All Actuaries Summit 2026
25 – 27 May 2026, Melbourne

Beyond pairwise correlation: capturing nonlinear and higher-order dependence with distance statistics

Prepared by Benjamin Avanzi, Guillaume Boglioni Beaulieu, Pierre Lafaye de Micheaux, Ho Ming Lee, Bernard Wong and Rui Zhou

Presented to the Actuaries Institute
2026 All-Actuaries Summit
25-27 May 2026

Actuaries Institute
ABN 69 000 423 656
Level 34, Australia Square, 264 George Street, Sydney NSW 2000
P +61 (0) 2 9239 6100 | actuaries.asn.au

Abstract

Measuring and modelling dependence between risks is crucial in many actuarial applications, such as when assessing diversification benefits or setting capital requirements. Current industry practice relies heavily on Pearson's correlation coefficient, despite well-known limitations. In particular, it captures only linear association, is restricted to pairwise relationships, and does not naturally extend to multivariate settings involving multiple random variables or random vectors (such as when one-hot encoding categorical variables). Sole reliance on Pearson's linear correlation may fail to detect important nonlinear, higher-order, and mutual dependence structures.

In this paper, we discuss and illustrate several distance-based dependence statistics which do not suffer from the same limitations as correlation, and discuss how they can be used in actuarial applications. In the bivariate setting, we consider the Hellinger correlation (Geenens & Lafaye de Micheaux, 2022) as a tool for measuring dependence between two continuous univariate random variables, and distance covariance (Székely et al., 2007) as a tool for detecting and testing dependence, especially when random vectors are involved. We then discuss extensions to higher-dimensional settings, including joint distance covariance for assessing mutual dependence across multiple random variables (or vectors), and the auto-distance correlation function for time series applications such as forecasting mortality rates over time.

Throughout the paper, we illustrate the use of these tools in actuarial contexts and we direct the reader to available software implementations. Overall, the paper aims to provide actuaries with a practical introduction to distance-based dependence statistics and to show how they can complement classical correlation-based tools in actuarial workflows.

Keywords: *Dependence modelling; Mutual dependence; Hellinger correlation; Distance covariance; Random vectors*

1. Introduction

1.1. Background and context

The work of actuaries often involves modelling dependence and/or formulating dependence assumptions, and both these tasks have long been recognised as challenging yet essential in actuarial practice. In fact, one of the first definitions of "independence" is due to a German mathematician who specialised in probability theory and actuarial mathematics (Bohlmann, 1901). Independence (or lack thereof) is a central consideration across many aspects of the insurance industry. For instance, when determining capital adequacy, insurers must consider the dependence between different risk categories (e.g., asset risk and insurance risk), primarily to account for diversification benefits (see, for instance, Australian Prudential Regulation Authority, 2023a). Dependence is also relevant for interactions such as risk accumulation, where multiple exposures are affected by a common event (Australian Prudential Regulation Authority, 2023c; CRO Forum, 2015). In reserving, widely used reserving models impose independence assumptions on different variables. For example, Mack's distribution-free chain ladder (Mack, 1993) assumes that each accident year's claims development is independent of every other accident year's claims development. In addition, in pricing insurance products, severities are often assumed to be mutually independent, typically independent and identically distributed (i.i.d.), meaning that severities are independent across claims and follow the same distribution. Moreover, severity and frequency are often assumed to be independent (as in the classic collective risk model). Dependence modelling is becoming increasingly important with the emergence of newer products such as cyber insurance, where incidents are likely to be associated (Awiszus et al., 2023; Sun et al., 2021). Overall, modelling dependence and validating dependence assumptions (including mutual independence) remain crucial tasks in actuarial practice.

Pearson's correlation coefficient is one of many numerical summaries of a particular linear form of association (Bradley, 1985), and has been widely used by actuaries as the default mea-

sure of dependence. It is often treated as a standard tool when assessing dependence between risks or lines of business (Avanzi et al., 2016; Denuit et al., 2006; P. O. J. Kelliher et al., 2020; P. Kelliher, 2022; R. A. Shaw et al., 2011). More generally, in economic capital and enterprise risk management (ERM) practice, dependence is often captured through variance–covariance aggregation, where a correlation matrix is a key input used to aggregate stand-alone risk measures and determines the diversification level and capital allocation (e.g., Bernard & Vanduffel, 2015). For instance, APRA's asset risk charge represents the minimum capital required to be held by an insurer against asset risks (Australian Prudential Regulation Authority, 2023b), and the aggregation formula is computed using correlations between asset risks. This is also the case under Solvency II, the prudential regulatory framework for EU (re)insurers. Its "standard formula" for computing the Solvency Capital Requirement (SCR) uses prescribed correlations (European Parliament and the Council of the European Union, 2009), although internal models may also be used to compute the SCR. In addition, under IFRS 17, when adjusting for uncertainty in insurance cash flows, correlation matrices are commonly used to reflect diversification across different risks (Jiang, 2020). Finally, in a typical modelling pipeline for pricing (e.g., GLMs), correlation is widely used to diagnose multicollinearity and redundancy among predictors. Interpretability tools such as partial dependence plots and SHAP-based feature attributions can also become unreliable when predictors are dependent (Molnar, 2020; Ng et al., 2025). In actuarial workflows, simple dependence diagnostics such as correlation matrices are therefore often used as an initial screen in model review (National Association of Insurance Commissioners, 2025). These examples illustrate that correlation remains the default operational summary of dependence across the industry.

Despite the widespread use of Pearson's correlation coefficient, its use presents several well-known (and some less well-known) issues, some of which are discussed in the landmark paper by Embrechts, McNeil, and Straumann (2002). In this paper, we revisit some of these pitfalls, but more importantly, we extend the discussion to additional, important issues for which solutions now exist, which we aim to introduce to the broader actuarial community. In particular, we focus on three main limitations of correlation: zero correlation does not characterise independence, correlation is inherently pairwise and cannot detect higher-order dependence, and correlation does not naturally extend to random vectors or mixed-type data. These issues are discussed further, together with actuarial examples, in Section 2.

1.2. Dependence as a "distance" from independence

The limitations of Pearson's correlation coefficient motivate the use of alternative dependence measures that can detect broader forms of dependence, including non-linear, non-monotonic, and higher-order (beyond pairwise) dependence, as well as dependence involving random vectors. In this paper, we review *distance-based dependence statistics* as complementary tools for addressing the limitations of correlation. Mutual independence of d random variables or random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d$ (and, in particular, pairwise independence when $d = 2$) means that the joint distribution $\mathbb{P}_{(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_d)}$ coincides with the product of the marginal distributions $\prod_{j=1}^{d} \mathbb{P}_{\boldsymbol{X}_j}$. We use the term *distance-based dependence statistic* to refer to quantities that assess departure from independence by quantifying the dissimilarity between the joint distribution and the product of the marginal distributions. In practice, this comparison is often carried out through suitable representations of these distributions, such as characteristic functions or copula densities, and dependence is then assessed through the discrepancy between these representations. Note that the resulting quantity is not necessarily a distance in the metric sense; see also Appendix A.

1.3. Purpose of the dependence analysis

Viewing dependence as departure from independence provides a common conceptual framework for the statistics considered in this paper. However, different actuarial tasks require differ-

ent types of dependence tools. It is therefore helpful to clarify the purpose of the dependence analysis before introducing specific statistics in detail.

Dependence tools in actuarial applications are mainly used for two distinct tasks: (i) *measuring* the strength of dependence (e.g., for aggregation of risks and/or diversification analyses, or for comparison purposes), and (ii) *testing* whether independence (or i.i.d.-type) assumptions are reasonable (e.g., when validating modelling assumptions). These two tasks place different requirements on the statistic which is used. As discussed in Geenens and Lafaye de Micheaux (2022), a dependence *measure* (task (i)) should provide an interpretable quantification of dependence strength, so that weak (but nonzero) dependence yields a correspondingly small value. In contrast, an independence *test* (task (ii)) aims at a binary decision, and a powerful test may use a statistic that departs from its null distribution as soon as any dependence is present, even if the underlying dependence is very weak. Consequently, interpreting a test statistic as a measure of dependence strength can be misleading. Conversely, an interpretable dependence measure may not be optimised for independence testing, so its finite-sample power may differ from that of statistics constructed primarily for testing. Which procedure is more powerful in practice depends on the underlying dependence structure of the data.

1.4. Contributions

This paper is tailored to an actuarial audience and makes several contributions. Through several actuarial examples, we show that correlation has important shortcomings: it may fail to detect nonlinear, non-monotonic, vector-level, and higher-order dependence, with material implications for actuarial quantities such as aggregate tail risk and for the assessment of modelling assumptions.

The paper introduces the concept of distance-based dependence statistics as complementary tools for actuarial dependence analysis. Dependence analysis in actuarial work typically involves two distinct tasks: measuring the strength of dependence and testing whether independence assumptions are reasonable. These two tasks require different types of statistics. We discuss several distance-based dependence statistics for these different use cases. In the bivariate setting, we discuss the Hellinger correlation as a useful tool for measuring dependence strength between two continuous univariate random variables, and distance covariance as a useful tool for detecting and testing dependence, especially when random vectors are involved. We then extend the discussion to joint distance covariance for assessing mutual dependence across multiple random variables or vectors, and to the (multivariate) auto-distance correlation function for time series model specification and diagnostics.

Concepts discussed in this paper are extensively illustrated. We discuss the main limitations of each statistic, and point to available `R` and `Python` implementations so that readers can readily incorporate these tools into actuarial workflows.

Table 1.1 summarises the main tasks and key properties of the dependence statistics discussed in this paper. In particular, it highlights the main differences between correlation-based and distance-based measures in terms of the types of dependence they can capture[1], whether they characterise independence (that is, whether independence can be inferred from their value), and whether they can be applied to random vectors or to higher-order dependence.

1.5. Paper structure

Section 2 discusses and illustrates the shortcomings of Pearson's correlation coefficient, with actuarial examples. Section 3 discusses distance-based dependence statistics for measuring

[1] Note that the distance statistics introduced here are not "directional". This makes sense, since a distance cannot be negative, and it also makes sense when considering that the notion of direction is not well defined for associations that are more complex than linear or monotonic.

Statistic	**Pearson's r**	**Spearman's ρ**	**Kendall's τ**	**Hellinger correlation**	**Distance covariance**	**Joint distance covariance**	**(Multivariate) Auto-distance correlation**
Main task	Measure linear association	Measure monotonic association	Measure monotonic association	Measure dependence strength	Test for pairwise dependence	Test for mutual dependence	Serial dependence diagnostics
Non-linear?	✗	✓	✓	✓	✓	✓	✓
Non-monotonic?	✗	✗	✗	✓	✓	✓	✓
$0 \iff$ independence?	✗	✗	✗	✓	✓	✓	✓
Directional?	✓	✓	✓	✗	✗	✗	✗
Margin-free?	✗	✓	✓	✓	✗	✗	✗
Random vectors?	✗	✗	✗	✗	✓	✓	✓
Mutual dependence ($d > 2$)?	✗	✗	✗	✗	✗	✓	✗

Table 1.1: This table summarises the main tasks and key properties of the distance-based dependence statistics reviewed in this paper. The properties include the types of dependence they can capture. Here, "Directional" means that the sign indicates a positive or negative relationship, and "Margin-free" means that the statistic remains unchanged under strictly monotone transformations of the marginals, i.e., it depends only on the underlying dependence structure. The symbols ✓ and ✗ indicate whether a given property is satisfied.

and testing dependence between two random variables or random vectors, focusing on the Hellinger correlation and distance covariance, respectively. Section 4 then considers the extension of distance covariance to the multivariate setting, which is useful for assessing mutual dependence among more than two random vectors. Section 5 introduces another extension of distance covariance, namely the auto-distance covariance, which is useful for time series applications involving both univariate and multivariate time series. Throughout, we explain how actuaries can incorporate these statistics into their workflows, and we discuss their main properties and limitations. We also illustrate the use of available `R` and `Python` packages for implementing them. Section 6 concludes.

2. What's wrong with pairwise correlation? Review and illustrations

As outlined in Section 1.1, we focus on three main limitations of Pearson's correlation coefficient in this paper. In this section, we illustrate each of these shortcomings through actuarial examples, highlighting the need for better dependence statistics.

2.1. Uncorrelatedness does not characterise independence

It is well known that zero Pearson correlation between two random variables does not, in general, imply independence. One important exception is the bivariate normal case, for which zero correlation *is* equivalent to independence (Embrechts et al., 2002). This is because the joint normal distribution is fully determined by its mean vector and covariance matrix (Muirhead, 1982). While other correlation coefficients such as Kendall's τ and Spearman's ρ are popular rank-based[2] alternatives (Androschuck et al., 2017; Peng & Wang, 2014; Venter, 2002), they

[2]Here, the 'rank' refers to an observation's corresponding empirical cdf value. It is a number between 0 and 1 that indicates how the observation "ranks" compared to the others. When considering the association between two variables, it is always advisable to plot their ranks rather than their actual observation values (for a pair (x_i, y_i), "$F(x_i)$" against "$G(y_i)$" rather than x_i against y_i, where F and G maps x's and y's to their ranks (ecdf), because

mainly summarise *monotonic* relationships (van den Heuvel & Zhan, 2022), that is, relationships in which one variable tends to move consistently in one direction as the other increases. In particular, even when these coefficients are statistically significant, this does not imply that the underlying relationship itself is linear or monotonic. More complex non-monotonic relationships, such as U-shaped patterns, may not be well summarised by these measures. As a result, these coefficients can provide evidence of association, but they do not in general provide a complete description of non-monotonic dependence, and a value of zero does not imply independence.

