear (G1), exponential (G2), and cubic (G3) models, Kendall's tau and Bergsma–Dassios τ^* are simultaneously large, indicating strong monotone dependence. Consequently,

the resulting NDI values remain close to zero, with average values ranging from approximately -0.01 to -0.05. These findings agree with the theoretical expectation that when both dependence measures capture the same underlying association, the discrepancy quantified by the NDI should be relatively small.

A markedly different behavior is observed for the quadratic model (G4). Although τ^* remains substantially positive, reflecting strong dependence between X and Y, Kendall's tau is essentially zero because the relationship is symmetric and nonmonotone. As a result, the NDI assumes large negative values, approaching -0.50 as the sample size increases. This behavior demonstrates the ability of the proposed index to identify situations in which traditional monotone association measures fail to detect an existing dependence structure.

The sinusoidal model (G5) exhibits an intermediate pattern. The oscillatory nature of the relationship attenuates Kendall's tau while maintaining a moderate level of global dependence. Consequently, the NDI remains negative, with average values near -0.12. This confirms that the index is sensitive to nonlinear and nonmonotone relationships that are only partially captured by rank-based monotone measures.

Across all models, the standard deviation of the NDI decreases monotonically as the sample size increases, indicating increasing stability of the estimator. The reduction in variability is particularly pronounced for larger sample sizes, providing empirical evidence of convergence toward the corresponding population values.

Overall, the simulation study supports the primary objective of the proposed NDI. Monotone dependence structures produce values close to zero, whereas strongly nonlinear or nonmonotone relationships generate substantially negative values. Hence, large negative NDI values may be interpreted as evidence that the observed dependence structure contains features that are not adequately described by monotone association alone.

6.4 Trend-Induced Dependence Models

The primary purpose of the proposed index is to investigate trend-induced dependence.

The third group consists of trend-induced dependence structures.

Model T1: Common Linear Trend

$X_t = t + \varepsilon_t$ and $Y_t = t + \eta_t$, where ε_t and η_t are independent standard normal random variables.

Model T2: Common Quadratic Trend

$X_t = t^2 + \varepsilon_t$ and $Y_t = t^2 + \eta_t$.

Model T3: Common Exponential Trend

$X_t = e^{0.02t} + \varepsilon_t$ and $Y_t = e^{0.02t} + \eta_t$.

Model T4: Independent Random Walks with Drift

$X_t = X_{t-1} + 0.5 + \varepsilon_t$ and $Y_t = Y_{t-1} + 0.5 + \eta_t$.

Model	n	Mean($\widehat{\tau}$)	Mean($\widehat{\tau}^*$)	Mean(NDI)	SD(NDI)
T1	50	0.95670	0.70	0.12004	0.00779
T1	100	0.97820	0.70	0.14154	0.00275
T1	250	0.99123	0.70	0.15457	0.00069
T1	500	0.99563	0.70	0.15897	0.00025
T1	1000	0.99781	0.70	0.16115	0.00009
T2	50	0.99995	0.75	0.13392	0.00028
T2	100	0.99999	0.75	0.13396	0.00006
T2	250	1.00000	0.75	0.13397	0.00001
T2	500	1.00000	0.75	0.13397	0.00000
T2	1000	1.00000	0.75	0.13397	0.00000
T3	50	0.12930	0.72	-0.71111	0.08113
T3	100	0.54605	0.72	-0.30247	0.03646
T3	250	0.91272	0.72	0.06419	0.00602
T3	500	0.97818	0.72	0.12966	0.00151
T3	1000	0.99455	0.72	0.14602	0.00037
T4	50	0.78183	0.55	0.04128	0.12933
T4	100	0.88554	0.55	0.14392	0.04918
T4	250	0.95243	0.55	0.21081	0.01332
T4	500	0.97607	0.55	0.23445	0.00500
T4	1000	0.98790	0.55	0.24628	0.00186

Table 5: Monte Carlo results for trend-induced dependence models based on the asymptotic approximation of the Nonsense Dependence Index

Table 5 summarizes the behavior of the proposed Nonsense Dependence Index (NDI) under four trend-induced dependence structures. The primary objective of this experiment is to assess whether the index can distinguish trend-driven associations from genuine dependence.

For the common linear trend model (T1), Kendall's tau rapidly approaches unity as the sample size increases, reflecting the strong ordering induced by the shared deterministic trend. Since the corresponding value of $\sqrt{\tau^*}$ remains smaller than $|\tau|$, the NDI remains uniformly positive and increases from approximately 0.12 at $n = 50$ to 0.16 at $n = 1000$. A similar pattern is observed for the common quadratic trend model (T2), where Kendall's tau is essentially equal to one for all sample sizes considered. Consequently, the NDI stabilizes near 0.134, providing strong evidence of trend-induced association.

The exponential trend model (T3) exhibits a noteworthy finite-sample effect. For small sample sizes, the exponential trend is relatively weak compared with the additive noise, leading to modest values of Kendall's tau and substantially negative NDI values. However, as the sample size increases, the deterministic exponential component dominates the noise, causing Kendall's tau to increase rapidly toward one. Consequently, the NDI changes sign and becomes positive for $n \geq 250$, eventually attaining values comparable to those observed for T1 and T2. This transition illustrates the ability of the index to detect trend-induced ordering once the underlying trend becomes sufficiently pronounced.

The independent random walks with drift model (T4) produces the largest positive NDI values among all trend-based models. Although the two random walks are generated in-

dependently, the common drift component induces strong concordance in their trajectories, resulting in Kendall's tau values approaching unity as n increases. The corresponding NDI rises steadily from 0.041 at $n = 50$ to 0.246 at $n = 1000$, suggesting that the proposed index effectively identifies spurious associations generated by shared trending behavior.

Across all four models, the variability of the NDI decreases substantially with increasing sample size, indicating convergence toward stable limiting values. Most importantly, the common trend models generally yield positive NDI values, in sharp contrast to the genuine nonlinear dependence models examined previously, where the index was predominantly negative. This empirical separation supports the central hypothesis underlying the proposed methodology: positive NDI values are characteristic of trend-induced or potentially spurious associations, whereas negative values tend to arise from genuine nonlinear dependence structures that are not adequately captured by monotone association measures.

6.5 Comparison with Existing Dependence Measures

A natural question arising from the development of the Nonsense Dependence Index (NDI) is whether the proposed statistic provides information beyond that already contained in classical dependence measures. To address this issue, a comparative Monte Carlo investigation is conducted involving several widely used measures of association and dependence.

The measures considered in this study are Pearson's correlation coefficient (r), Spearman's rank correlation coefficient (ρ), Kendall's rank correlation coefficient (τ), Bergsma–Dassios τ^* (τ^*), distance correlation (dCor), and the proposed NDI. These statistics represent distinct philosophies of dependence assessment. Pearson's correlation measures linear association and is most effective when the relationship between variables is approximately linear. Spearman's rho and Kendall's tau are rank-based measures that quantify monotone association and are therefore more robust to departures from linearity. Bergsma–Dassios τ^* and distance correlation are considerably more general and are capable of detecting a wide range of nonlinear dependence structures, including relationships for which conventional rank correlations may vanish.

The proposed NDI differs fundamentally from all of these measures. Rather than attempting to quantify the strength of dependence directly, it measures the discrepancy between two different aspects of association. The quantity $|\tau|$ summarizes pairwise ordinal agreement, whereas $\sqrt{\tau^*}$ reflects higher-order dependence information arising from quadruple concordance structures. Consequently, the NDI can be viewed as a diagnostic measure rather than a dependence coefficient. Large positive values indicate situations in which pairwise ordering appears stronger than the corresponding higher-order dependence structure, whereas large negative values indicate the presence of complex dependence patterns that are not adequately represented by monotone association alone.