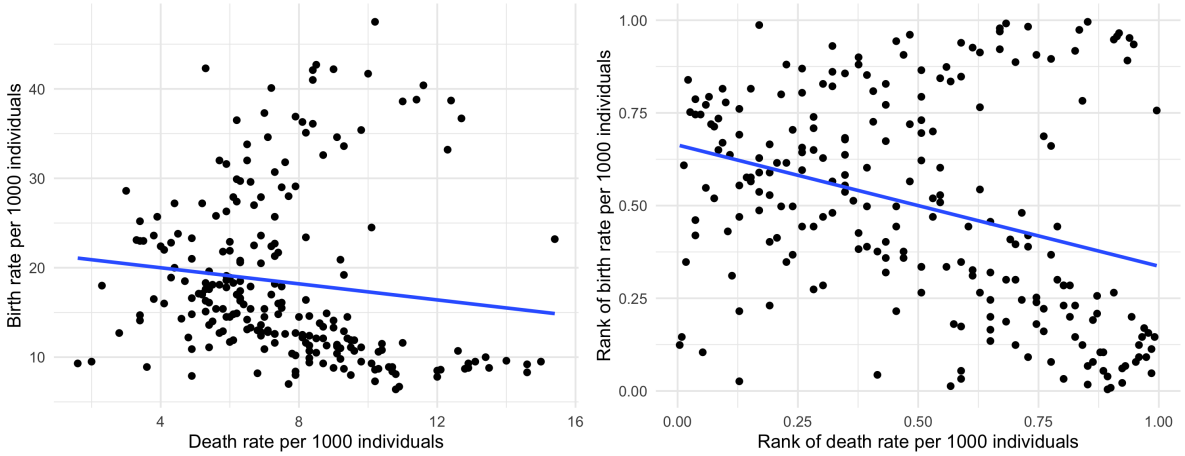


Figure 2.1: This figure displays the birth rate against the death rate in 2020 across countries. The left panel shows the raw values, while the right panel shows the corresponding empirical CDF values (ranks). The blue line represents the fitted linear regression line of birth rate on death rate. The scatterplots show a clear C-shaped relationship between the birth rate and the death rate, whereas the fitted regression line shows only a slight downward trend. This is consistent with the correlation coefficients being slightly negative (Pearson: -0.13, Spearman: -0.33, Kendall: -0.24). However, these negative values arise from the overall downward tendency in the points rather than from a genuinely negative linear or monotonic relationship between the birth rate and the death rate.

We illustrate this pitfall using the birth and death rate data from different countries in the first trimester of 2020 (Central Intelligence Agency, 2020). The birth and death rates are measured as the numbers of births and deaths per 1000 individuals, respectively, across 229 countries. The left panel of Figure 2.1 plots the raw birth rate against the death rate, and the right panel plots the corresponding empirical CDF values (ranks), with each point representing a country. The blue line is the fitted linear regression line of birth rate on death rate. The scatterplot displays a clear C-shaped relationship between the birth and death rates. We compute several association coefficients and perform hypothesis tests for

$$\begin{aligned} H_0 &: \theta = 0, \\ H_1 &: \theta \neq 0, \end{aligned} \tag{2.1}$$

where θ is the corresponding association coefficient. Details of the hypothesis tests are provided in Appendix D on page 45. Here, the Pearson correlation coefficient is insignificant, with $r = -0.13$, which is consistent with the fitted regression line being fairly flat and not indicating a strong linear relationship. In contrast, Spearman's ρ and Kendall's τ are significant, with values of $\rho = -0.33$ and $\tau = -0.24$, respectively. This can be understood from the rank plot in the right panel, since Spearman's ρ is the correlation of the ranks (an indication of the slope of

this takes the effect of their marginal distributions away. In fact, in a bivariate case, the density of those "ranks" in a scatterplot is exactly what a copula density aims to represent.

the blue line through the ranks). Although the relationship is clearly non-monotonic, the rank plot still exhibits an overall downward tendency, which leads to significant negative values of Spearman's ρ and Kendall's τ. On one hand, Pearson correlation fails to capture the C-shaped relationship between the birth and death rates. On the other hand, although Spearman's ρ and Kendall's τ are statistically significant, they summarise the relationship as a weak negative monotonic trend, even though the dependence in Figure 2.1 is clearly non-monotonic and C-shaped.

This graphical example shows that a small or insignificant correlation does not rule out meaningful dependence. In addition, even when a correlation coefficient is statistically significant, this does not necessarily mean that the relationship between the two variables is in fact linear or monotonic. Rather, correlation can miss important features of the dependence structure, as illustrated above. Misinterpreting an insignificant correlation as independence can also have serious consequences, for example by leading to a material underestimation of tail risk in business operations, as we illustrate in Appendix B.1 on page 40. The issue becomes even more problematic when dependence is only visible through the joint behaviour of multiple variables rather than through any single pair, since pairwise correlation may fail to reveal such dependence altogether, as we discuss next.

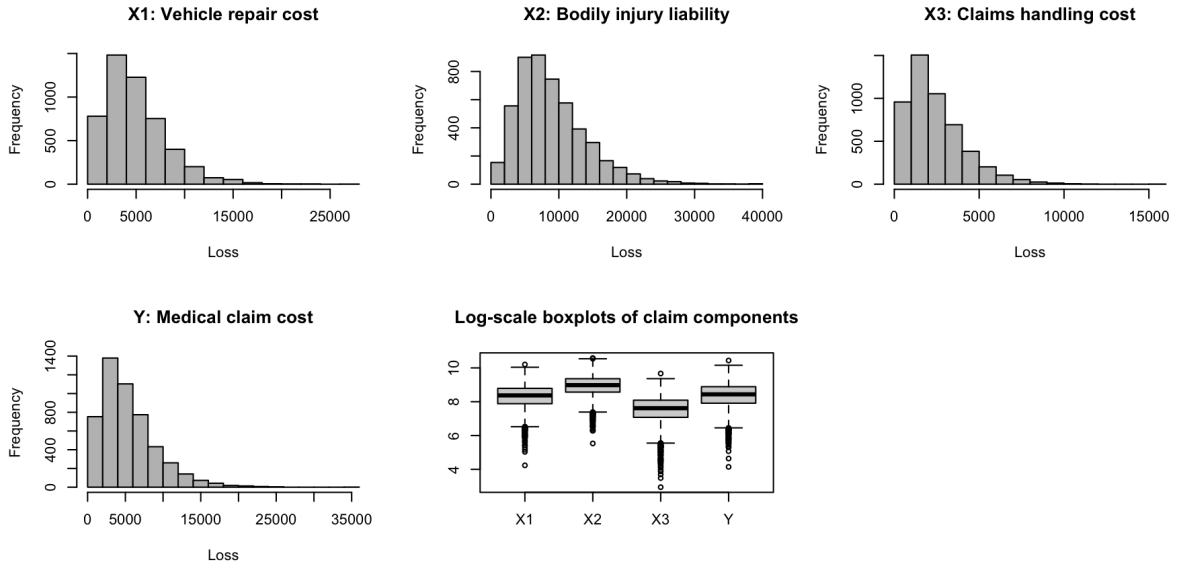


Figure 2.2: Distributions of the motor claim components $\boldsymbol{X} = (X_1, X_2, X_3)$ and the medical claim Y.

2.2. Correlation is inherently pairwise and cannot detect higher-order dependence

Correlation coefficients measure linear or monotonic relationships between pairs of random variables, and are therefore inherently pairwise summaries of dependence. This becomes problematic when the relevant dependence is visible only through the joint behaviour of several variables rather than through any single pair; this is further explained in Section 4 and illustrated below. In such cases, *pairwise* correlation may suggest little or no dependence, even though the variables are not *jointly* ("multi-wise") independent. We illustrate this issue using a simulated multi-line insurance policy. Note that this example is constructed to illustrate the limitations of Pearson correlation; consequently, the dependence structure is intentionally exaggerated and may not fully reflect day-to-day insurance operations.

The policy involves two lines of business: motor claims, with three components,

$$\boldsymbol{X} = (X_1, X_2, X_3) = \text{(vehicle repair cost, bodily injury liability, claims handling cost)},$$

and medical claims, where Y denotes the medical claim cost recorded under another line of business. Figure 2.2 displays the simulated claim distributions using histograms and boxplots.

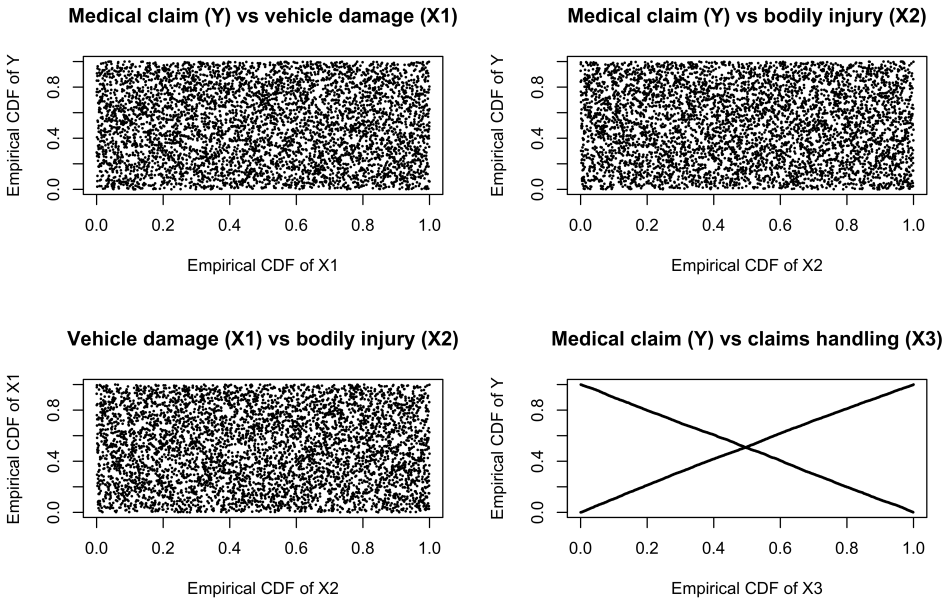


Figure 2.3: Empirical CDF scatterplots between each motor claim component and the medical claim cost Y. The plots of (X_1, X_2), (X_1, Y), and (X_2, Y) show no obvious pattern, consistent with pairwise independence and the near-zero Pearson correlations reported in Table 2.1. By contrast, the plot of (X_3, Y) exhibits a cross-shaped pattern, even though its Pearson correlation is also close to zero. This illustrates the first limitation of correlation discussed in Section 2.1.

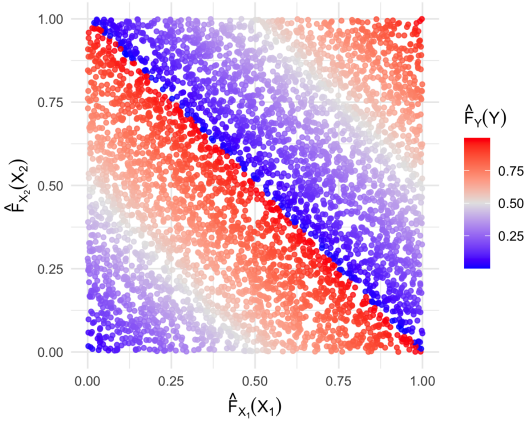
ρ	X_1	X_2	X_3	Y
X_1	1.0000	-0.0259	0.0013	-0.0100
X_2	-0.0259	1.0000	0.0068	-0.0200
X_3	0.0013	0.0068	1.0000	0.0761
Y	-0.0100	-0.0200	0.0761	1.0000

Table 2.1: Pearson correlation coefficients between the simulated claim components. The correlations between X_i and Y for $i = 1, 2, 3$ are close to zero, indicating little or no linear dependence between individual claim components across the two lines of business. The correlations within the motor line are also close to zero. All correlations except that for the pair (X_3, Y) are statistically insignificant. Although the correlation for (X_3, Y) is statistically significant, its magnitude remains small, illustrating that Pearson correlation may provide only a weak linear summary even when a non-trivial, significant dependence structure is present.

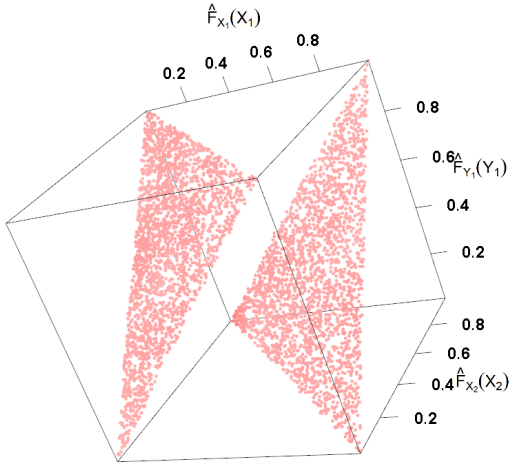
Most importantly, the simulated data are constructed so that (X_1, X_2), (X_1, Y), and (X_2, Y) are pairwise independent (theoretically). This is evidenced by the empirical CDF scatterplots in Figure 2.3. In particular, there is no obvious pattern in the plots of (X_1, X_2), (X_1, Y) and (X_2, Y), which is consistent with the near-zero and statistically insignificant Pearson correlation coefficients reported in Table 2.1. Thus, if one inspects the data only through pairwise relationships, there appears to be little evidence of dependence (indeed, there is none even theoretically).

Furthermore, note that the plot of (X_3, Y) exhibits a cross-shaped pattern, yet its Pearson correlation remains close to zero. This can be intuitively understood by the fact that a regression line would be flat.

However, pairwise independence of (X_1, X_2), (X_1, Y), and (X_2, Y) does not imply that (X_1, X_2, Y) are jointly independent. To make this hidden dependence more visible, Figure 2.4a plots the empirical CDF values (i.e., based on ranks) of X_1 and X_2 as a scatterplot, with each point coloured by the empirical CDF of Y. The idea of representing the third variable Y by colour is taken from Boglioni Beaulieu (2023). When the colours are ignored, the points fill the unit square without any obvious pattern (this is the bottom left scatterplot in Figure 2.3), which is consistent with the independence of X_1 and X_2. However, once the colour scale is taken into account, a pronounced structure becomes visible: points with similar colours align along diagonal bands. Figure 2.4b illustrates this more clearly: the point cloud lies on a structured surface rather than being scattered randomly in three dimensions, indicating that the joint values of (X_1, X_2) constrain the value of Y. This shows that the value of Y depends on the joint behaviour of X_1 and X_2, even though no such dependence is visible from any single pair alone. In other words, looking at one pair at a time can miss dependence that only appears when the variables are considered together. This illustrates the second pitfall of correlation: higher-order dependence among more than two random variables cannot, in general, be detected from pairwise correlations (even, associations in general) alone. Although situations in which variables are pairwise independent but mutually dependent may not be frequently observed in practice, the consequences of ignoring this association can still be severe. This is particularly important in actuarial modelling, where assumptions are often stated in terms of *mutual* independence.



(a) Empirical CDF values of X_1 and X_2, with each point coloured by the empirical CDF value of Y. Ignoring the colours, the empirical CDF plot of X_1 and X_2 shows no obvious pattern, suggesting independence between them. However, once the colour scale of Y is incorporated, a clear diagonal structure becomes visible. This shows that Y depends on the joint behaviour of X_1 and X_2, so that (X_1, X_2, Y) are not jointly independent, even though this is not apparent from any single pair alone.



(b) 3D scatterplot for the multi-line simulated example with the empirical CDF values of X_1, X_2, and Y. As shown in the plot, the points lie on two planes, and knowing the values of X_1 and X_2 restricts the possible values of Y. This indicates that Y depends jointly on (X_1, X_2).