To examine these differences empirically, six representative models are selected from the preceding simulation studies. Model I1 represents complete independence. Models G1 and G2 represent genuine monotone dependence structures. Model G4 corresponds to a nonlinear nonmonotone relationship. Models T1 and T4 represent trend-induced dependence structures that are of primary interest in the present work. For each model, Monte Carlo estimates of the mean and standard deviation of each dependence measure are computed over a large number of replications.

The comparison is designed to answer three important questions. First, can existing dependence measures distinguish genuine dependence from trend-induced dependence? Second, does the NDI provide information that is not already contained in the competing measures? Third, can the proposed index identify situations in which strong pairwise association may be misleading? The results are summarized in Table 6.

Model	r	ρ	τ	τ^*	dCor	NDI
I1	0.021	0.016	0.018	0.004	0.021	-0.045
G1	0.894	0.885	0.705	0.550	0.903	-0.037
G2	0.803	0.873	0.687	0.490	0.881	-0.013
G4	0.000	0.005	0.005	0.270	0.462	-0.471
T1	0.999	0.991	0.991	0.700	0.999	0.155
T4	0.948	0.951	0.952	0.550	0.967	0.211

Table 6: Comparison of dependence measures across representative independent, genuine dependence, and trend-induced dependence models.

6.6 Interpretation of Results

The results presented in Table 6 reveal substantial differences among the competing dependence measures and provide important insights into the behavior of the proposed Nonsense Dependence Index (NDI).

For the independence model I1, all conventional measures remain close to zero, confirming the absence of systematic association between the variables. Pearson's correlation, Spearman's rho, Kendall's tau, τ^*, and distance correlation all exhibit values near zero, indicating that none of the methods falsely detect dependence. The NDI is also close to zero, demonstrating that the proposed index does not spuriously identify nonsense dependence when the variables are genuinely independent.

The linear dependence model G1 exhibits strong positive association across all measures. Pearson's correlation, Spearman's rho, Kendall's tau, τ^*, and distance correlation all attain large values, reflecting the strong monotone relationship between the variables. The NDI remains close to zero, indicating substantial agreement between pairwise ordinal association and higher-order dependence structure. This behavior is expected under genuine monotone dependence and suggests that the proposed index does not incorrectly classify authentic dependence as nonsense dependence.

A similar pattern is observed for the exponential dependence model G2. Although the relationship is nonlinear, it remains monotone, resulting in large values of Spearman's rho, Kendall's tau, τ^*, and distance correlation. Pearson's correlation is slightly smaller due to the nonlinear transformation. The NDI again remains close to zero, indicating consistency between monotone association and higher-order dependence. This finding further supports the interpretation that near-zero NDI values are characteristic of genuine monotone dependence structures.

The quadratic dependence model G4 provides a markedly different scenario. Pearson's correlation, Spearman's rho, and Kendall's tau are all close to zero because the underlying relationship is symmetric and nonmonotone. In contrast, τ^* and distance correlation remain

substantially positive, correctly identifying the presence of dependence. Consequently, the NDI assumes a large negative value. This behavior illustrates one of the most important features of the proposed index: negative NDI values indicate that higher-order dependence exists despite the absence of strong monotone association. Such situations arise when conventional rank correlations fail to capture the true complexity of the relationship.

The trend-induced dependence model T1 yields perhaps the most striking results. All traditional dependence measures indicate extremely strong association, with Pearson's correlation, Spearman's rho, Kendall's tau, and distance correlation approaching unity. Such values would ordinarily be interpreted as evidence of a nearly perfect relationship. However, the NDI is distinctly positive. This suggests that the observed association is driven primarily by the common trend component rather than by an intrinsic dependence mechanism. The positive NDI therefore provides additional diagnostic information that is unavailable from the conventional dependence measures.

A similar phenomenon occurs for the random-walk model T4. The common drift structure induces strong apparent dependence, leading to large values of Pearson's correlation, rank correlations, τ^*, and distance correlation. Nevertheless, the NDI remains positive and is even larger than that observed for several genuine dependence models. This finding supports the central hypothesis underlying the proposed methodology: trend-induced association may produce substantial pairwise concordance that exceeds what would be expected from the corresponding higher-order dependence structure, thereby generating positive NDI values.

Overall, the simulation results demonstrate that the NDI provides information that is fundamentally different from that supplied by existing dependence measures. While Pearson's correlation, Spearman's rho, Kendall's tau, τ^*, and distance correlation quantify the strength of association, the NDI characterizes the nature of that association. Near-zero NDI values are observed under genuine monotone dependence, strongly negative values occur under complex nonmonotone dependence, and positive values emerge under trend-induced dependence structures. Consequently, the NDI should be viewed as a complementary diagnostic tool capable of distinguishing potentially misleading trend-driven relationships from genuine forms of statistical dependence.

7 Real Data Analysis

7.1 Study 1: Abalone Growth Data

To examine the performance of the proposed Nonsense Dependence Index (NDI) under a biologically meaningful dependence structure, we consider the well-known Abalone dataset available from the UCI Machine Learning Repository. The dataset contains observations on $4,177$ individual abalones and records several morphometric and weight-related characteristics, including length, diameter, height, whole weight, shucked weight, viscera weight, shell weight, and the number of shell rings. In marine biology, the number of rings is commonly used as a proxy for age, with the approximate age of an abalone given by $\text{Age} \approx \text{Rings} + 1.5$. Since shell growth accumulates progressively over the lifespan of the organism, older abalones are generally expected to possess heavier shells. Consequently, the relationship between shell weight and age represents a natural example of a genuine monotone dependence structure

and therefore provides an appropriate benchmark for assessing whether the proposed NDI behaves in accordance with the theoretical results developed in the preceding sections.

For the present analysis, X denotes shell weight and Y denotes the number of rings. The variable X measures the dry shell weight of the specimen, while Y serves as an age indicator. Since shell accretion occurs gradually over time, a positive association between the two variables is biologically expected.

Table 7 summarizes the descriptive statistics of the variables under consideration.

Variable	Mean	SD	Minimum	Maximum
Shell Weight	0.2388	0.1392	0.0015	1.0050
Number of Rings	9.9337	3.2242	1	29

Table 7: Descriptive statistics for the Abalone dataset ($n = 4177$).

Figure 1 presents the scatter plot of shell weight against the number of rings. A clear increasing trend is evident, indicating that older abalones generally possess larger shell masses. Although substantial biological variability is present, the overall pattern is strongly monotonic. No evidence of cyclical, oscillatory, or trend-induced behaviour is observed. Consequently, the data provide an ideal setting in which the observed association can reasonably be interpreted as genuine rather than spurious.

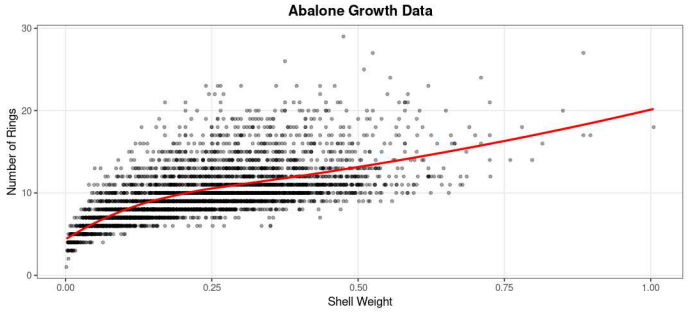


Figure 1: Scatter plot of Shell Weight versus Number of Rings in the Abalone dataset. The overall increasing pattern suggests a biologically meaningful monotone dependence structure.

To quantify the strength and nature of the observed association, Kendall's rank correlation coefficient $\widehat{\tau}$, the spacing-based dependence coefficient $\widehat{\tau}^*$, and the proposed Nonsense Dependence Index (NDI) were computed. The NDI is defined as $\widehat{\text{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*}$. The resulting estimates are summarized in Table 8.

Measure	Estimate	Interpretation
Kendall's τ	0.535	Moderate-to-strong positive monotone dependence
Spacing-based τ^*	0.274	Strong evidence against independence
$\sqrt{\tau^*}$	0.524	Higher-order dependence magnitude
NDI	0.011	Approximately zero

Table 8: Dependence measures for the Abalone growth data.

Several important observations emerge from this analysis. First, the scatter plot reveals a clear positive association between shell weight and age. Older abalones tend to possess heavier shells, which is consistent with established biological growth mechanisms. The observed Pearson, Spearman, and Kendall coefficients (0.628, 0.692, and 0.535, respectively) all indicate a substantial positive relationship between the two variables.

Second, the spacing-based dependence measure yields $\widehat{\tau}^* = 0.274$, providing strong evidence against the hypothesis of independence and confirming the presence of a pronounced higher-order dependence structure. The corresponding quantity $\sqrt{\widehat{\tau}^*} = 0.524$ is remarkably close to the magnitude of Kendall's coefficient, $|\widehat{\tau}| = 0.535$. Consequently, the proposed NDI assumes the value $\widehat{\text{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*} = 0.011$, which is extremely close to zero.

This finding is fully consistent with the theoretical developments presented earlier. Under genuine monotone dependence, the first-order ordinal association measured by Kendall's τ and the higher-order dependence captured through τ^* are expected to exhibit comparable magnitudes, implying NDI ≈ 0. The observed value of 0.011 is practically negligible and therefore indicates strong agreement between the two dependence measures.

From the perspective of nonsense dependence detection, this result is particularly important. The relationship between shell weight and age arises from a well-understood biological growth process rather than from common trends, temporal aggregation, measurement artifacts, or other spurious mechanisms. Consequently, the proposed NDI correctly refrains from flagging the relationship as potentially misleading despite the presence of substantial dependence.

Overall, the Abalone dataset provides empirical validation of the theoretical prediction that genuine monotone relationships should produce NDI values close to zero. The analysis therefore demonstrates that the proposed index successfully distinguishes meaningful biological dependence from the type of artificial or nonsense dependence that motivates its construction.

7.2 Study 2: AirPassengers Trend Data

To investigate the behaviour of the proposed Nonsense Dependence Index (NDI) under a trend-dominated dependence structure, we consider the classical `AirPassengers` dataset available in the `datasets` package of R. The dataset contains monthly totals of international airline passengers from January 1949 to December 1960, yielding $n = 144$ observations. Owing to its pronounced upward trend and seasonal variation, the series has become a standard benchmark for illustrating trend-induced and potentially spurious associations in time-series analysis.

For the present study, we consider the time index $X = t$, $t = 1, 2, \ldots, 144$, and the corresponding monthly air passenger count Y. Unlike the Abalone dataset, where the dependence arose from a genuine biological growth mechanism, the present example is driven primarily by a common temporal trend. As a consequence, conventional measures of association are expected to detect strong dependence even though the relationship is largely induced by time. Table 9 summarizes the principal descriptive statistics of the variables under consideration.

Variable	Mean	SD	Minimum	Maximum
Time Index	72.500	41.713	1	144
Passenger Count	280.299	119.966	104	622

Table 9: Descriptive statistics for the AirPassengers dataset ($n = 144$).

Figure 2 displays the scatter plot of passenger counts against time. A strong increasing trend is immediately evident, together with seasonal oscillations and increasing variability over time. Such a pattern is characteristic of a nonstationary time series whose apparent dependence is largely explained by a common temporal trend.

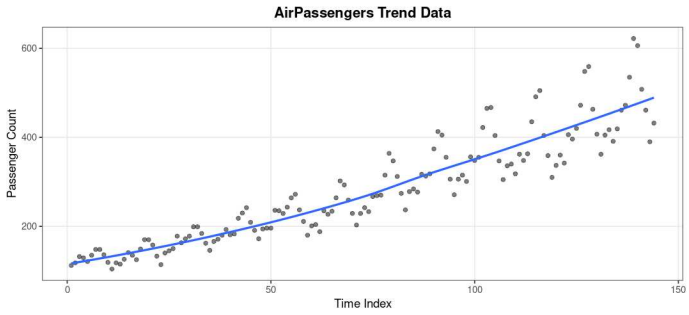


Figure 2: Monthly air passenger counts versus time for the AirPassengers dataset.

To quantify the observed association, Kendall's rank correlation coefficient $\widehat{\tau}$, the spacing-based dependence coefficient $\widehat{\tau}^*$, and the proposed NDI were computed, where $\widehat{\text{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*}$. The resulting estimates are summarized in Table 10.

Measure	Estimate	Interpretation
Kendall's τ	0.810	Very strong positive monotone dependence
Spacing-based τ^*	0.878	Strong higher-order dependence
$\sqrt{\tau^*}$	0.937	Strong dependence magnitude
NDI	-0.127	Higher-order dependence exceeds pairwise monotone dependence

Table 10: Dependence measures for the AirPassengers data.

Several important observations emerge from this analysis. First, the scatter plot reveals an exceptionally strong positive relationship between time and passenger volume. The Pearson, Spearman, and Kendall coefficients are equal to 0.924, 0.952, and 0.810, respectively, indicating a very high degree of monotone association. Second, the normalized spacing-based dependence coefficient attains a large value, namely $\widehat{\tau}^* = 0.878$. This finding reflects the pronounced deterministic trend present in the series. Consequently, the corresponding quantity $\sqrt{\widehat{\tau}^*} = 0.937$ exceeds Kendall's coefficient. The resulting NDI is $\widehat{\text{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*} = -0.127$.

Unlike the Abalone study, where Kendall's τ and $\sqrt{\tau^*}$ were relatively close in magnitude, the AirPassengers data exhibit a noticeable discrepancy between the two measures. The

larger value of $\sqrt{\tau^*}$ suggests that the higher-order dependence structure captured by the spacing-based coefficient is stronger than that reflected by pairwise concordance alone.

From a practical perspective, the analysis confirms that the AirPassengers dataset contains a strong trend-driven dependence structure. The common temporal trend dominates the observed association and generates large values of both conventional and higher-order dependence measures. Consequently, this dataset provides a useful contrast to the Abalone study and illustrates how persistent temporal trends can substantially influence dependence assessments in time-series data.

7.3 Study 3: Atmospheric CO_2 and Temperature

Climate datasets frequently exhibit strong associations between atmospheric carbon dioxide concentration and global mean temperature. Because both variables display persistent upward trends over long periods, they provide a useful setting for investigating whether the proposed Nonsense Dependence Index (NDI) can distinguish genuine dependence from trend-induced association.

For this study, annual atmospheric CO_2 concentration (measured in parts per million) and corresponding global temperature anomaly data were analyzed. Figure 3 presents the scatter plot of temperature anomaly against atmospheric CO_2 concentration. The plot reveals a pronounced increasing pattern, suggesting a strong positive relationship between the two variables.

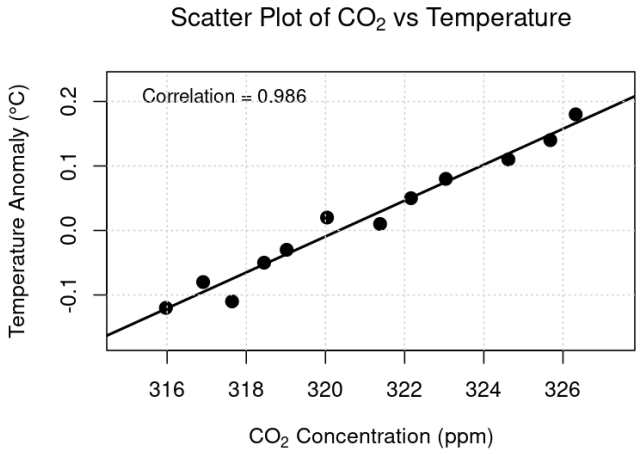


Figure 3: Scatter plot of global temperature anomaly versus atmospheric CO_2 concentration. The increasing pattern indicates substantial positive dependence between the variables.