Figure 2.4: 2D (left) and 3D (right) scatterplots for the multi-line simulated example, based on the empirical CDF values of X_1, X_2, and Y.

For example, the classic collective risk model for the aggregate loss

$$S = \sum_{i=1}^{N} X_i \tag{2.2}$$

where N is the number of claims and X_i is the individual claim amount, typically assumes that claim severities are i.i.d., while the Cramér–Lundberg model in ruin theory assumes mutual independence among claim sizes, together with independence between claim sizes and the claim count. It is therefore important to distinguish between *pairwise* and *mutual* independence (Bohlmann, 1909). Pairwise independence means that every pair of variables is independent, whereas mutual independence is a stronger condition: knowing the values of any subset of the random variables does not affect the distribution of the remaining variables.

Following the definition in Kolmogorov (2018), for d random variables $X_1, X_2, \ldots, X_d$,[3] we say that $X_1, \ldots, X_d$ are *mutually independent* if

$$\mathbb{P}(X_1 \in A_1, \ldots, X_d \in A_d) = \prod_{i=1}^{d} \mathbb{P}(X_i \in A_i). \tag{2.3}$$

Here, each A_i represents any subset of the possible outcomes of the corresponding variable X_i. In contrast, pairwise independence requires only that $X_i \perp X_j$ for every $i \neq j$ and for all sets A_i, A_j:

$$\mathbb{P}(X_i \in A_i, X_j \in A_j) = \mathbb{P}(X_i \in A_i)\mathbb{P}(X_j \in A_j), \tag{2.4}$$

which does not in general imply mutual independence. In other words, mutual independence implies pairwise independence—(2.3) implies (2.4), but the converse is false: random variables can be pairwise independent without being mutually independent.

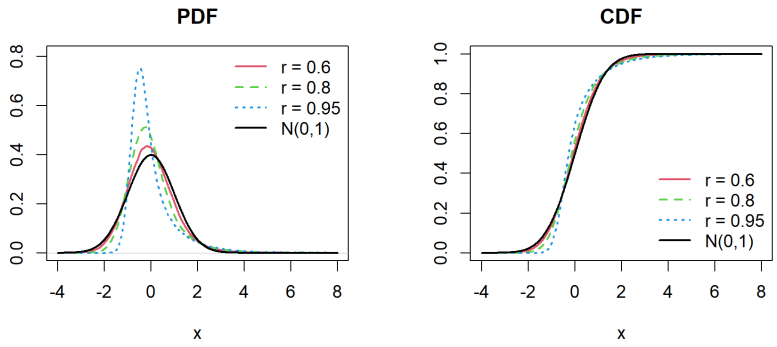


Figure 2.5: This figure compares the probability density functions (PDFs; left panel) and cumulative distribution functions (CDFs; right panel) of the standard normal distribution $N(0,1)$ and the limiting distribution S (i.e. the distribution approached by the normalised sum as the number of random variables tends to infinity) for a sequence of pairwise independent but mutually dependent random variables, as defined in Equation (2.6), for $r = 0.6, 0.8, 0.95$. The figure illustrates that the limiting distribution is non-Gaussian, and hence that the CLT can fail when the random variables are not mutually independent.

A further practical consequence is that the central limit theorem (CLT), which is often relied upon in actuarial applications, can fail for sequences that are pairwise independent but not mutually independent. As illustrated by Avanzi et al. (2021) and Boglioni Beaulieu et al. (2021), pairwise independence alone does not guarantee asymptotic normality. They construct a sequence $X_1, \ldots, X_n$ of pairwise independent, but not mutually independent, standard normal

[3]The definition also applies when we have random vectors; see Section 2.3.

random variables such that the limiting distribution of

$$\sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.5}$$

coincides with that of the random variable

$$S = \sqrt{1-r^2}\,Z + r\chi, \tag{2.6}$$

where $Z \sim N(0,1)$, and where χ is a standardised chi-squared random variable independent of Z, and $0 < r^2 < 1$. Figure 2.5 compares the distribution of S with the standard normal distribution $N(0,1)$, showing that the CLT can fail when the random variables are not mutually independent. Therefore, relying solely on pairwise diagnostics can be insufficient when models or approximations require mutual independence, even when the sample size is large. More generally, dependence that is invisible from every individual pair may become visible only when the variables are considered jointly. In such situations, the relevant object of interest is not a single variable, but a group of variables taken together, that is, a random vector. This leads naturally to the next problem: correlation does not provide a natural summary of dependence when the object of interest is a random vector.

2.3. Correlation does not naturally extend to random vectors and mixed-type data

In actuarial applications, the object of interest is often inherently multivariate. Examples include vectors of claim components, such as in the multi-line example in the previous subsection, policyholder risk profiles used in pricing models that combine numerical and categorical variables (Goldburd et al., 2025; Kafková & Křivánková, 2014; Shi & Shi, 2023), and risk modules that themselves consist of multiple sub-risk modules when determining capital requirements (International Association of Insurance Supervisors, 2024). In particular, actuarial data are often of mixed type, with numerical variables appearing alongside categorical variables. In practice, categorical variables are frequently represented through one-hot encoding, so that even a single rating factor may become a vector of indicator variables. In such settings, the relevant dependence is often not between two scalar random variables, but between a random vector and another variable, or between two random vectors.

However, pairwise dependence statistics such as correlation are designed to quantify the relationship between (only) two scalar random variables at a time. If one relies only on correlation, then the dependence between two random vectors $\boldsymbol{X} = (X_1, \ldots, X_p)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_q)$ is reduced to the collection of pairwise relationships

$$(X_i, Y_j), \quad i = 1, \ldots, p, \quad j = 1, \ldots, q.$$

This can be inadequate, because dependence between random vectors concerns whether the two multivariate objects are associated as wholes, rather than whether each component is associated with another one in isolation. The issue is even more pronounced when the vectors are of mixed type, since pairwise correlation is most interpreted for scalar numerical variables and does not provide a unified summary of dependence for multivariate objects containing categorical components. This is illustrated schematically in Figure 2.6 for two two-dimensional random vectors. Pairwise correlations (left panel) only describe separate component-by-component relationships, represented by the black solid edges. By contrast, a vector dependence measure asks whether $\boldsymbol{X}$ and $\boldsymbol{Y}$ are dependent as wholes. This includes dependence between the vectors as wholes (black solid edge), as well as how the vector-level dependence may relate to individual components (grey solid edges), beyond the pairwise component-to-component associations (grey dashed edges). Most importantly, independence between $\boldsymbol{X}$ and $\boldsymbol{Y}$ implies

independence between each pair (X_i, Y_j), that is,

$$\boldsymbol{X} \perp \boldsymbol{Y} \implies X_i \perp Y_j, \quad \forall i = 1, \ldots, p, \quad j = 1, \ldots, q. \tag{2.7}$$

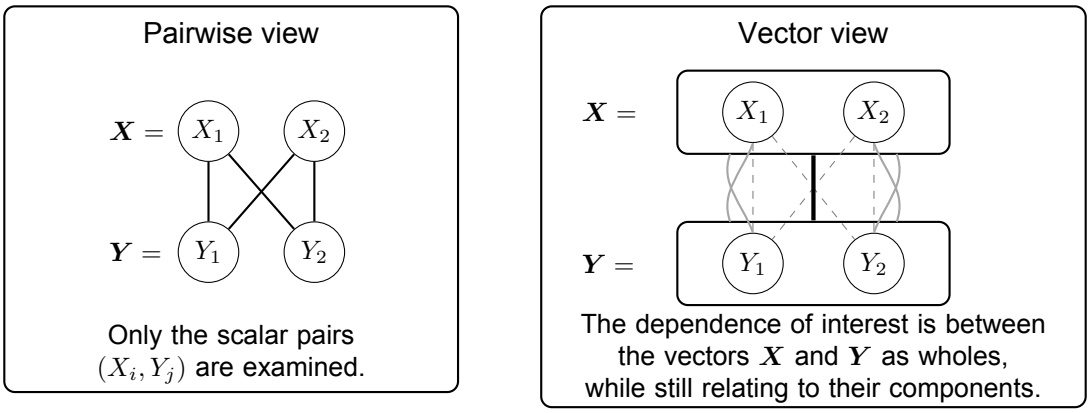


Figure 2.6: Schematic illustration of pairwise versus vector dependence. Left: pairwise dependence only captures the separate associations between each component pair (X_i, Y_j) of the two random vectors. Right: the dependence of interest is between $\boldsymbol{X} = (X_1, X_2)$ and $\boldsymbol{Y} = (Y_1, Y_2)$ as multivariate objects. In addition to the pairwise component-to-component associations (grey dashed edges), vector dependence also allows dependence between the vectors as wholes (black solid edge) and between a vector and an individual component of the other vector (grey solid edges).

This point is also reflected in the multi-line example above. There, the medical claim cost Y is constructed to depend on the joint behaviour of vehicle repair cost X_1 and bodily injury liability X_2, rather than on either component separately. Intuitively, medical costs are often related to both accident severity and bodily injury severity. A claim with substantial vehicle damage but limited bodily injury may lead to relatively small medical expenses, while bodily injury with limited accident severity may also not generate the same level of treatment cost. It is when these features are considered together that the relationship with medical cost becomes more apparent. Hence, the relevant object of interest is the dependence between the vector (X_1, X_2) and Y, rather than the separate pairwise associations. Furthermore, in practice those severities are likely to be encoded with categorical variables.

3. Bivariate distance-based dependence statistics

We first look at the limitation discussed in Section 2.1, namely that a correlation of 0 between two random variables does not, in general, imply independence. This limitation motivates the use of dependence measures that remain informative beyond linear association. In this section we focus on the bivariate setting, which provides the simplest framework for introducing the key ideas before extending them to higher-order dependence. In particular, we consider two complementary statistics: the Hellinger correlation, which is well-suited to quantifying dependence strength between two continuous random variables through a measure bounded between 0 and 1, and distance covariance, which is primarily useful for detecting and testing general dependence between random variables or random vectors. These two statistics reflect the two related objectives in dependence analysis discussed in Section 1.3: (i) measuring the strength of association when variables are dependent, and (ii) testing whether they are independent. However, the distinction is not strict. For instance, the Hellinger correlation can also be used for testing independence. It cannot, however, be used on vectors or mixed-type variables (see Section

2.3); in this case, distance covariance can be used for the first objective instead (although with some caveats as discussed below).

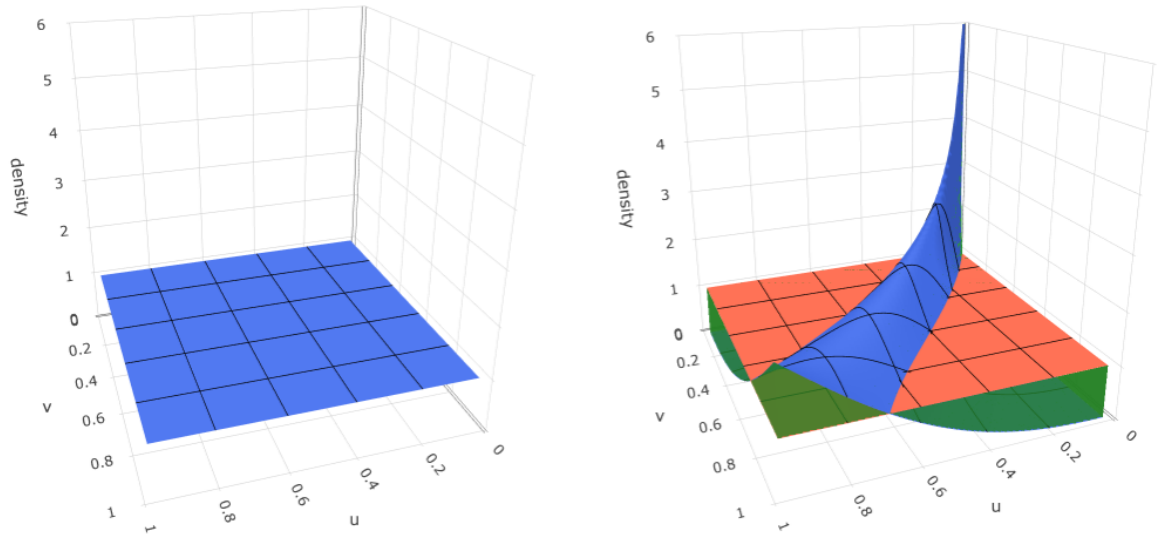


Figure 3.1: This figure plots the surfaces of the square root copula density $\sqrt{c(u,v)}$ in blue and the independence copula density in red, which equals 1 for all $u, v \in [0,1]$, for two continuous random variables (X, Y). In the left panel, $c(u,v)$ is the independence copula density, so the two surfaces overlap and the gap between them is 0. In contrast, in the right panel, $c(u,v)$ is the Clayton copula density (blue surface), while the independence copula density $c(u,v) = 1$ is shown by the red surface. The green shaded volume between the two surfaces illustrates the discrepancy captured by the squared Hellinger distance. More precisely, $2\mathcal{H}^2(X,Y)$ is given by the integrated squared gap between $\sqrt{c(u,v)}$ and 1 over the unit square, as stated in Equation (3.1).

3.1. Hellinger correlation

In actuarial applications, one is often interested in comparing the strength of dependence across different pairs of continuous variables arising within the same portfolio or model, such as losses from two lines of business or different claim components in a multivariate loss setting. Since these quantities are often measured on different scales and may be transformed during preprocessing, it is useful to employ a dependence measure that focuses directly on the underlying dependence structure rather than on the marginal scales. The Hellinger correlation (Geenens & Lafaye de Micheaux, 2022) is particularly appealing in this context, as it is copula-based and therefore margin-free in the continuous case; see also footnote 1 above.

3.1.1 Definition

The Hellinger correlation (Geenens & Lafaye de Micheaux, 2022) is designed to measure dependence between two continuous random variables. For a statistic to be a meaningful dependence measure, Rényi (1959) proposed several desirable properties. Most importantly, it should be margin-free, meaning that strictly monotonic transformations of the random variables do not affect its value. It should also attain the value “1” under perfect dependence (the notion of “perfect dependence” is further discussed in Appendix C). Motivated by these requirements, Geenens and Lafaye de Micheaux (2022) proposed the Hellinger correlation. Being

copula-based, it depends only on the *copula* of (X, Y) and is therefore invariant under strictly monotonic marginal transformations in the continuous case.

The Hellinger correlation is constructed from the squared Hellinger distance $\mathcal{H}^2(X, Y)$ between the joint measure $\mathbb{P}_{XY}$ and the product measure $\mathbb{P}_X \times \mathbb{P}_Y$. For $X, Y \in \mathbb{R}$ with copula density[4] $c(u, v)$, the squared Hellinger distance is given by

$$\mathcal{H}^2(X, Y) = \frac{1}{2} \iint_{[0,1]\times[0,1]} \left(\sqrt{c(u, v)} - 1\right)^2 du\, dv. \tag{3.1}$$

Under independence, the copula density is equal to 1, so $\mathcal{H}^2(X, Y) = 0$. More generally, $\mathcal{H}^2(X, Y)$ quantifies how far the square root of the copula density departs from the independence copula density.

The Hellinger correlation $\eta(X, Y)$ is then defined as a normalised version of the Hellinger distance:

$$\eta(X, Y) = \left(\frac{2}{\left(1 - \mathcal{H}^2(X, Y)\right)^2}\right) \sqrt{\left(1 - \mathcal{H}^2(X, Y)\right)^4 + \sqrt{4 - 3\mathcal{H}^2(X, Y)} - 2}. \tag{3.2}$$

This normalisation calibrates the measure so that the Hellinger correlation coincides with the absolute Pearson correlation $|\rho|$ when (X, Y) follows a bivariate Gaussian distribution. Hence, η can be interpreted on a familiar correlation-like scale, while remaining copula-based and therefore margin-free. Details of the empirical estimator are deferred to Appendix E.1.

Figure 3.1 provides an intuitive illustration of the squared Hellinger distance by plotting the surfaces of the square root copula density, $\sqrt{c(u, v)}$, and the independence copula density, which is equal to 1, under two scenarios. In the first scenario, $c(u, v)$ is the independence copula density, so the two surfaces overlap and the Hellinger distance is equal to 0. In the second scenario, $c(u, v)$ is taken to be the Clayton copula density

$$c(u, v) = (\theta + 1)(uv)^{-\theta-1} \left(u^{-\theta} + v^{-\theta} - 1\right)^{-(2\theta+1)/\theta} \tag{3.3}$$

with $\theta = 5$. In this case, the copula exhibits dependence and therefore differs from the independence copula. The squared Hellinger distance then quantifies the discrepancy between the two surfaces over the unit square (the green volume).

The Hellinger correlation satisfies several desirable properties. In particular, for two univariate continuous random variables X and Y, it characterises independence, that is,

$$\eta(X, Y) = 0 \quad \iff \quad X \text{ and } Y \text{ are independent.} \tag{3.4}$$

Moreover,

$$0 \leq \eta(X, Y) \leq 1, \tag{3.5}$$

so the Hellinger correlation places pairwise dependence strengths on a common (comparable) scale. As discussed above, it reaches 1 only under perfect dependence (i.e., when X and Y can be "generated" by a *single* uniform random variable U), and its value remains unchanged under rescaling or strictly monotonic transformations of the marginals, such as the logarithmic transform (often used in actuarial workflows). Another advantage of the Hellinger correlation is that, since it is formulated in terms of the copula, it does not require specific moments to be finite.[5] This is mainly a theoretical advantage, as it ensures that the measure remains well

[4] Recall that the joint density of two continuous random variables X and Y can be written as $f(x, y) = c(F(x), G(y))f(x)g(y)$, where $F(x) \equiv u$ and $G(y) \equiv v$ are the ranks. This illustrates well how focusing on $c(u, v)$ takes the margins ($f(x)$ and $g(y)$) 'out of the equation'.

[5] Remember that Pearson's correlation coefficient requires the first two moments of both distributions to be finite to exist.

defined even in heavy-tailed settings. In the next section, we provide illustrations on the use of the Hellinger correlation.

3.1.2 Illustrations

We illustrate both the estimation and testing aspects of the Hellinger correlation using the `R` package `HellCor` (Geenens & Lafaye de Micheaux, 2020). For two univariate random variables $X, Y \in \mathbb{R}$, the Hellinger correlation serves not only as a measure of dependence, but also as a test statistic for

$$\begin{aligned} H_0 &: X \perp Y, \\ H_1 &: X \not\perp Y. \end{aligned} \tag{3.6}$$

The associated test approximates the null distribution of the empirical Hellinger correlation by Monte Carlo simulation from the independence copula; details are provided in Appendix E.2. In practice, the procedure is fully data-driven and has been implemented efficiently in `HellCor`, although the computational cost increases with the sample size. Both the estimate and its associated p-value can be obtained with `HellCor::HellCor`, with the option `pval.comp = TRUE` used to compute the latter. Below, we illustrate this procedure using the example introduced in Section 2.

Birth–death We first revisit the birth and death rate example shown in Figure 2.1. As discussed in Section 2, the Pearson correlation coefficient between birth rate and death rate is not statistically significant. In contrast, the Hellinger correlation reveals a clear dependence structure. In this example, the Hellinger correlation is 0.6903, indicating moderate to strong dependence. This is further supported by the hypothesis test in Equation (3.6), which gives a p-value of less than 0.01, leading to rejection of the null hypothesis of independence. Hence, the Hellinger correlation is able to detect the nonlinear dependence between birth rate and death rate that Pearson correlation fails to capture.

Multi-line simulated data We next return to the multi-line example in Section 2. There, Pearson's correlation coefficient failed to detect the dependence structure between claims handling cost (X_3) and medical claim costs (Y), even though Figure 2.3 displays a clear "cross" pattern for (X_3, Y). Here we apply the Hellinger correlation to the dataset to illustrate its ability to detect such non-linear, non-comonotonic dependence structures.

Table 3.1 reports the Hellinger correlation matrix for all pairs of variables in the motivating example. Focusing on the cross-line pairs (X_i, Y) for $i = 1, 2, 3$, we observe that the Hellinger correlation for (X_3, Y) is close to 1, indicating very strong dependence between these pairs. This demonstrates the ability of the Hellinger correlation to detect non-linear and non-monotonic dependence between continuous random variables.

$\eta(X_i, Y)$	X_1	X_2	X_3	Y
X_1	1.0000	0.0345	0.0139	0.0387
X_2	0.0345	1.0000	0.0191	0.0201
X_3	0.0139	0.0191	1.0000	0.9975
Y	0.0387	0.0201	0.9975	1.0000

Table 3.1: Hellinger correlation matrix for the multi-line simulated data. The pairwise Hellinger correlations are all close to zero, except for the pair (X_3, Y), which is close to 1, suggesting a strong dependence between them. This is consistent with the scatterplots in Figure 2.3: there is no obvious pattern for the pairs (X_1, Y) and (X_2, Y), whereas the pair (X_3, Y) exhibits a clear cross pattern.

However, as noted in Section 2, X_1, X_2, and Y are not mutually independent, as shown in Figure 2.4a. At the same time, the Hellinger correlation is close to zero for each pair among them, and the corresponding pairwise tests also fail to reject independence. This is not a failure of the Hellinger correlation, since these pairs are constructed to be pairwise independent. Rather, it points to an important limitation of the Hellinger correlation: it is designed to measure dependence only between a pair of continuous random variables, not between random vectors or among more than two random variables. We discuss the limitations of the Hellinger correlation further in the next section.

3.1.3 Limitations

A practical limitation of the Hellinger correlation is that it is developed for *univariate continuous* random variables. Since the margin-free property relies on the copula representation in the continuous case, the Hellinger correlation defined above is naturally restricted to continuous marginals. This is an important limitation in actuarial settings, where dependence may involve discrete variables or random vectors, such as claim counts in the collective risk model framework, the number of sudden braking events in telematics insurance, or categorical rating factors (e.g., gender, occupation, or vehicle type), which are often represented as discrete indicators or one-hot encoded vectors.

It is also important to acknowledge that, as noted by Geenens and Lafaye de Micheaux (2022), the Hellinger correlation is not primarily designed as a test statistic for independence. Its main role is to quantify dependence strength, in a way that makes it comparable across cases (independently of their marginal distributions). This does not necessarily mean that it is not powerful as a test statistic in practice; its power will depend on the underlying dependence structure and the sample size (Serdar et al., 2021). Since measuring dependence strength and testing independence are related but distinct objectives, it is natural to also consider statistics that are particularly suited for the latter purpose. This leads to distance covariance in the next section, which is especially convenient for independence testing and, in addition, applies to random vectors of any dimensions and types.

3.2. Distance covariance

As discussed in Section 1.1, detecting association between random vectors is useful in many actuarial applications, since the relevant objects are often naturally multivariate. Examples include risk aggregation for capital and solvency purposes, reserving across multiple lines of business, and fairness considerations in insurance pricing with machine learning algorithms, where categorical protected features may be one-hot encoded into a random vector (Cai et al., 2025; Grari et al., 2022; International Association of Insurance Supervisors, 2024; Kearns et al., 2018; H. M. Lee et al., 2025; Shi, 2014). For this reason, we introduce another distance-based dependence statistic, distance covariance, which is particularly useful when the objects of interest for dependence analysis are random vectors.

3.2.1 Definition

Distance covariance (Székely et al., 2007) provides a useful tool for testing the independence between two random variables/vectors. For two random variables or random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ of arbitrary dimension and of various types (e.g. continuous, discrete, or one-hot encoded categorical), one can quantify the distance between their joint characteristic function $\varphi_{\boldsymbol{XY}}$ and the product of their marginal characteristic functions $\varphi_{\boldsymbol{X}}\varphi_{\boldsymbol{Y}}$ via the following formula

$$\|\varphi_{\boldsymbol{XY}}(\boldsymbol{s},\boldsymbol{t}) - \varphi_{\boldsymbol{X}}(\boldsymbol{s})\varphi_{\boldsymbol{Y}}(\boldsymbol{t})\|_w^2 = \int_{\mathbb{R}^{p+q}} |\varphi_{\boldsymbol{XY}}(\boldsymbol{s},\boldsymbol{t}) - \varphi_{\boldsymbol{X}}(\boldsymbol{s})\varphi_{\boldsymbol{Y}}(\boldsymbol{t})|^2 \, w(\boldsymbol{s},\boldsymbol{t}) \, d\boldsymbol{s} \, d\boldsymbol{t}, \tag{3.7}$$

using a weight function $w(\cdot, \cdot)$ to ensure that the integral is finite. Think of the characteristic function as one unique way of specifying a distribution (similar to moment-generating functions) which satisfies $\varphi_{\boldsymbol{XY}} = \varphi_{\boldsymbol{X}}\varphi_{\boldsymbol{Y}}$ if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent; further details on characteristic functions are provided in Appendix F.1. Although distance covariance is defined through characteristic functions, in practice the sample estimator is not obtained by numerically evaluating the integral in (3.7) directly. Instead, standard implementations use an equivalent formulation based on pairwise Euclidean distance matrices, which leads to a practical and readily computable estimator; see Appendix F.1 for details. Several weight functions have been considered. For instance, Bilodeau and Lafaye de Micheaux (2005), under the assumption of Gaussian marginals, used a Gaussian weight function and proposed a semi-parametric test of independence (or serial independence) between random vectors, extending earlier results of Deheuvels (1983) and Feuerverger (1993). Székely et al. (2007) adapted these ideas to the non-serial setting by using a special weight function, and Székely and Rizzo (2012) appear to suggest that this weight should be preferred. However, Fan et al. (2017) studied alternative weight functions and argued that, in a hypothesis testing context, the "best" choice of weight depends on the alternative distribution. Székely et al. (2007) denoted (3.7) by $dCov^2(\boldsymbol{X}, \boldsymbol{Y})$ and called the square root of this quantity distance covariance. In this paper, we adopt the weight function used by Székely et al. (2007), since it is the standard choice in the literature and is also the one implemented in commonly used `R` packages. Again, details of this weight function are provided in Appendix F.1.

Figure 3.2 displays the real part of the joint characteristic function and the product of the marginal characteristic functions for two random variables with standard Gaussian marginals. Under independence (left panel), the two characteristic functions coincide, so their difference is zero, resulting in zero distance covariance. This is consistent with the interpretation of distance covariance as a distance between the joint characteristic function and the product of the marginal characteristic functions, as discussed in Section 1.2. In contrast, when the two Gaussian marginals are coupled through a Clayton copula with parameter $\theta = 5$, as in Equation (3.3) and Figure 3.1, the variables are no longer independent, and the two characteristic functions no longer coincide. Distance covariance captures this discrepancy through the gap between the two functions, illustrated by the shaded area in the right panel. In addition, Székely et al. (2007) also defined a standardised version of the distance covariance, called the distance correlation, which is given by

$$0 \leq dCor^2(\boldsymbol{X}, \boldsymbol{Y}) = \frac{dCov^2(\boldsymbol{X}, \boldsymbol{Y})}{\sqrt{dCov^2(\boldsymbol{X}, \boldsymbol{X}) dCov^2(\boldsymbol{Y}, \boldsymbol{Y})}} \leq 1, \tag{3.8}$$

but contrary to the Hellinger correlation, which is copula-based (and hence margin-free), distance correlation is defined directly from the variables rather than from their copula. Its magnitude therefore does not reflect the dependence structure alone, and is not interpretable as a measure of dependence strength. We leave the detailed definition and estimation of distance covariance to Appendices F.1 and F.2.

By quantifying the distance between characteristic functions, distance covariance can capture general forms of association, including non-linear and non-monotonic dependence between two random variables/vectors. This is an advantage over the association coefficients discussed in Section 1. Most importantly, this characterises independence, that is,

$$dCov^2(\boldsymbol{X}, \boldsymbol{Y}) = 0 \iff \boldsymbol{X} \text{ and } \boldsymbol{Y} \text{ are independent.} \tag{3.9}$$

The notion of independence between two random vectors was introduced in Section 2.3. This property follows from the fact that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent if and only if their joint characteristic function coincides with the product of their marginal characteristic functions as explained earlier. Moreover, distance covariance is defined for real-valued random vectors of arbitrary dimension.