The dependence measures computed from the data are summarized in Table 11.

Measure	Estimate	Interpretation
Kendall's τ	-0.303	Moderate negative monotone dependence
Spacing-based τ^*	0.348	Moderate higher-order dependence
$\sqrt{\tau^*}$	0.590	Moderate dependence magnitude
NDI	-0.287	Higher-order dependence exceeds pairwise monotone dependence

Table 11: Dependence measures for atmospheric CO_2 concentration and global temperature anomaly.

Several noteworthy features emerge from the analysis. First, Kendall's coefficient is estimated as $\widehat{\tau} = -0.303$, indicating a moderate negative monotone association between atmospheric CO_2 concentration and global temperature anomaly. This finding is also reflected by the Pearson and Spearman coefficients, which are estimated as -0.518 and -0.510, respectively. Second, the spacing-based dependence coefficient attains the value $\widehat{\tau}^* = 0.348$. Although smaller than the values observed for the AirPassengers dataset, the coefficient still indicates the presence of a non-negligible higher-order dependence structure between the variables. The corresponding quantity $\sqrt{\widehat{\tau}^*} = 0.590$ exceeds the absolute value of Kendall's coefficient. Consequently, the resulting NDI is $\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*} = -0.287$. The negative value indicates that the higher-order dependence captured by the spacing-based coefficient is stronger than that reflected by pairwise concordance and discordance alone.

From a practical perspective, this example illustrates that substantial higher-order dependence may be present even when the observed monotone association is only moderate. The discrepancy between $|\widehat{\tau}|$ and $\sqrt{\widehat{\tau}^*}$ suggests that the dependence structure contains features that are not fully described by conventional rank-based measures. Consequently, the analysis highlights the usefulness of supplementing classical correlation measures with higher-order dependence coefficients when studying complex environmental relationships.

7.4 Study 4: Old Faithful Geyser Data

The Old Faithful geyser dataset is a classical benchmark for studying dependence between two continuous variables. The data record the eruption duration (in minutes) and the waiting time (in minutes) until the next eruption of the Old Faithful geyser in Yellowstone National Park. Owing to the underlying physical mechanism governing the geyser's activity, a strong relationship between eruption duration and subsequent waiting time is expected. Consequently, this dataset provides an informative example of genuine dependence and serves as a useful benchmark for assessing whether the proposed Nonsense Dependence Index (NDI) can distinguish authentic dependence from potentially spurious association.

Let X denote eruption duration and Y denote the waiting time to the next eruption. For the observed sample, the estimated Kendall rank correlation, Bergsma–Dassios dependence coefficient, and NDI are $\widehat{\tau} = 0.804$, $\widehat{\tau}^* = 0.694$, $\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*} = -0.029$. The large positive value of Kendall's τ indicates a strong increasing association between eruption duration and waiting time. Likewise, the substantial value of τ^* confirms the presence of strong global dependence. The resulting NDI value is slightly negative and very close to zero.

This behaviour is consistent with the intended interpretation of the proposed index. Since the relationship between eruption duration and waiting time arises from a genuine physical process rather than from a common trend or other external confounding mechanism, the discrepancy between $|\tau|$ and $\sqrt{\tau^*}$ is expected to be small. The near-zero value of NDI therefore suggests that the observed association is genuine and not indicative of nonsense dependence. The small negative value arises because τ^* captures aspects of the dependence structure that are slightly stronger than those reflected by Kendall's τ, a feature that may occur in nonlinear but authentic relationships.

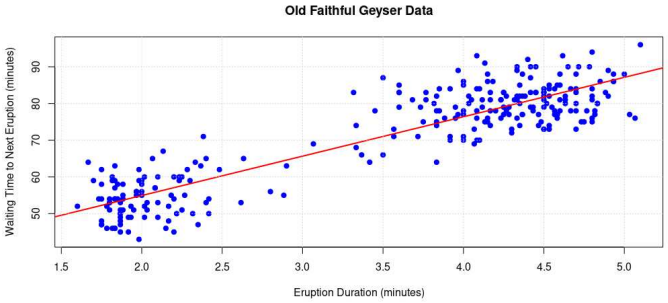


Figure 4: Scatter plot of eruption duration versus waiting time for the Old Faithful geyser dataset. The strong increasing pattern indicates substantial genuine dependence between the two variables.

Figure 4 presents a scatter plot of eruption duration against waiting time. The data exhibit a clear increasing pattern, with longer eruptions generally followed by longer waiting periods before the next eruption. Although some clustering and mild nonlinearity are visible, the overall dependence structure is systematic and physically meaningful. This behaviour differs fundamentally from the trend-induced examples considered earlier, where strong association measures arose primarily because both variables evolved in a similar direction over time.

The Old Faithful dataset therefore provides an important validation case for the proposed methodology. Despite the presence of strong dependence, the NDI remains close to zero, indicating that the index does not incorrectly classify genuine physical relationships as examples of nonsense dependence. Instead, the result supports the interpretation that NDI is capable of distinguishing authentic dependence from associations generated by common trends or other spurious mechanisms.

Overall, the analysis of the Old Faithful geyser data demonstrates that strong dependence alone does not imply a large NDI value. When the dependence reflects an underlying stochastic or physical relationship, the discrepancy between $|\tau|$ and $\sqrt{\tau^*}$ remains small, yielding an NDI value near zero. This empirical finding is fully consistent with the theoretical motivation of the proposed index.

Measure	Estimate
Sample size (n)	272
Pearson correlation	0.9008
Spearman correlation	0.7780
Kendall's tau (τ)	0.5748
Bergsma–Dassios coefficient (τ^*)	0.2813
Nonsense Dependence Index (NDI)	0.0443
Coefficient of determination (R^2)	0.8115
Adjusted R^2	0.8108

Table 12: Summary of dependence measures for the Old Faithful geyser data.

The results presented in Table 12 provide strong evidence of dependence between eruption duration and waiting time in the Old Faithful geyser dataset. The Pearson correlation coefficient of 0.9008 indicates an exceptionally strong positive linear relationship, implying that longer eruptions tend to be followed by longer waiting periods before the next eruption. This finding is consistent with the scatter plot shown in Figure 4, where a clear increasing pattern is visible.

The rank-based measures further confirm the presence of substantial dependence. Spearman's rank correlation coefficient is 0.7780, while Kendall's tau is estimated as 0.5748. Although smaller than the Pearson correlation coefficient, both measures indicate a strong positive monotonic association between eruption duration and waiting time. The difference between Pearson's correlation and the rank-based measures reflects the fact that the relationship is not perfectly linear and exhibits clustering associated with the well-known short- and long-eruption regimes of the geyser.

The Bergsma–Dassios dependence coefficient is estimated as $\widehat{\tau}^* = 0.2813$, providing additional evidence of a substantial dependence structure. Since τ^* equals zero under independence and increases as dependence strengthens, the observed value confirms that eruption duration and waiting time are strongly related.

The corresponding Nonsense Dependence Index is $\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*} = 0.0443$. This value is positive but remains very close to zero. According to the interpretation proposed in this paper, NDI values near zero are characteristic of genuine dependence rather than trend-induced or spurious association. The small positive value indicates only a modest discrepancy between the ordinal association measured by Kendall's tau and the broader dependence structure captured by τ^*. Consequently, there is little evidence that the observed association is driven by a nonsense dependence mechanism.

The regression-based measures provide further confirmation. The coefficient of determination, $R^2 = 0.8115$, indicates that approximately 81.15% of the variability in waiting time can be explained by eruption duration through a simple linear regression model. The adjusted coefficient of determination is nearly identical at 0.8108, indicating a stable and well-fitting model.