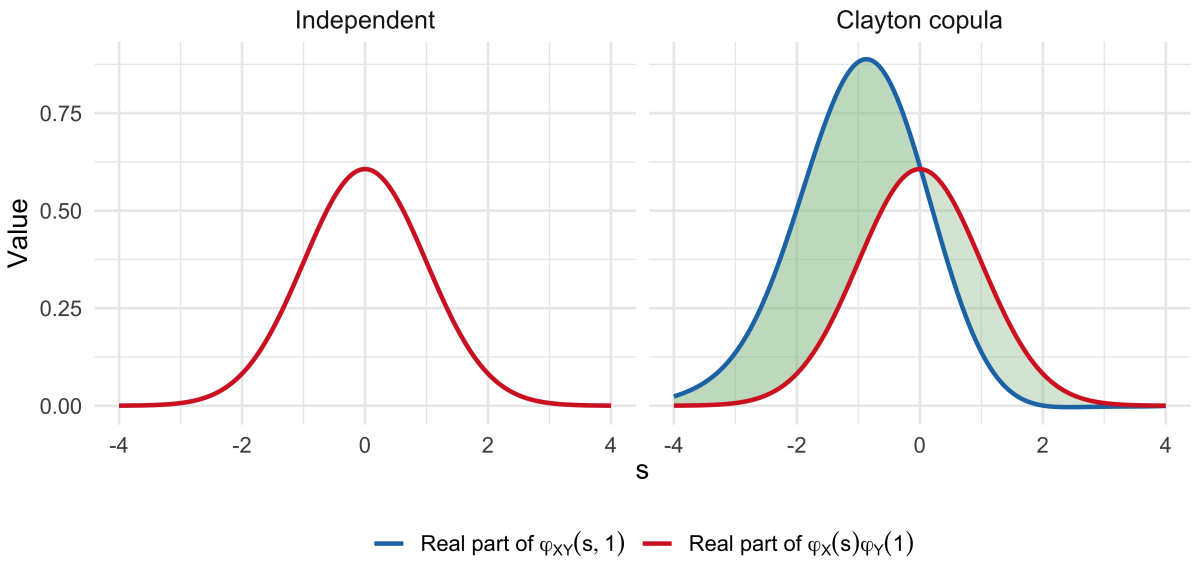


Figure 3.2: This figure compares the real part of the joint characteristic function (blue) and the real part of the product of the marginal characteristic functions (red), with t fixed at 1, for two random variables with standard Gaussian marginals. The left panel shows the independent case, for which $\varphi_{XY}(s,1) = \varphi_X(s)\varphi_Y(1)$, so the two curves coincide. The right panel shows the dependent case, where the Gaussian marginals are coupled through the same Clayton copula with parameter $\theta = 5$ as in (3.3) and Figure 3.1. In this case, the joint characteristic function no longer coincides with the product of the marginals, and the shaded region illustrates what distance covariance is designed to capture. Note that a larger shaded area does not necessarily indicate stronger dependence, as distance covariance is defined directly from the variables rather than from their copula, so its magnitude does not reflect the dependence structure alone.

This is particularly useful, since many quantities of interest in actuarial applications are naturally multivariate, as discussed in Section 2.3. In the next section, we illustrate the use of distance covariance through the motivating examples introduced earlier.

3.2.2 Illustrations

For illustration, we make use of the `R` package `dcov` (Hang, 2020) to compute distance covariance and distance correlation. In particular, the functions `dcov::dcov` and `dcov::dcor` are used to compute these statistics, while `dcov::dcor.test` is used to perform a permutation test based on distance correlation under the null hypothesis in Equation (3.6). In practice, the computational cost increases mainly with the sample size, and permutation testing further increases the runtime because the statistic must be recomputed over many random permutations of the data, with the number of permutations specified by the user. Further details on the implementation and the computational cost of computing the statistic and carrying out the permutation test are provided in Appendix F.3.

Birth–Death We again consider the birth rate versus death rate example from Section 2. We compute both the distance covariance and the distance correlation between birth rate and death rate. The distance covariance is 1.2321. However, direct interpretation of distance covariance is less informative, since it is not scale-free. We therefore turn to the distance correlation, which is equal to 0.1255. Even so, this value is interpretable only to the extent that we know it will be between 0 and 1, but it is not comparable to other pairs, because unlike the Hellinger

correlation, distance correlation is not margin-free. We discuss this limitation of distance-based dependence measures further in Section 3.2.3. Hence, we focus primarily on the test result under the hypotheses stated in Equation (3.6). The corresponding permutation test gives a p-value of < 0.0002, leading to a rejection of the null hypothesis of independence. This result is consistent with our findings in Section 3.1.2 and shows that both the Hellinger correlation and the distance correlation are able to detect the nonlinear dependence between birth rate and death rate.

Multi-line simulated data. We continue with our motivating example involving multiple lines of business. Tables 3.2 and 3.3 display the distance correlation matrix, and the matrix of permutation-test p-values for independence for every pair of variables under the hypotheses

$$\begin{aligned} H_0 &: X_i \perp Y, \\ H_1 &: X_i \not\perp Y, \end{aligned} \tag{3.10}$$

for $i = 1, 2, 3$. When we inspect the distance correlation, all values are close to zero except for the pair (X_3, Y), for which the distance correlation is 0.0913, substantially larger than for the other pairs. In particular, the dependence between (X_3, Y) is exactly the relationship that correlation fails to detect. A similar conclusion is supported by the corresponding permutation-test p-values, which lead us to reject the null hypothesis of independence for (X_3, Y) (with a p-value $<$0.0020) and therefore indicate dependence between the two variables. However, despite being the largest pairwise value, a distance correlation of 0.0913 may still appear numerically small. If a value of this magnitude arose for Pearson or Hellinger correlation, it would typically suggest weak dependence, which is not consistent with the clear cross-shaped pattern in Figure 2.3. This further illustrates that distance correlation is more useful for *detecting* dependence (as a binary answer yes/no) than for interpreting the strength of dependence.

$dCor^2(X_i, Y_j)$	X_1	X_2	X_3	Y
X_1	1.0000	0.0009	0.0004	0.0006
X_2	0.0009	1.0000	0.0004	0.0008
X_3	0.0004	0.0004	1.0000	0.0913
Y	0.0006	0.0008	0.0913	1.0000

Table 3.2: Distance correlation matrix for the multi-line simulated data. Similar to the Hellinger correlation, the distance correlations are all close to 0, except for the pair (X_3, Y), which is substantially larger than the others. This indicates little pairwise dependence for (X_1, Y) and (X_2, Y), but a noticeably stronger dependence for (X_3, Y).

p-value	X_1	X_2	X_3	Y
X_1	1.0000	0.1377	0.8423	0.4491
X_2	0.1218	1.0000	0.8643	0.2136
X_3	0.8483	0.8543	1.0000	$<$0.0020
Y	0.3852	0.1457	$<$0.0020	1.0000

Table 3.3: Permutation-test p-value matrix based on distance covariance for the motivating example. The p-value is significant only for the pair (X_3, Y), which is consistent with Figure 2.3, where only (X_3, Y) exhibits an obvious cross pattern.

As discussed earlier, distance covariance has the additional advantage of being able to detect dependence between random vectors of arbitrary dimension and type. This is an important advantage over both correlation and the Hellinger correlation. As illustrated in Figure 2.4a and

2.4b, there is clear dependence involving the full collection (X_1, X_2, Y). We therefore compute the distance covariance, distance correlation, and permutation-test p-value between the random vector (X_1, X_2) and the scalar Y. The null hypothesis is

$$\begin{aligned} H_0 &: (X_1, X_2) \perp Y, \\ H_1 &: (X_1, X_2) \not\perp Y. \end{aligned} \tag{3.11}$$

The distance correlation between (X_1, X_2) and Y is given by 0.0384 with a p-value of $<$0.0020, indicating a significant relationship between them. Using distance covariance/correlation, we detect joint dependence between the random vector (X_1, X_2) and Y that is not captured by statistics limited to pairwise associations between individual variables. Note that the vector (X_1, X_2) is constructed by concatenation, as in the $CCdCov$ framework of H. M. Lee et al. (2025); see also Section 4.1.

3.2.3 Limitations

As we can see in the previous illustrations, a practical consideration is that distance covariance itself is not scale-free: if we rescale the units of $\boldsymbol{X}$ or $\boldsymbol{Y}$ (e.g., measuring losses in dollars rather than thousands of dollars), the value of distance covariance changes. In practice, this issue can be addressed by reporting the distance correlation, which is normalised to lie in $[0, 1]$. However, some care is still needed when interpreting distance correlation as a general-purpose measure of dependence strength. As noted by Geenens and Lafaye de Micheaux (2022), distance correlation is still not margin-free. In addition, for two real-valued univariate random variables X and Y, the statistic attains its maximum only in the case of perfect linear dependence (Edelmann et al., 2021), that is, when $Y = aX + b$ for some real numbers $a \neq 0$ and b. This should be kept in mind, since perfect dependence may also arise in non-linear forms, and distance covariance may not be able to reflect this by attaining its maximum.

Hence, we view distance covariance primarily as a useful tool for testing independence between random vectors, rather than as a unified and directly interpretable measure of association strength for reporting across settings. Instead, the Hellinger correlation detailed in Section 3.1 is typically fit for a universal measure of dependence strength between two random variables.

With that said, two additional remarks are worth adding. First, similar to the correlation coefficient, distance covariance is defined under the assumption that the first moments of both random vectors are finite. In principle, this condition may fail under heavy-tailed models (Embrechts et al., 1997). In such cases, distance covariance is not guaranteed to be finite or to converge, and the characterisation that "independence holds if and only if distance covariance is 0" may no longer apply, since it relies on these moment conditions. However, this is more of a technical condition and is likely not a practical limitation in most standard actuarial applications.

Moreover, the estimation of distance covariance is not robust to extreme outliers (Leyder et al., 2026). It can be driven to a high value by a single, or a few, extreme observations, even when the two random variables are independent at the population level. This limitation is not unique to distance covariance and also applies to the correlation coefficient, but it is still worth noting.

3.3. Choosing between Hellinger correlation and distance covariance

The Hellinger correlation and distance covariance serve different primary purposes. When the goal is to obtain an interpretable summary of dependence strength between two continuous univariate random variables, the Hellinger correlation is typically the more appropriate choice. Because it is copula-based, it is margin-free, and its values lie in $[0, 1]$, which makes it convenient for reporting and comparing dependence across settings.

In contrast, distance covariance is typically more useful when the main objective is to detect dependence, especially when random vectors are involved. Its main advantage is that it applies directly to random vectors of arbitrary dimension, so it can detect dependence between a scalar and a vector, or between two random vectors, without reducing the problem to separate pairwise comparisons. This is particularly useful in actuarial applications, where the relevant objects are often naturally multivariate or of mixed type.

In the case of testing independence between two continuous univariate random variables, it may be useful to consider both statistics. Although the Hellinger correlation is more naturally interpreted as a measure of dependence strength, one cannot single out one of the two as always preferred for testing in this setting. Their power may differ across dependence structures, so considering both can provide a more complete assessment of whether dependence is present.

A common feature of both the Hellinger correlation and distance covariance is that they are *non-negative*. Consequently, unlike Pearson's correlation coefficient, they do not provide a sign or direction of association. This is natural, since for non-linear or non-monotonic dependence it is often unclear what a "direction" should mean.

Finally, note that different dependence structures may yield similar values of the same statistic. For this reason, numerical summaries are often best complemented by graphical diagnostics, such as the 2D and 3D coloured plots used in our earlier illustrations.

4. Multivariate dependence

In Section 3, we introduced the Hellinger correlation and distance covariance as two complementary distance-based statistics. The Hellinger correlation is well suited for measuring and reporting the strength of dependence between two univariate continuous variables, while distance covariance is particularly useful for testing independence between *two* random vectors. When multiple random variables are involved, one may concatenate them into random vectors and assess dependence between the resulting vector pair, as in the multi-line simulated example in Section 3.2.2; see again H. M. Lee et al. (2025) and Section 4.1. In practice, however, the relevant grouping into two random vectors may be unclear, and dependence may remain hidden if it is not captured by the chosen grouping. More generally, any bivariate dependence measure, even when applied to suitably chosen random vectors, only assesses dependence between two specified objects and therefore does not in general resolve the broader question of mutual independence. This is particularly relevant in actuarial applications where modelling assumptions are often stated in terms of mutual independence. We therefore introduce a distance-based dependence statistic that assesses higher-order dependence directly among multiple random variables or vectors, namely the joint distance covariance (Chakraborty & Zhang, 2019).

4.1. Joint distance covariance

As shown in Equation (2.3), mutual independence of d random vectors means that their joint probability distribution coincides with the product of their marginal distributions. An analogous characterisation holds in terms of characteristic functions: for random vectors $\boldsymbol{X}_j \in \mathbb{R}^{p_j}$, $j \in \{1, \ldots, d\}$, mutual independence is equivalent to

$$\varphi_{(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_d)}(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_d) = \prod_{j=1}^{d} \varphi_{\boldsymbol{X}_j}(\boldsymbol{t}_j), \qquad \forall\, \boldsymbol{t}_j \in \mathbb{R}^{p_j},\ j = 1, \ldots, d. \tag{4.1}$$

In the case $d = 2$, distance covariance is already constructed by comparing the joint characteristic function with the product of the marginal characteristic functions; see Equation (3.7). Motivated by this same idea, Fan et al. (2017) and Chakraborty and Zhang (2019) extended the original construction to the setting with $d > 2$. Chakraborty and Zhang (2019) first define the

kth-order distance covariance by

$$dCov^2(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k) = \int_{\mathbb{R}^{\sum_{j=1}^k p_j}} \left| \mathbb{E}\left[\prod_{j=1}^{k} \left(\varphi_{\boldsymbol{X}_j}(\boldsymbol{t}_j) - e^{i\boldsymbol{t}_j^\top \boldsymbol{X}_j} \right) \right] \right|^2 w(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_k)\, d\boldsymbol{t}_1 \cdots d\boldsymbol{t}_k, \quad (4.2)$$

where w is the weight function analogous to that of Equation (3.7) Székely et al. (2007). When $k = 2$, this reduces to the distance covariance defined in Equation (3.7). Extending this idea, the joint distance covariance for d random vectors is given by

$$\begin{aligned} JdCov^2(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d) = &\sum_{(j_1,j_2)\in\mathcal{J}_2^d} dCov^2(\boldsymbol{X}_{j_1}, \boldsymbol{X}_{j_2}) + \sum_{(j_1,j_2,j_3)\in\mathcal{J}_3^d} dCov^2(\boldsymbol{X}_{j_1}, \boldsymbol{X}_{j_2}, \boldsymbol{X}_{j_3}) \\ &+ \cdots + dCov^2(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d), \end{aligned} \quad (4.3)$$

where $\mathcal{J}_k^d = \{(\tau_1, \ldots, \tau_k) \in \{1, \ldots, d\}^k : \tau_1 < \tau_2 < \cdots < \tau_k\}$ denotes the collection of increasing index tuples of size k chosen from $\{1, \ldots, d\}$. Further technical details on the definition (e.g., the weight function w) and the estimation of the joint distance covariance are provided in Appendices G.1 and G.2. Fan et al. (2017) employed a different strategy using the Mobius transformation, and considered various weights as well. We do not consider that approach here, since the weight function used by Chakraborty and Zhang (2019) naturally extends the original definition of distance covariance and provides an estimator with a construction similar to that of distance covariance.