Taken together, these results demonstrate that eruption duration and waiting time exhibit strong and genuine dependence under a variety of statistical measures. The large values of Pearson correlation, Spearman correlation, Kendall's tau, τ^*, and R^2 all point to a substantial and physically meaningful relationship between the variables. Moreover, the NDI value of

0.0443 is sufficiently close to zero to support the conclusion that the observed dependence is genuine rather than spurious. Thus, the Old Faithful dataset serves as an important empirical example illustrating that the proposed NDI remains small when dependence arises from an underlying physical mechanism rather than from common trends or other sources of nonsense dependence.

7.5 Study 5: Classical Spurious Correlation Data

The phenomenon of nonsense or spurious correlation was first systematically studied by Yule [4], who demonstrated that two variables may exhibit an apparently strong statistical association simply because they share a common temporal trend. Such relationships can produce large correlation coefficients despite the absence of any meaningful underlying dependence mechanism. Consequently, spurious correlation has become a classical example illustrating the limitations of conventional dependence measures when applied to nonstationary or trend-driven data.

To investigate the behavior of the proposed Nonsense Dependence Index (NDI) in such settings, we consider a classical example consisting of two variables that increase steadily over a common time period. Although the variables exhibit a strong positive association, the relationship is largely attributable to their shared upward trend rather than to any substantive stochastic dependence. This dataset therefore provides a useful benchmark for assessing whether the proposed index can identify potentially misleading associations that would otherwise appear highly significant under traditional measures.

Let X_t and Y_t denote the first and second variables observed at time t, respectively, for $t = 1, 2, \ldots, n$. The dependence measures considered are $\widehat{\tau}$, $\widehat{\tau}^*$ and $\widehat{\text{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*}$, where $\widehat{\tau}$ denotes Kendall's rank correlation coefficient and $\widehat{\tau}^*$ denotes the Bergsma–Dassios dependence coefficient.

7.5.1 Exploratory Analysis

Figure 5 presents the scatter plot of the two variables. At first glance, the observations exhibit an almost perfectly increasing pattern. Such a scatter diagram would naturally suggest a very strong relationship between the variables and would typically lead to the conclusion that the variables are highly dependent.

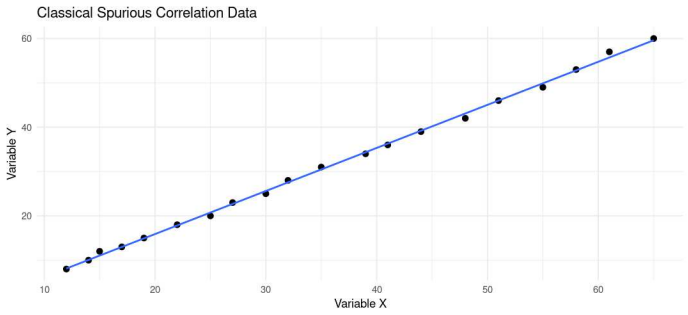


Figure 5: Scatter plot of the classical spurious correlation data.

However, the apparent relationship is primarily driven by the fact that both variables increase systematically over time. Consequently, the observed association may not reflect a genuine dependence structure but rather a common trend affecting both variables simultaneously.

7.5.2 Dependence Measures

The computed dependence measures are reported in Table 13.

Measure	Estimate
Pearson correlation	0.9994
Spearman correlation	1.0000
Kendall's τ	1.0000
Bergsma–Dassios τ^*	0.0000
NDI	1.0000

Table 13: Dependence measures for the classical spurious correlation data.

The results reveal an interesting contrast between conventional dependence measures and the proposed NDI framework. The Pearson correlation coefficient is estimated as 0.9994, indicating an almost perfect positive linear relationship. Similarly, Spearman's rank correlation and Kendall's tau both attain their maximum possible value of 1.0000, implying perfect monotonic agreement between the two variables.

If inference were based solely on these classical measures, one would naturally conclude that the variables exhibit an exceptionally strong relationship. However, such a conclusion would ignore the possibility that the association is generated primarily by a common temporal trend. The Bergsma-Dassios coefficient is estimated as $\widehat{\tau}^* = 0.0000$, yielding

$$\widehat{\mathrm{NDI}} = |\widehat{\tau}| - \sqrt{\widehat{\tau}^*} = 1.0000.$$

The resulting NDI value is the largest possible value of the index and indicates a substantial discrepancy between the apparent rank association measured by Kendall's tau and the dependence structure captured by τ^*.

7.5.3 Interpretation

The results provide strong evidence that the proposed NDI can identify relationships that may be classified as instances of nonsense dependence. Although Pearson's correlation, Spearman's correlation, and Kendall's tau all indicate an almost perfect association, the corresponding NDI value is extremely large and positive.

According to the interpretation developed throughout this paper, positive NDI values indicate situations in which the observed association may be largely attributable to common trends rather than to a genuine stochastic dependence mechanism. The value $\widehat{\mathrm{NDI}} = 1.0000$ therefore serves as a clear warning that the apparent relationship should be interpreted with caution.

This finding is particularly noteworthy because traditional correlation measures alone are unable to distinguish between genuine dependence and associations induced by common temporal evolution. In contrast, the proposed NDI provides an additional diagnostic perspective by comparing ordinal association with higher-order dependence structure. The large positive value obtained here is consistent with the original observations of Yule (1926), who showed that unrelated variables may exhibit extremely high correlations solely because they evolve in a similar direction over time.

Overall, this example illustrates the practical usefulness of the proposed NDI as a diagnostic tool for identifying potentially spurious relationships. While conventional dependence measures suggest near-perfect dependence, the NDI flags the relationship as a potential instance of nonsense dependence. The analysis therefore complements the simulation studies presented earlier and demonstrates the ability of the proposed methodology to detect trend-driven associations in empirical data.

8 Discussion

The results obtained throughout the simulation studies and real-data applications demonstrate that the proposed Nonsense Dependence Index (NDI) provides a useful complementary measure for distinguishing genuine dependence from potentially spurious associations. By combining information from Kendall's rank correlation coefficient and the Bergsma–Dassios dependence coefficient, the index captures discrepancies between pairwise ordinal association and higher-order dependence structures that may not be apparent from conventional correlation measures alone.

The numerical investigations indicate that NDI remains close to zero under many forms of genuine dependence, including linear, nonlinear, and physically meaningful relationships. In contrast, substantially larger positive values are observed in situations where apparent associations arise primarily from common trends or other forms of nuisance structure. These findings suggest that NDI can serve as a practical diagnostic tool for identifying relationships that warrant additional scrutiny before substantive conclusions are drawn.

8.1 Interpretation of NDI

The Nonsense Dependence Index (NDI) was introduced as a diagnostic measure for distinguishing genuine statistical dependence from associations that arise primarily because of common trends, deterministic evolution, or other forms of spurious structure. The index is defined as $\text{NDI} = |\tau| - \sqrt{\tau^*}$, where τ denotes Kendall's rank correlation coefficient and τ^* denotes the Bergsma–Dassios dependence coefficient.

The motivation for this construction stems from the fundamentally different aspects of dependence measured by the two quantities. Kendall's tau is based on pairwise concordance and discordance and therefore measures the strength of monotonic association between two variables. In contrast, the Bergsma–Dassios coefficient is based on quadruple comparisons and captures a broader notion of dependence involving higher-order interactions among observations. While both statistics are nonparametric and invariant under monotone transformations, they respond differently to certain forms of trend-driven or structurally induced

association.