The formulation of joint distance covariance (JdCov) consists of an exhaustive sum of kth-order distance covariance terms for $k = 2, \ldots, d$. For each k, the corresponding term aggregates the dependence contributions associated with all subsets of size k among the d random vectors. Thus, $JdCov$ combines dependence information across all orders, from pairwise dependence up to the full "d-way" dependence structure.

For example, when $d = 3$, the 3rd-order distance covariance considers only the dependence structure that is visible when $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_3$ are considered jointly. It does not separately account for the pairwise associations among the three random vectors. As a result, even if the 3rd-order distance covariance is zero, this does not imply pairwise independence for every pair of random vectors. Conversely, the pairwise distance covariance terms may all be zero even when the 3rd-order distance covariance is nonzero, that is, when dependence is present only considering the three random vectors together. We illustrate this further with an example in Appendix G.4. This is precisely why $JdCov$ combines all 2nd- and 3rd-order distance covariance terms (and so on), so that it provides an exhaustive quantity of dependence among the three random vectors.

Figure 4.1 provides a clearer view of the associations captured by the joint distance covariance. It illustrates the dependence structures captured by the d-th order distance covariance for four random vectors $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, and $\boldsymbol{X}_4$. For $d = 2$ (top left), it captures all pairwise dependence, namely the dependence between the six pairs of random vectors formed from $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4)$, represented by differently coloured edges. The 3rd-order distance covariance (top right) then captures dependence across triples of random vectors, so there are four 3-way dependence terms to consider, represented by the coloured faces of the tetrahedron. The 4th-order distance covariance (bottom left) captures the dependence that is visible only when all four random vectors are considered jointly, represented by the tetrahedron itself. Finally, the joint distance covariance (bottom right) combines the dependence contributions from all orders.

Similar to the distance covariance, $JdCov$ is able to characterise mutual independence at 0 among d random vectors for $d \geq 2$ (arguably a very strong requirement). It is also able to capture both nonlinear and non-monotonic dependence, as the bivariate distance covariance does, even in cases where the random vectors/variables are all pairwise independent.

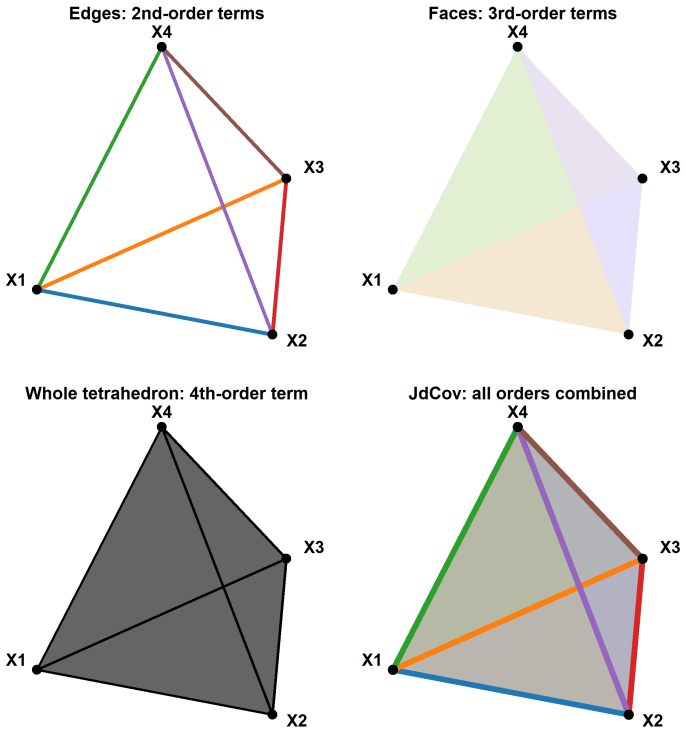


Figure 4.1: This figure provides an intuition for what the higher-order distance covariance terms and the joint distance covariance capture for four random vectors $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, and $\boldsymbol{X}_4$. The top-left panel represents the second-order terms (the edges), capturing the pairwise associations among all six pairs of random vectors. The top-right panel represents the third-order terms (the faces), capturing the associations among all four triplets of random vectors. The bottom-left panel represents the fourth-order term (the tetrahedron), capturing the association among all four random vectors jointly. The bottom-right panel illustrates the joint distance covariance, which combines the second-, third-, and fourth-order terms.

4.2. Illustrations

The illustration of the joint distance covariance is implemented in `python`, and the code is available on GitHub. Details of the corresponding hypothesis test for mutual independence based on joint distance covariance are provided in Appendix G.3. Alternatively, in `R`, the `jdcov` package (Chakraborty et al., 2019) provides functions for both estimation and testing of joint distance covariance.

Fairness gerrymandering We illustrate how joint distance covariance can be useful in the context of "fairness issues" in machine learning. Let $\hat{Y}$ denote a model prediction and let S denote a protected feature, such as gender or ethnicity. A common fairness criterion is *demographic parity* (Hardt et al., 2016; Mehrabi et al., 2021; Xin & Huang, 2024), which requires the model prediction to be independent of the protected feature. That is, the distribution of $\hat{Y}$ should not depend on the value of S. Mathematically, for every set A of possible values of $\hat{Y}$,

$$\mathbb{P}(\hat{Y} \in A \mid S) = \mathbb{P}(\hat{Y} \in A) \iff \hat{Y} \perp S. \tag{4.4}$$

In practice, however, there are often multiple protected features, say $S_1, \ldots, S_d$. Suppose

that a model satisfies demographic parity for each protected feature separately, that is,

$$\hat{Y} \perp S_k, \qquad k = 1, \ldots, d. \tag{4.5}$$

This does not guarantee fairness at the intersection of the protected features. For example, a model may satisfy demographic parity marginally for gender and race, while still being unfair for intersectional subgroups such as Caucasian females or African-American males. This phenomenon is known as *fairness gerrymandering* (Kearns et al., 2018), and arises from the fact that

$$\hat{Y} \perp S_k \text{ for all } k = 1, \ldots, d \quad \not\Longrightarrow \quad \hat{Y} \perp (S_1, \ldots, S_d). \tag{4.6}$$

This makes fairness gerrymandering a natural context in which to illustrate the use of joint distance covariance. Indeed, the issue is precisely that marginal independence with each protected feature does not imply mutual independence when the protected features are considered jointly. To capture this, H. M. Lee et al. (2025) proposed using

$$JdCov^2(\hat{Y}, S_1, \ldots, S_d) \tag{4.7}$$

as a regulariser. If this regularisation is made sufficiently strong, then theoretically we obtain

$$JdCov^2(\hat{Y}, S_1, \ldots, S_d) = 0, \tag{4.8}$$

which implies mutual independence among the model prediction and all protected features. In particular,

$$\hat{Y} \perp (S_{i_1}, S_{i_2}, \ldots, S_{i_k}) \qquad \forall \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, d\},\ 1 \leq k \leq d. \tag{4.9}$$

Hence, the model prediction is independent of any subset of the protected features, including each protected feature marginally. This directly addresses the fairness gerrymandering issue, since it enforces fairness not only for each protected feature separately but also for all intersectional subgroups.

At the same time, this example also reveals an important feature of the joint distance covariance. Since $JdCov$ captures the full mutual dependence structure among all inputs, the condition in Equation (4.8) also depends on the dependence structure among the protected features themselves. In particular, Equation (4.8) can hold only when the protected features are mutually independent. Thus, in this application, $JdCov$ is valuable because it highlights the full mutual dependence structure underlying fairness gerrymandering, but it may be too strong if the modelling goal is only to enforce independence between the prediction and the protected attributes.

For this reason, H. M. Lee et al. (2025) also introduced the concatenated distance covariance as a complementary regulariser. Since distance covariance can be applied to random vectors, we may concatenate the protected features into a single random vector $(S_1, \ldots, S_d)$, as in the multi-line simulated example in Section 3.2.2, and use

$$CCdCov(\hat{Y}, (S_1, \ldots, S_d)) = dCov^2(\hat{Y}, (S_1, \ldots, S_d)) \tag{4.10}$$

as the regulariser. If the regularisation strength is sufficiently large, then we obtain

$$dCov^2(\hat{Y}, (S_1, \ldots, S_d)) = 0, \tag{4.11}$$

which can be achieved by adjusting $\hat{Y}$ alone, without imposing any independence requirement on the protected features themselves. By the defining property of distance covariance,

$$dCov^2(\hat{Y}, (S_1, \ldots, S_d)) = 0 \iff \hat{Y} \perp (S_1, \ldots, S_d). \tag{4.12}$$

This in turn implies (4.9). Hence, the model prediction is again independent of any subset of

the protected features, including each protected feature marginally.

We illustrate this more clearly using the dataset `pg15training` from the `R` package `CASdatasets` (Dutang & Charpentier, 2026), which comprises 100,000 third-party liability policies for a private motor insurance product in France. Details of the dataset are provided in Appendix G.5. We consider two protected features, namely `Region`, which reflects the residential region of the policyholder and may proxy for ethnicity or race, and `Female`, which equals 1 if the policyholder is female and 0 if male. Here, we fit three Poisson regression models: one without regularisation, one with joint distance covariance,

$$JdCov^2(\hat{Y}, \texttt{Region}, \texttt{Female}),$$

and one with

$$CCdCov(\hat{Y}, (\texttt{Region}, \texttt{Female}))$$

as the regulariser, where the model prediction $\hat{Y}$ is the predicted claims frequency. Figure 4.2 displays the distributions of the predicted claim frequency across different gender-region groups, using kernel density estimates in the top row and empirical cumulative distribution functions in the bottom row. As we can see, when there is no regularisation (left column), the predicted claim frequencies across gender-region groups are more dispersed, indicating that the distribution of model predictions is not independent of the protected features. However, when we fit the model with regularisation, the prediction distributions become much closer one to another. This is consistent with the role of the regularisers. For $CCdCov$, smaller values correspond to weaker dependence between $\hat{Y}$ and the concatenated protected-feature vector $(\texttt{Region}, \texttt{Female})$, and the limiting case $CCdCov(\hat{Y}, (\texttt{Region}, \texttt{Female})) = 0$ implies independence. For $JdCov$, smaller values indicate weaker overall mutual dependence among $\hat{Y}$, `Region`, and `Female`, although the condition $JdCov^2(\hat{Y}, \texttt{Region}, \texttt{Female}) = 0$ is stronger, since it also depends on the dependence structure between the protected features themselves. Theoretically, with sufficiently strong regularisation, these distributions would overlap as the joint distance covariance (or similarly, $CCdCov$) is driven to zero, so that the model predictions become independent of the protected groups; that is, the prediction distributions remain the same across all protected groups.

We refer the reader to H. M. Lee et al. (2025) for further details and additional illustrations of this setting.

4.3. Limitations

Again, since $JdCov$ is built on the distance covariance, it shares similar limitations. First, finite first moments of the random variables are required. In addition, it does not indicate which subset of variables drives the dependence, nor what structure that dependence has. Also, it is not margin-free, as the formulation does not involve the copula of the random variables. Hence, $JdCov$ is not invariant under monotonic transformations applied to any of the random variables or vectors. Even though there is a copula version of distance multivariance (discussed in Remark 4.1), which is a generalised version of $JdCov$ formulated using copulas (Böttcher, 2020), it may still not be a suitable measure. As discussed in Böttcher (2020), the copula distance multivariance is not directly interpretable, in the sense that its values do not correspond to a clear strength or structure of mutual association. Also, similar to the limitation of the Hellinger correlation, the copula distance multivariance does not support quantifying the association when any of the random vectors/variables is discrete. Hence, $JdCov$ is best used as a test for the existence of mutual association, rather than as a universal measure. Since our main purpose is testing mutual independence rather than interpreting the magnitude of $JdCov$ as a dependence measure, differences in scale among the kth-order terms are less problematic here: what matters is whether the resulting statistic differs from zero.

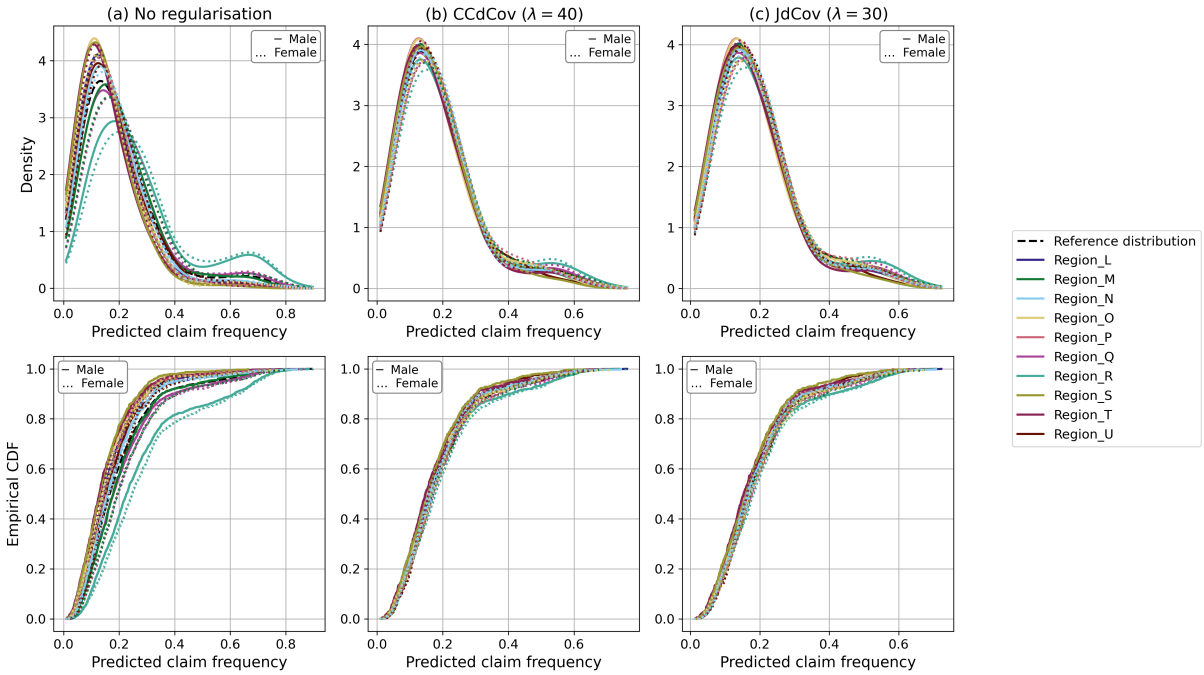


Figure 4.2: Predicted claim frequency distributions using the `pg15training` data under three regularisation settings: (a) no regularisation, (b) regularisation using $CCdCov$, and (c) regularisation using $JdCov$. The top row shows kernel density estimates and the bottom row shows the corresponding empirical CDFs. Each colour corresponds to a specific region, and line styles distinguish gender (solid for male and dashed for female). Without regularisation, the predicted claim frequency distributions are more dispersed across gender-region groups. With either regularisation, the group-specific distributions move closer together, consistent with weaker dependence between the model predictions and the protected features.

Remark 4.1 (Distance multivariance)**.** *Distance multivariance was developed by Böttcher et al. (2019) (borrowing substantial material from Fan et al., 2017) around the same time that Chakraborty and Zhang (2019) developed the joint distance covariance. The main difference between the two measures is that, in distance multivariance, the weight function (w in Equations* (3.7) *and* (4.2)*) is replaced by a Lévy measure. With this replacement, instead of computing the Euclidean norm between the joint characteristic function and the product of the marginal characteristic functions, the choice of Lévy measure allows this Euclidean norm to be replaced by other distance metrics, such as the Minkowski distance. Hence, the joint distance covariance can be viewed as a special case of distance multivariance.*

The reason we choose the joint distance covariance is that it provides a natural extension of the original distance covariance, with a similar choice of the weight function w. In addition, under the analogous weight function, the estimator of the joint distance covariance can also be computed using the $\mathcal{U}$-distance matrix.

5. Time series modelling and diagnostics

Distance-based dependence statistics are not only useful for measuring and testing nonlinear and non-monotonic association between random variables or random vectors; there are also several extensions that can be useful in other actuarial workflows. In this section, we introduce one of the extensions, which is the auto-distance correlation function, and illustrate how it can be used in practice.

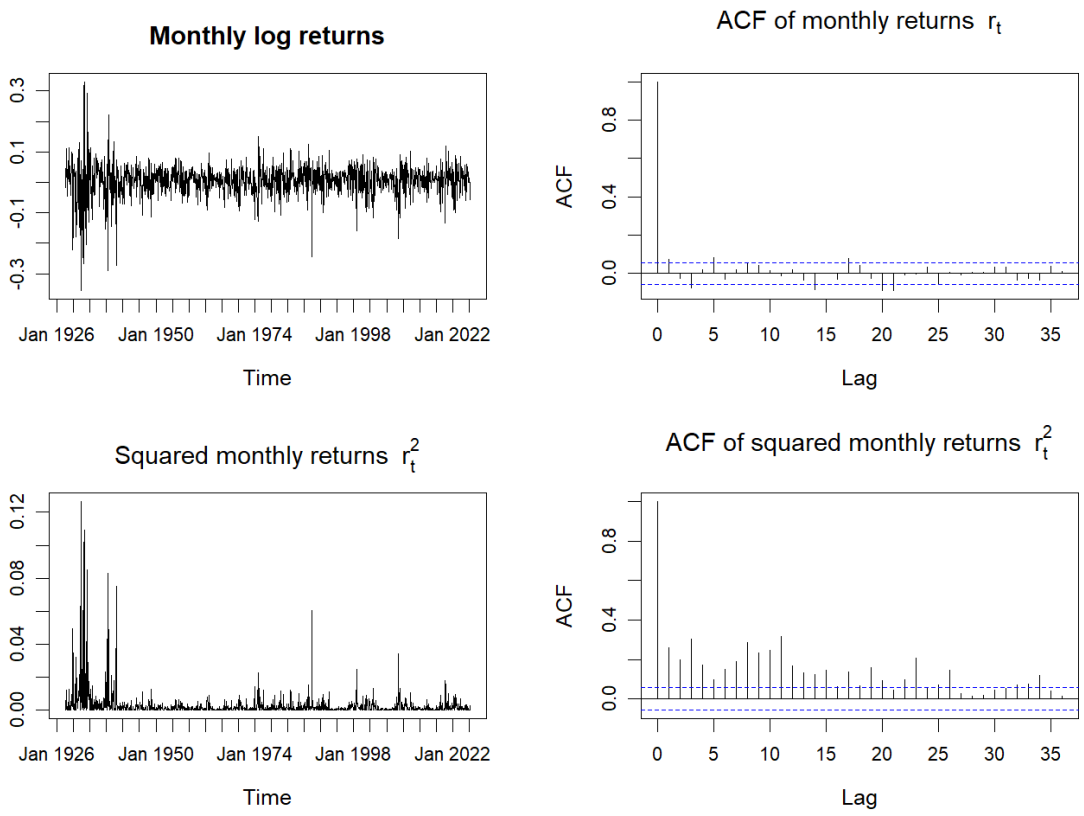


Figure 5.1: S&P-500 monthly log returns and squared log returns from 1926 onward, together with their sample autocorrelation functions. While the ACF of the log returns is insignificant or only marginally significant at most nonzero lags, the squared returns exhibit clear serial dependence. This illustrates that weak linear dependence, as indicated by the ACF, does not rule out nonlinear time dependence.

5.1. Background

Time series modelling has been an important tool in the actuarial community, with wide applications across different fields. For instance, the Lee–Carter model in mortality modelling requires fitting a time series and projecting based on the fitted time series (R. Lee, 2000; Li et al., 2023; Macdonald et al., 2007; Peters et al., 2021); time series methods have also been used in claims reserving (Atherino et al., 2010; Chukhrova & Johannssen, 2021; Usman & Chan, 2022), modelling customers' lapse behaviours (Lai & Lu, 2017), and many other actuarial applications.

Many of these applications rely on the autocorrelation function (ACF). In particular, the ACF is mainly used as a diagnostic tool for model specification, model selection, and assessing model fit. For a (weakly) stationary time series $\{X_t \mid t \in T\}$, where T is some time index set, the ACF at lag h is defined as

$$\rho(h) = \frac{\mathrm{Cov}(X_t, X_{t+h})}{\mathrm{Var}(X_t)}. \tag{5.1}$$

The ACF measures the linear serial dependence between observations that are h time steps apart. Since it is based on the correlation coefficient, it can only detect linear serial dependence. In fact, the time series literature has long noted that a near-zero ACF does not rule out serial dependence, since a process may be uncorrelated across lags but still exhibit nonlinear dependence (McLeod & Li, 1983). This is particularly relevant when dependence appears in higher moments, as in autoregressive conditional heteroskedasticity (ARCH) models (Engle,

1982). Figure 5.1 plots S&P-500 monthly stock returns downloaded from Yahoo Finance over the period 1926-01-01 to 2026-03-20. In particular, the ACF of the monthly log returns (top right) is insignificant for most lags, with only a few showing marginal significance. However, the ACF of the squared returns (bottom right) is clearly positive at small lags and appears to decay as the lag increases. This illustrates that the ordinary ACF may fail to detect serial dependence when the dependence is nonlinear.

For vector time series, the ACF can still be used for model diagnostics. However, an ACF matrix must be computed across the components of the vector time series for each lag h, which can be difficult to interpret. In particular, each lag yields a $p \times p$ matrix of auto- and cross-correlations for a vector time series of dimension p, so standard correlation-based diagnostics may require examining up to p^2 plots (Kathari & Tangirala, 2020). Moreover, these diagnostics inherit the same limitation as the univariate ACF, in that they may fail to detect nonlinear dependence structures.

5.2. (Multivariate) Auto-distance correlation

Motivated by the limitations of the original ACF, Zhou (2012) developed the auto-distance correlation function (ADCF), which is capable of detecting general serial dependence, including nonlinear dependence, in both univariate and vector time series. Let $\{\boldsymbol{X}_t \mid t \in T\}$ be a strictly stationary p-dimensional time series, with $\boldsymbol{X}_t = (X_{t;1}, X_{t;2}, \ldots, X_{t;p}) \in \mathbb{R}^p$. The ADCF at lag h is defined as the distance correlation between $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t+h}$:

$$\mathcal{R}_{\boldsymbol{X}}(h) = \frac{\mathcal{V}_{\boldsymbol{X}}(h)}{\mathcal{V}_{\boldsymbol{X}}(0)}, \quad \text{provided that } \mathcal{V}_{\boldsymbol{X}}(0) \neq 0, \tag{5.2}$$

where $\mathcal{V}_{\boldsymbol{X}}(h)$ is the auto-distance covariance, given by

$$\mathcal{V}^2_{\boldsymbol{X}}(h) = \int_{\mathbb{R}^p \times \mathbb{R}^p} \left|\varphi_{\boldsymbol{X}_t, \boldsymbol{X}_{t+h}}(\boldsymbol{s}, \boldsymbol{t}) - \varphi_{\boldsymbol{X}_t}(\boldsymbol{s})\varphi_{\boldsymbol{X}_{t+h}}(\boldsymbol{t})\right|^2 w(\boldsymbol{s}, \boldsymbol{t})\, d\boldsymbol{s}\, d\boldsymbol{t}, \tag{5.3}$$

which is defined analogously to the distance covariance in Equation (3.7), applied to the pair $(\boldsymbol{X}_t, \boldsymbol{X}_{t+h})$. For vector time series, the ADCF provides a compact scalar summary of serial dependence across lags, avoiding the need to inspect a $p \times p$ matrix of pairwise correlations at each lag. The ADCF inherits several attractive properties from distance correlation: it takes values in $[0, 1]$, and most importantly,

$$\mathcal{R}_{\boldsymbol{X}}(h) = 0 \iff \boldsymbol{X}_t \perp \boldsymbol{X}_{t+h}. \tag{5.4}$$

A particularly useful feature of the ADCF is that, for several important time-series model classes, its pattern across lags plays a role analogous to that of the classical ACF in linear settings. In Zhou's analysis, for vector ARMA-type linear models whose coefficients decay exponentially, the ADCF also decays exponentially with the lag. For finite-order moving-average-type models, the ADCF exhibits a cut-off pattern: in particular, Zhou's vector MA(1) example cuts off at lag 1, and more generally, a multivariate nonlinear moving average process of order m has ADCF that cuts off at lag m. Therefore, beyond detecting nonlinear serial dependence, the ADCF can also provide useful qualitative guidance for model specification.

In addition, Fokianos and Pitsillou (2018) introduced component-wise auto-distance covariance measures for multivariate time series. For components $r, m \in \{1, \ldots, p\}$ and lag j, define

$$\mathcal{V}^2_{rm}(j) = \int_{\mathbb{R}^2} \left|\varphi_j^{(r,m)}(u, v) - \varphi^{(r)}(u)\varphi^{(m)}(v)\right|^2 w(u, v)\, du\, dv, \tag{5.5}$$

where $\varphi_j^{(r,m)}(u, v)$ is the joint characteristic function of $X_{t;r}$ and $X_{t+j;m}$, and $\varphi^{(r)}(u)$ and $\varphi^{(m)}(v)$

are their corresponding marginal characteristic functions, with w defined analogously to Equation (5.3). The corresponding component-wise auto-distance correlation is

$$\mathcal{R}_{rm}(j) = \frac{\mathcal{V}_{rm}(j)}{\sqrt{\mathcal{V}_{rr}(0)\mathcal{V}_{mm}(0)}}, \quad \text{provided that } \mathcal{V}_{rr}(0)\mathcal{V}_{mm}(0) \neq 0. \tag{5.6}$$

This quantity measures the dependence between the r-th component at time t and the m-th component at time $t+j$. In particular, when $r = m$, $\mathcal{R}_{rm}(j)$ describes the serial dependence of the r-th component across lag j; when $r \neq m$, it captures the lagged cross-dependence between two different components. Hence, these component-wise measures provide a more detailed view of the dependence structure in a multivariate time series, complementing the scalar ADCF by showing which variables are associated with which lagged variables. The detail of the estimation for ADCF and multivariate ADCF is provided in Appendices H.1 and H.2 respectively.

5.3. Illustrations

For these illustrations, we use the `dCovTS` package (Tsagris et al., 2026) in `R`. The ADCF can be computed using `dCovTS::ADCF` and plotted using `dCovTS::ADCFplot`. In addition, the ADCF plot reports the critical value for the test under the null hypothesis of no serial dependence, namely,

$$\boldsymbol{X}_t \perp \boldsymbol{X}_{t+h} \quad \forall\, h \geq 1. \tag{5.7}$$

The critical value is obtained via the wild bootstrap as described in Fokianos and Pitsillou (2018). Similar to the estimation of distance covariance, the computational cost is driven mainly by the sample size, while the bootstrap procedure increases the runtime further by repeating the same calculation.

S&P-500 We refer back to the S&P-500 example above. Figure 5.1 shows that the ACF of the monthly returns does not exhibit meaningful autocorrelation: the sample autocorrelations are insignificant or only marginally significant at most lags. This suggests that the ACF does not reveal substantial *linear* serial dependence in the series. By contrast, the ADCF detects clear serial dependence: as shown in Figure 5.2, the ADCF is significant (based on the wild bootstrap test) for more than the first 10 lags, indicating the presence of nonlinear serial dependence that is not captured by the ACF.

We fitted a generalised ARCH (GARCH(1,1)) model, which was used as a benchmark model in Awartani and Corradi (2005), to the monthly log returns of the S&P-500 index. The model is specified as

$$\begin{aligned} r_t &= \mu + \epsilon_t, \\ \sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \end{aligned} \tag{5.8}$$

Here, r_t denotes the monthly log return at time t, μ is the mean return under this specification, and ϵ_t is the shock in returns at time t. The conditional variance of this shock is denoted by $\sigma_t^2 = \text{Var}(\epsilon_t \mid \mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1}$ is the information available up to time $t-1$. Under the GARCH$(1,1)$ specification, this conditional variance evolves over time as a function of a constant term $\omega > 0$, the previous squared shock ϵ_{t-1}^2, and the previous conditional variance σ_{t-1}^2, with coefficients α and β. Intuitively, the GARCH model is designed to capture the tendency for periods of high volatility to be followed by further periods of high volatility. The details of the fitted model is provided in Appendix H.3.

In Figure 5.3, we plot the ADCF of the (squared) standardised residuals. As shown in the figure, the ADCFs are insignificant or only marginally significant at most nonzero lags, suggesting that the GARCH(1,1) model is able to capture most of the serial dependence in the shock term, as well as most of the dependence in the conditional variance.