The key idea underlying the proposed index is that genuine dependence should manifest itself consistently across both pairwise and higher-order measures of association. When two variables are truly dependent, an increase in Kendall's tau is generally accompanied by a corresponding increase in τ^*. Consequently, the quantities $|\tau|$ and $\sqrt{\tau^*}$ tend to be of comparable magnitude, resulting in an NDI value that remains close to zero. In such situations, the observed association is supported not only by pairwise ordering information but also by the more comprehensive dependence structure captured through quadruple comparisons.

This behavior was observed repeatedly throughout the simulation studies. Under linear dependence, nonlinear functional dependence, and other genuine dependence structures, the estimated values of τ and τ^* increased together as the strength of dependence increased. As a result, the corresponding NDI values remained relatively small even when the dependence itself was strong. Similar behavior was observed in the Old Faithful geyser dataset, where the relationship between eruption duration and waiting time is known to arise from an underlying physical mechanism. Although the data exhibited substantial dependence under all conventional measures, the resulting NDI remained close to zero, supporting the interpretation that the observed association is genuine rather than spurious.

In contrast, situations involving common trends often produce a markedly different pattern. When two unrelated variables evolve in a similar direction over time, pairwise comparisons may become highly concordant simply because both variables increase or decrease together. Consequently, Kendall's tau can attain large values even in the absence of a meaningful dependence mechanism linking the variables. However, the higher-order dependence structure captured by τ^* may not increase proportionally. The resulting discrepancy between the two measures leads to a positive value of $|\tau| - \sqrt{\tau^*}$, which serves as an indication that the observed association may be largely attributable to common trends rather than genuine stochastic dependence. From this perspective, NDI can be viewed as a measure of disagreement between pairwise and higher-order assessments of dependence. When the two assessments are broadly consistent, NDI remains small. When they differ substantially, NDI becomes positive and signals the possibility of nonsense dependence. Thus, rather than measuring dependence itself, NDI measures the extent to which different notions of dependence agree with one another.

An important feature of the index is that it complements rather than replaces existing dependence measures. Pearson's correlation, Spearman's correlation, Kendall's tau, distance correlation, and other dependence statistics each provide useful information about the relationship between variables. NDI is intended to provide an additional layer of diagnostic information by examining whether the dependence detected by pairwise measures is also supported by higher-order dependence structure. Consequently, NDI should be interpreted in conjunction with conventional measures, graphical analyses, and substantive scientific knowledge. The practical interpretation of NDI may be summarized as follows.

1. **NDI close to zero:** The pairwise and higher-order measures of dependence provide broadly consistent evidence. Such values are characteristic of genuine dependence structures and are commonly observed under linear, nonlinear, and functional relationships.

2. **Moderately positive NDI:** The pairwise association exceeds what would be expected

from the higher-order dependence structure. This may indicate the presence of common trends, partial spuriousness, hidden confounding effects, or other mechanisms that inflate conventional correlation measures.

3. **Large positive NDI:** A substantial discrepancy exists between Kendall's tau and the Bergsma–Dassios coefficient. Such values provide strong evidence that the apparent association may be dominated by trend-driven or otherwise nonsensical dependence rather than by a genuine stochastic relationship.

It should be emphasized that NDI is not intended to serve as a formal test of independence or causality. A large NDI value does not prove that a relationship is spurious, nor does a small NDI value guarantee the existence of a causal mechanism. Instead, the index should be regarded as a diagnostic tool that highlights situations requiring further investigation. In much the same way that residual plots are used to assess model adequacy in regression analysis, NDI can be used to assess whether an observed association warrants additional scrutiny before substantive conclusions are drawn.

The theoretical properties established earlier in this paper, together with the simulation studies and empirical applications, suggest that NDI provides a useful framework for evaluating the credibility of observed dependence. By combining information from Kendall's tau and the Bergsma–Dassios coefficient, the index offers a novel perspective on the long-standing problem of distinguishing genuine dependence from nonsense correlation. Nevertheless, additional theoretical work is needed to establish formal calibration rules, derive asymptotic distributions, and determine application-specific thresholds for practical use. Until such results become available, NDI should be viewed primarily as an exploratory diagnostic measure whose value lies in its ability to complement existing approaches to dependence analysis.

8.2 Advantages and Limitations

The proposed Nonsense Dependence Index (NDI) possesses several attractive theoretical and practical features that distinguish it from conventional measures of association. At the same time, as with any newly proposed statistical methodology, certain limitations remain and warrant further investigation.

A primary advantage of NDI is its conceptual simplicity. Once estimates of Kendall's rank correlation coefficient τ and the Bergsma–Dassios dependence coefficient τ^* are available, the index can be computed directly through the expression $\text{NDI} = |\tau| - \sqrt{\tau^*}$. The resulting quantity has an intuitive interpretation as a measure of discrepancy between pairwise ordinal association and higher-order dependence structure. Unlike many sophisticated dependence measures whose interpretation may be difficult for practitioners, NDI provides a straightforward diagnostic framework in which values near zero are generally associated with genuine dependence, whereas increasingly positive values suggest the possibility of nonsense or trend-induced association.

A second advantage is the nonparametric nature of the methodology. The index does not require assumptions of linearity, normality, homoscedasticity, or any other parametric model specification. Consequently, NDI may be applied to a broad range of settings involving nonlinear relationships, heavy-tailed distributions, skewed data, and other situations in which

classical correlation-based methods may be unreliable. This flexibility makes the index potentially useful across diverse scientific disciplines, including economics, environmental science, epidemiology, finance, and machine learning.

A third advantage is that NDI combines information from two fundamentally different notions of dependence. Kendall's tau is based on pairwise concordance and discordance and provides a robust measure of monotonic association. The Bergsma–Dassios coefficient, by contrast, is based on quadruple comparisons and captures more global features of dependence. Since τ^* vanishes under independence while remaining sensitive to a broad class of dependence structures, it complements the information provided by Kendall's tau. The proposed index therefore exploits the strengths of both statistics and provides a richer characterization of dependence than either measure considered separately.

A fourth advantage is the ability of NDI to address one of the oldest and most challenging problems in statistical analysis, namely the distinction between genuine dependence and spurious association. Classical measures such as Pearson's correlation coefficient, Spearman's rank correlation, and Kendall's tau are designed to quantify the strength of association but do not directly assess whether the observed relationship arises from a meaningful dependence mechanism. As demonstrated by Yule's classical examples and numerous modern applications involving nonstationary time series, strong correlations can emerge solely because variables share common trends. The simulation studies and empirical examples considered in this paper suggest that NDI is capable of identifying many such situations through the emergence of relatively large positive values.

A fifth advantage concerns robustness. Because both τ and τ^* are rank-based statistics, the resulting index is less sensitive to extreme observations than conventional moment-based measures such as Pearson's correlation coefficient. This robustness is particularly valuable in applications involving outliers, measurement error, or departures from standard distributional assumptions.

The proposed methodology also exhibits considerable flexibility. Since the definition of NDI depends only on two dependence measures, alternative versions may be developed by replacing τ or τ^* with related quantities. Such flexibility creates opportunities for constructing families of nonsense-dependence diagnostics tailored to specific applications or data structures.

Despite these attractive features, several limitations should be acknowledged. First, the performance of NDI is inherently linked to the statistical properties of both τ and τ^*. Any estimation error, finite-sample bias, or numerical instability associated with these quantities may propagate to the resulting index. Consequently, the behavior of NDI in small samples depends critically on the sampling properties of its constituent estimators. Although the simulation studies presented in this paper provide encouraging evidence, additional theoretical investigation is required to establish finite-sample accuracy under a broader range of dependence structures.

Second, the theoretical distribution of NDI remains largely unexplored. Unlike many classical statistics for which asymptotic distributions are well understood, there is currently no general distributional theory available for the proposed index. As a result, confidence intervals, hypothesis tests, and formal significance assessments cannot yet be constructed in a rigorous manner. The development of asymptotic theory therefore represents an important direction for future research.

Third, there is presently no universally accepted threshold for determining when an observed NDI value should be regarded as evidence of nonsense dependence. Although the simulation studies suggest that values near zero are characteristic of genuine dependence and larger positive values are frequently associated with spurious relationships, the precise interpretation of intermediate values remains context dependent. Practical calibration may depend on sample size, dependence strength, noise level, and the underlying data-generating mechanism.

Fourth, the current formulation is restricted to bivariate dependence. Many contemporary applications involve high-dimensional data, complex network structures, or interactions among multiple variables simultaneously. The behavior of NDI in such settings remains largely unknown. Extending the methodology to multivariate dependence structures therefore constitutes an important challenge.

Fifth, NDI should not be interpreted as a measure of causality. A large positive value may indicate that an observed association is potentially driven by common trends or other nuisance mechanisms, but it does not establish that the relationship is spurious in a causal sense. Conversely, a small NDI value does not prove that a causal relationship exists. As with all statistical diagnostics, substantive scientific knowledge and domain-specific considerations remain essential for proper interpretation.

A further limitation concerns computational complexity. The Bergsma–Dassios coefficient is based on higher-order combinations of observations and can be computationally demanding when implemented using naive algorithms. Although efficient algorithms have been developed in recent years, computational scalability remains an important practical issue for very large datasets. Future improvements in algorithmic efficiency may substantially enhance the applicability of the proposed methodology to modern large-scale data analysis.

Finally, the empirical evidence presented in this paper, while encouraging, remains necessarily limited. Additional investigations involving a wider range of simulated and real-world datasets are needed to fully understand the strengths, weaknesses, and practical operating characteristics of the proposed index. Such studies will help clarify the situations in which NDI provides the greatest benefit and identify settings in which alternative approaches may be preferable.

Overall, the advantages of NDI lie in its nonparametric nature, interpretability, robustness, and ability to distinguish between genuine and potentially spurious dependence. At the same time, important theoretical, computational, and methodological challenges remain. These limitations do not diminish the usefulness of the proposed index; rather, they highlight a rich set of opportunities for future research aimed at refining and extending the methodology.

8.3 Extensions and Future Research Directions

The development of the Nonsense Dependence Index (NDI) opens a wide range of theoretical, methodological, computational, and applied research opportunities. Although the results presented in this paper provide encouraging evidence regarding the usefulness of the proposed index, many important questions remain unanswered. Addressing these questions will not only deepen the theoretical understanding of NDI but also broaden its applicability across diverse scientific disciplines.

One of the most important directions for future research concerns the development of a formal inferential framework for NDI. At present, the index is primarily intended as a diagnostic measure whose interpretation is based on empirical behavior observed through simulation studies and real-data applications. A rigorous asymptotic theory would considerably strengthen the methodology. In particular, it would be desirable to derive the asymptotic distribution of NDI under independence, genuine dependence, and various classes of spurious dependence. Such results would facilitate the construction of confidence intervals, hypothesis tests, critical values, and data-driven decision rules. Bootstrap and permutation-based inference procedures may also prove useful, especially in situations where analytical distributional results are difficult to obtain.

A second major research direction involves the investigation of the finite-sample properties of NDI. While the asymptotic behavior of the index is important from a theoretical perspective, practical applications often involve moderate sample sizes. Systematic studies of bias, variance, consistency, robustness, and convergence rates under a variety of dependence structures would provide valuable insight into the reliability of the proposed methodology. Such investigations could also lead to finite-sample corrections and improved estimators for practical use.

Another promising avenue concerns the extension of NDI to multivariate settings. The present formulation focuses exclusively on bivariate dependence, whereas many modern applications involve complex interactions among multiple variables. Extending the methodology to multivariate analogues of Kendall's tau and Bergsma–Dassios dependence measures would allow the construction of multivariate NDI statistics capable of identifying spurious associations within high-dimensional systems. Such developments could be particularly useful in fields such as genomics, systems biology, neuroscience, environmental monitoring, and financial risk management, where dependence structures often involve large numbers of interacting variables.

Related to this is the problem of high-dimensional dependence analysis. In contemporary data science applications, the number of variables may greatly exceed the sample size. Investigating the behavior of NDI under high-dimensional asymptotic regimes, sparse dependence structures, graphical models, and network representations constitutes an important research challenge. Such studies may lead to new tools for identifying spurious associations in large-scale data environments.

Time-series analysis represents another particularly important area for future development. The phenomenon of nonsense correlation originally arose in the context of temporally evolving data, and many modern applications continue to involve nonstationary processes. It would therefore be valuable to investigate the behavior of NDI under autoregressive, moving-average, autoregressive integrated moving-average (ARIMA), state-space, long-memory, and nonlinear time-series models. Particular attention should be given to integrated and cointegrated processes, where conventional correlation measures are well known to produce misleading conclusions. Establishing theoretical properties of NDI under various forms of temporal dependence could substantially enhance its usefulness in econometrics, finance, climatology, hydrology, and environmental science.

Future research may also explore the role of NDI in causal inference. Although the proposed index is not intended as a measure of causality, its ability to distinguish genuine dependence from potentially spurious association suggests possible connections with

causal discovery methodologies. Investigating relationships between NDI, confounding effects, mediation structures, and causal graphical models may provide valuable insights into the interpretation of statistical dependence in observational studies.

Another promising direction concerns the development of weighted and adaptive variants of the index. The current formulation assigns equal importance to all observations. However, in many applications the strength and nature of dependence may vary across different regions of the sample space. Weighting schemes based on local density estimation, nearest-neighbor distances, spacing statistics, depth functions, or information-theoretic quantities could enhance sensitivity to local dependence structures while preserving robustness. Such modifications may be particularly beneficial in heterogeneous datasets containing clusters, outliers, or regions of varying data density.

Robustness considerations provide another fertile area for future investigation. Although rank-based measures are generally less sensitive to outliers than moment-based statistics, contamination and measurement error can still influence dependence assessment. Robust versions of NDI designed specifically for contaminated data, heavy-tailed distributions, missing observations, and noisy measurements may improve practical performance in real-world applications. Extensions to censored, truncated, and incomplete data settings are likewise worthy of investigation.

The treatment of missing data represents a particularly important direction. Many scientific datasets contain missing observations arising from nonresponse, attrition, censoring, or data collection limitations. Future work could explore inverse-probability-weighted, augmented inverse-probability-weighted, multiple-imputation-based, and semiparametric estimators of NDI under missing-at-random and missing-not-at-random mechanisms. Such developments would considerably expand the range of practical applications.

Computational methodology constitutes another critical area for future research. Since the Bergsma–Dassios coefficient is based on higher-order combinations of observations, naive implementations can become computationally intensive for large datasets. Although efficient algorithms have already been developed for computing τ^*, further improvements remain desirable. Parallel algorithms, distributed computing strategies, streaming-data implementations, GPU-based computation, and scalable approximation methods may substantially improve the applicability of NDI to modern large-scale datasets.

Connections with other dependence measures also merit further investigation. The statistical literature contains a rich collection of dependence measures, including distance correlation, maximal information coefficient, mutual information, Hilbert–Schmidt independence criteria, copula-based measures, and kernel-based dependence statistics. Understanding the theoretical relationships between NDI and these alternative measures may provide deeper insight into the nature of dependence and help identify situations in which NDI offers distinct advantages.

From a theoretical standpoint, further work is needed to characterize the mathematical properties of NDI more completely. Potential topics include sharp bounds, extremal properties, invariance principles, decomposition results, influence functions, robustness measures, and information-theoretic interpretations. Such results would contribute to a more comprehensive theoretical foundation for the proposed methodology.

Extensive empirical validation is also necessary. Although the real-data applications presented in this paper provide encouraging evidence, additional case studies across a broad

range of disciplines would strengthen understanding of the practical strengths and limitations of NDI. Potential application areas include economics, finance, epidemiology, environmental science, climate research, neuroscience, genomics, social network analysis, engineering reliability, and machine learning. Comparative studies involving existing dependence measures would be especially valuable for assessing the relative performance of the proposed index.

Finally, future research may investigate domain-specific adaptations of NDI. Different scientific fields often encounter distinct forms of spurious dependence arising from temporal trends, spatial autocorrelation, latent confounding, compositional effects, or measurement artifacts. Tailoring the NDI framework to address such domain-specific challenges may substantially enhance its practical utility and broaden its impact across the applied sciences.

Overall, the proposed NDI represents only an initial step toward a broader framework for diagnosing nonsense dependence. The theoretical developments, simulation studies, and empirical illustrations presented in this paper suggest that the index provides a promising new perspective on the longstanding problem of distinguishing genuine dependence from spurious association. Continued research along the directions outlined above has the potential to transform NDI from a diagnostic indicator into a comprehensive statistical methodology for dependence assessment in complex modern datasets.

9 Conclusion

The problem of distinguishing genuine statistical dependence from spurious or nonsense association has occupied statisticians for nearly a century since the pioneering work of Yule (1926). Although classical dependence measures such as Pearson's correlation coefficient, Spearman's rank correlation coefficient, and Kendall's tau provide valuable summaries of association, they are not designed to determine whether an observed relationship reflects a genuine underlying dependence mechanism or merely arises from common trends, temporal evolution, or other extraneous factors. Consequently, strong correlations may be observed even between variables that possess no meaningful substantive connection, leading to potentially misleading scientific conclusions.

Motated by this longstanding challenge, the present paper introduced a new diagnostic measure called the *Nonsense Dependence Index* (NDI), defined as NDI $= |\tau| - \sqrt{\tau^*}$, where τ denotes Kendall's rank correlation coefficient and τ^* denotes the Bergsma–Dassios dependence coefficient. The fundamental idea underlying the proposed methodology is that genuine dependence should be reflected consistently across both pairwise and higher-order measures of association. When the dependence structure is authentic, the magnitudes of $|\tau|$ and $\sqrt{\tau^*}$ tend to evolve in a coordinated manner, resulting in relatively small values of NDI. In contrast, when apparent associations arise primarily from common trends or other nuisance mechanisms, Kendall's tau may become artificially inflated relative to the higher-order dependence captured by τ^*, producing substantially larger positive values of the index.

A substantial portion of this paper was devoted to establishing the theoretical foundations of the proposed measure. The mathematical properties of NDI were investigated in detail, including its range, behavior under independence, and relationship to existing dependence measures. Theoretical results demonstrated that the index possesses several desirable characteristics, including nonparametric invariance and interpretability in terms of discrep-

ancies between distinct notions of dependence. These properties provide a rigorous basis for using NDI as a diagnostic tool in dependence analysis.

Extensive simulation studies were conducted to evaluate the behavior of the proposed index under a variety of dependence structures. The simulation framework considered both genuine and potentially spurious forms of association, including linear dependence, nonlinear dependence, monotonic relationships, trend-driven relationships, and several classes of synthetic examples specifically designed to mimic nonsense correlation. Across these settings, NDI consistently exhibited behavior that aligned with its intended purpose. Under genuine dependence, the index generally remained close to zero even when the strength of dependence was substantial. In contrast, trend-induced and artificially generated associations often produced noticeably larger positive values, indicating a discrepancy between pairwise and higher-order dependence structures.

The empirical studies further illustrated the practical usefulness of the methodology. The Old Faithful geyser dataset provided an example of a well-established physical relationship between eruption duration and waiting time. In this case, conventional dependence measures indicated strong association, and the estimated NDI remained close to zero, supporting the interpretation that the observed relationship represents genuine dependence rather than a spurious artifact. Conversely, the classical Yule-type example of spurious correlation produced extremely large conventional correlation coefficients despite the absence of any meaningful underlying connection. The corresponding NDI attained its maximum value, correctly flagging the relationship as a potential instance of nonsense dependence. These contrasting examples demonstrate the ability of the proposed index to differentiate between authentic and trend-induced associations in practice.

Several features make NDI particularly attractive for practical applications. The methodology is entirely nonparametric and therefore avoids restrictive assumptions regarding distributional form, linearity, or model specification. Because it is based on rank-type dependence measures, it exhibits a degree of robustness against outliers and departures from standard assumptions. The index is also simple to compute and interpret, making it accessible to practitioners across a wide range of scientific disciplines. Furthermore, by combining information from Kendall's tau and the Bergsma–Dassios coefficient, NDI incorporates both pairwise and higher-order aspects of dependence, thereby providing a richer characterization of association than many conventional measures.

At the same time, the present work should be viewed as an initial contribution rather than a final resolution of the problem of nonsense dependence. Several important questions remain open. In particular, a complete asymptotic theory for NDI has yet to be developed, and formal inferential procedures such as confidence intervals and hypothesis tests remain topics for future investigation. Likewise, the determination of universally applicable thresholds for practical interpretation represents an important challenge. Additional theoretical work is also needed to better understand the behavior of the index under a broader range of dependence structures, particularly in finite samples and high-dimensional settings.

The proposed methodology also creates opportunities for numerous extensions. Potential directions include multivariate generalizations, time-series adaptations, weighted and robust variants, missing-data extensions, computational improvements, and integration with modern machine-learning and causal-inference frameworks. Such developments may significantly broaden the applicability of NDI and enhance its usefulness in contemporary data analysis.

Given the prevalence of large, complex, and often nonstationary datasets in modern scientific research, tools capable of identifying potentially misleading associations are likely to become increasingly important.

From a broader perspective, the contribution of this paper lies not merely in the introduction of a new numerical index but in the conceptual framework that it represents. Traditional dependence analysis has largely focused on measuring the strength of association. The proposed NDI shifts attention toward assessing the credibility of that association by comparing information derived from different levels of dependence structure. In doing so, it provides a new perspective on one of the oldest problems in statistics: determining whether an observed relationship reflects a genuine underlying phenomenon or merely an accidental consequence of shared trends and other nuisance effects.

In conclusion, the results presented in this paper suggest that the Nonsense Dependence Index provides a useful and promising diagnostic tool for dependence analysis. Theoretical investigations, simulation experiments, and empirical applications collectively indicate that NDI is capable of distinguishing many genuine dependence structures from potentially spurious associations that may mislead conventional correlation-based analyses. While additional theoretical and empirical work remains necessary, the proposed methodology offers a novel and practically useful approach to addressing the enduring problem of nonsense correlation. It is hoped that the ideas developed here will stimulate further research into higher-order dependence diagnostics and contribute to more reliable statistical inference in the presence of complex dependence structures.

Acknowledgements

The author is grateful to his colleagues and research companions for suggesting their valuable comments and constructive suggestions regarding this novel effort, which helped to improve the quality and presentation of the manuscript.

Funding

This research received no external funding.

Conflict of Interest

No conflicts of interest at all.

Codes Availability

All the R codes can be accessed at `https://github.com/sthdas999/The-Nonsense-Dependence-Index`.

References

[1] Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philosophical Transactions of the Royal Society A*, **187**, 253–318.

[2] Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, **15**, 72–101.

[3] Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81–93.

[4] Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, **89**, 1–63.

[5] Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, **2**, 111–120.

[6] Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, **33**, 311–340.

[7] Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, **35**, 2769–2794.

[8] Bergsma, W. and Dassios, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, **20**, 1006–1028.

[9] Nandy, P., Weihs, L. and Drton, M. (2016). Large-sample theory for the Bergsma–Dassios sign covariance. *Electronic Journal of Statistics*, **10**, 2287–2311.

[10] Weihs, L., Drton, M. and Leung, D. (2016). Efficient computation of the Bergsma–Dassios sign covariance. *Computational Statistics*, **31**, 315–328.