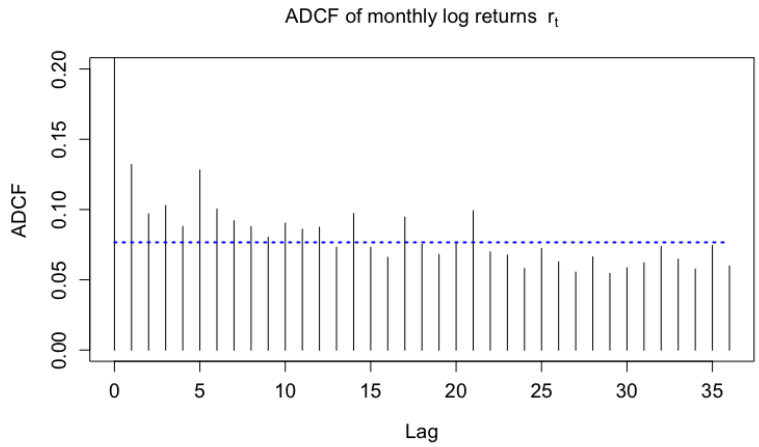


Figure 5.2: ADCF of the monthly returns of the S&P-500 from 1926 onward. The dashed line indicates the wild bootstrap critical value, so ADCF values above this line are deemed significant. The ADCF continues to indicate significant serial dependence at lags beyond 10, which is not detected by the ordinary ACF shown in Figure 5.1.

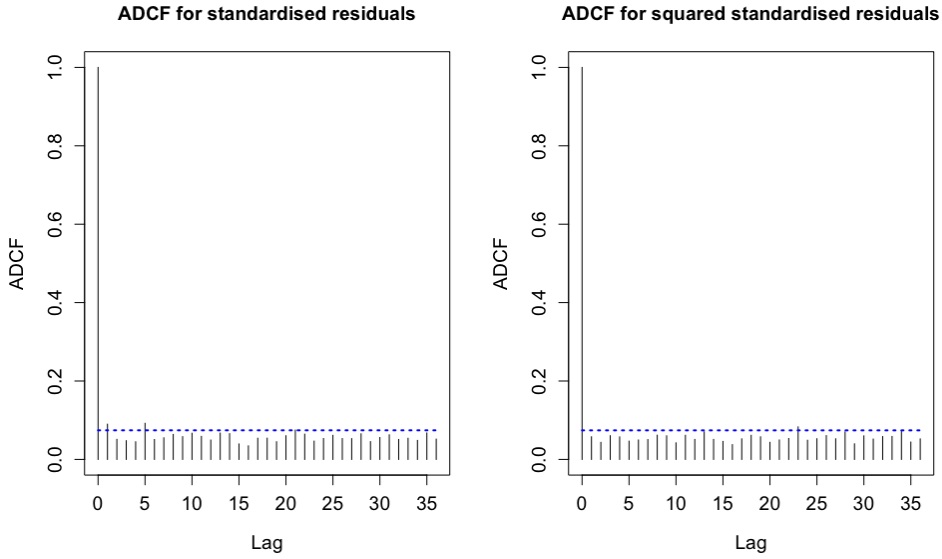


Figure 5.3: ADCF for the standardised residuals (left) and squared standardised residuals (right) of the fitted GARCH(1,1) model for the monthly log returns of the S&P-500 index. Both ADCFs are insignificant or only marginally significant at most nonzero lags. The ADCF for the standardised residuals suggests that the GARCH model captures most of the serial dependence in the shock term, while the ADCF for the squared standardised residuals suggests that the model captures most of the dependence in the conditional variance.

Los Angeles County mortality data Here we consider another example based on the Los Angeles County mortality data. Shumway et al. (1988) collected weekly data on cardiovascular mortality (`cmort`), temperature (`tempr`), and pollutant particulates (`part`) for Los Angeles County over the period 1970–1979. The left column of Figure 5.4 displays the three original time series, each of which exhibits clear annual seasonality. To remove this seasonal component, we apply a seasonal difference at lag 52, that is, $y_t - y_{t-52}$. The right column of Figure 5.4 displays the corresponding seasonally differenced series, for which the seasonal pattern is largely removed. We therefore fit the VAR model to the seasonally differenced data in the analysis below.

Augmented Dickey–Fuller tests reject the null hypothesis of a unit root for all three seasonally differenced series, supporting the use of these differenced series in the analysis below.

Figure 5.5 displays the component-wise ADCF across the three series. The diagonal panels in Figure 5.5 correspond to the ADCF of each individual series, whereas the off-diagonal panels describe the lagged cross-dependence between different variables. At nonzero lags, several panels still show significant dependence, with the stronger signals concentrated mainly in the first row. This suggests that, after seasonal differencing, some serial dependence remains, particularly for `cmort`, while lagged cross-dependence among the other pairs is relatively weak. Nevertheless, the series are not fully serially independent, which supports fitting a low-order dynamic multivariate model.

In this example, we therefore consider fitting a vector autoregression (VAR) model to the three series. Specifically, we fit the model

$$\Delta_{52}\boldsymbol{X}_t = \boldsymbol{c} + \boldsymbol{A}_1\Delta_{52}\boldsymbol{X}_{t-1} + \boldsymbol{A}_2\Delta_{52}\boldsymbol{X}_{t-2} + \boldsymbol{\varepsilon}_t, \tag{5.9}$$

where

$$\Delta_{52}\boldsymbol{X}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-52}, \qquad \boldsymbol{X}_t = (\texttt{cmort}_t, \texttt{tempr}_t, \texttt{part}_t)^\top.$$

Here, $\boldsymbol{c}$ is a vector of constants, $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are 3×3 coefficient matrices, and $\boldsymbol{\varepsilon}_t$ is the residual vector. Details of the fitted model are given in Appendix H.4. After fitting the model, we examine the ADCF (and multivariate ADCF) of the residuals, as shown in Figure 5.6. Here, the spikes largely fall below the bootstrap critical value, suggesting little remaining serial or cross-dependence and supporting an adequate model fit.

Overall, these illustrations highlight the ADCF (and its multivariate extension) as a practical complement to classical ACF/CCF tools, particularly for detecting nonlinear serial and cross-dependence. More broadly, a useful workflow is first to apply the ADCF to the raw series to assess dependence across lags, then fit an appropriate dynamic model when dependence is present, and finally re-examine the (multivariate) ADCF of the residuals as a diagnostic check for remaining serial or cross-dependence.

5.4. Limitations

The limitations are aligned with those of distance-based dependence measures more generally. Unlike the ACF or CCF, which summarise *linear* association and can sometimes be interpreted in terms of dependence structure under specific model classes, the (multivariate) ADCF provides a scalar summary of dependence and does not directly reveal the form of the underlying dependence. Moreover, while distance covariance itself is not scale-free, the ADCF is normalised (and hence scale-free) but it is not margin-free, meaning it is not invariant under strictly monotonic marginal transformations. As a result, the (multivariate) ADCF is best interpreted primarily as a tool for detecting and testing serial (and cross-)dependence, and for diagnosing remaining dependence in fitted model residuals, rather than as a fully interpretable measure of dependence structure.

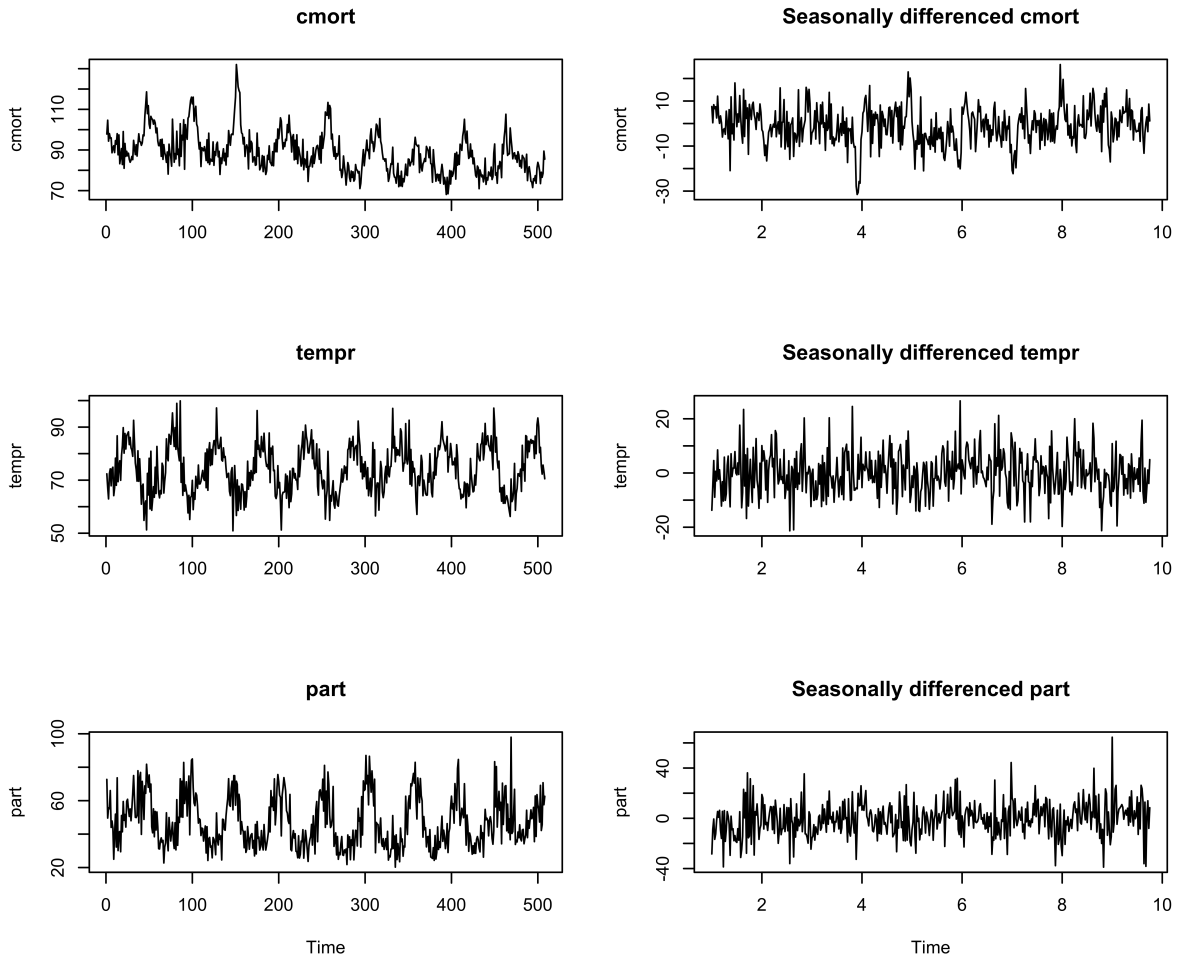


Figure 5.4: The left column displays the weekly time series of mortality (`cmort`; top), temperature (`tempr`; middle), and pollutant particulates (`part`; bottom) for Los Angeles County over the period 1970–1979, as collected by Shumway et al. (1988). The right column displays the corresponding seasonally differenced series obtained using a lag-52 difference.

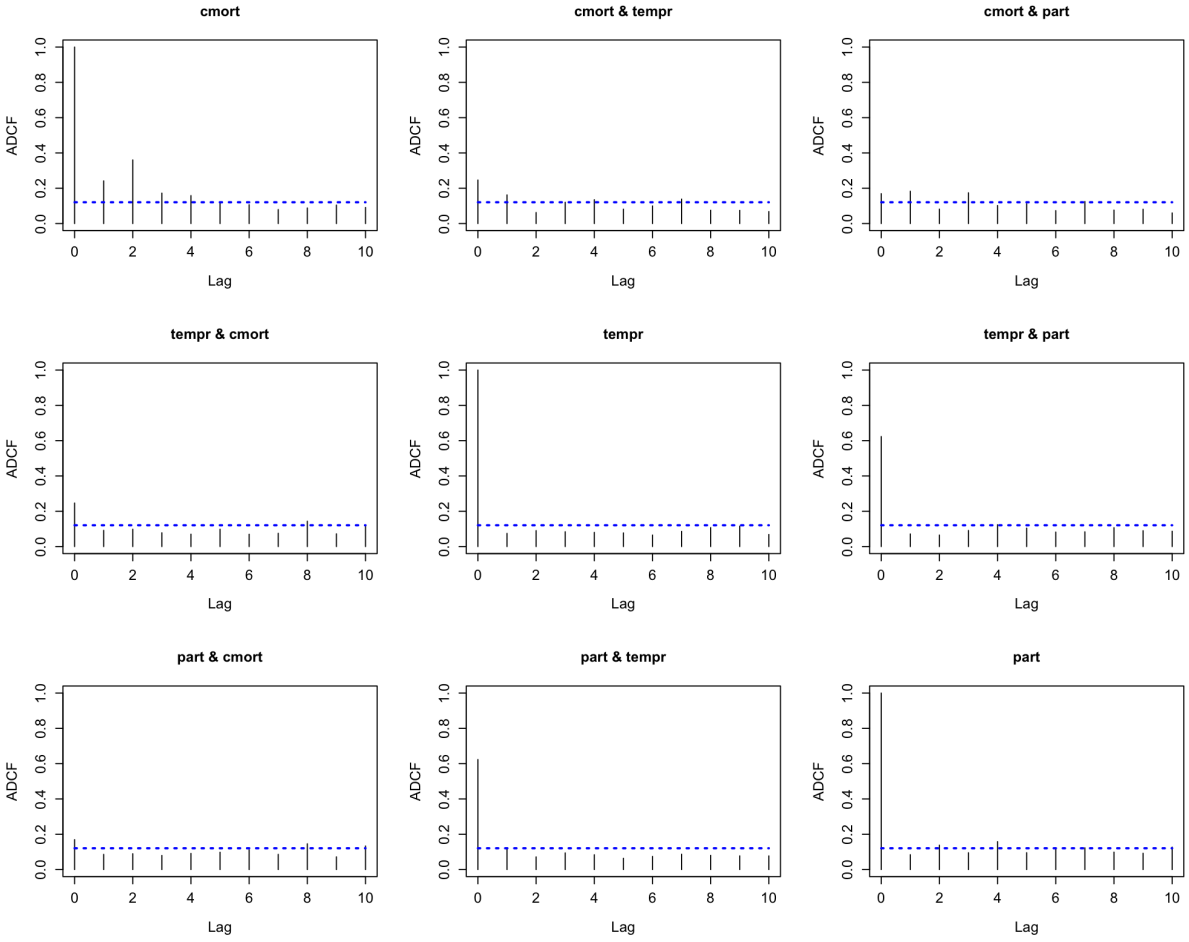


Figure 5.5: Component-wise ADCF for the seasonally differenced series (`cmort`, `tempr`, `part`) up to lag 10. The dashed line indicates the wild bootstrap critical value. At nonzero lags, several panels show significant dependence, with most of the stronger signals appearing in the first row. This suggests that, after seasonal differencing, some serial dependence remains, particularly for `cmort`, while lagged cross-dependence among the other pairs is relatively weak.

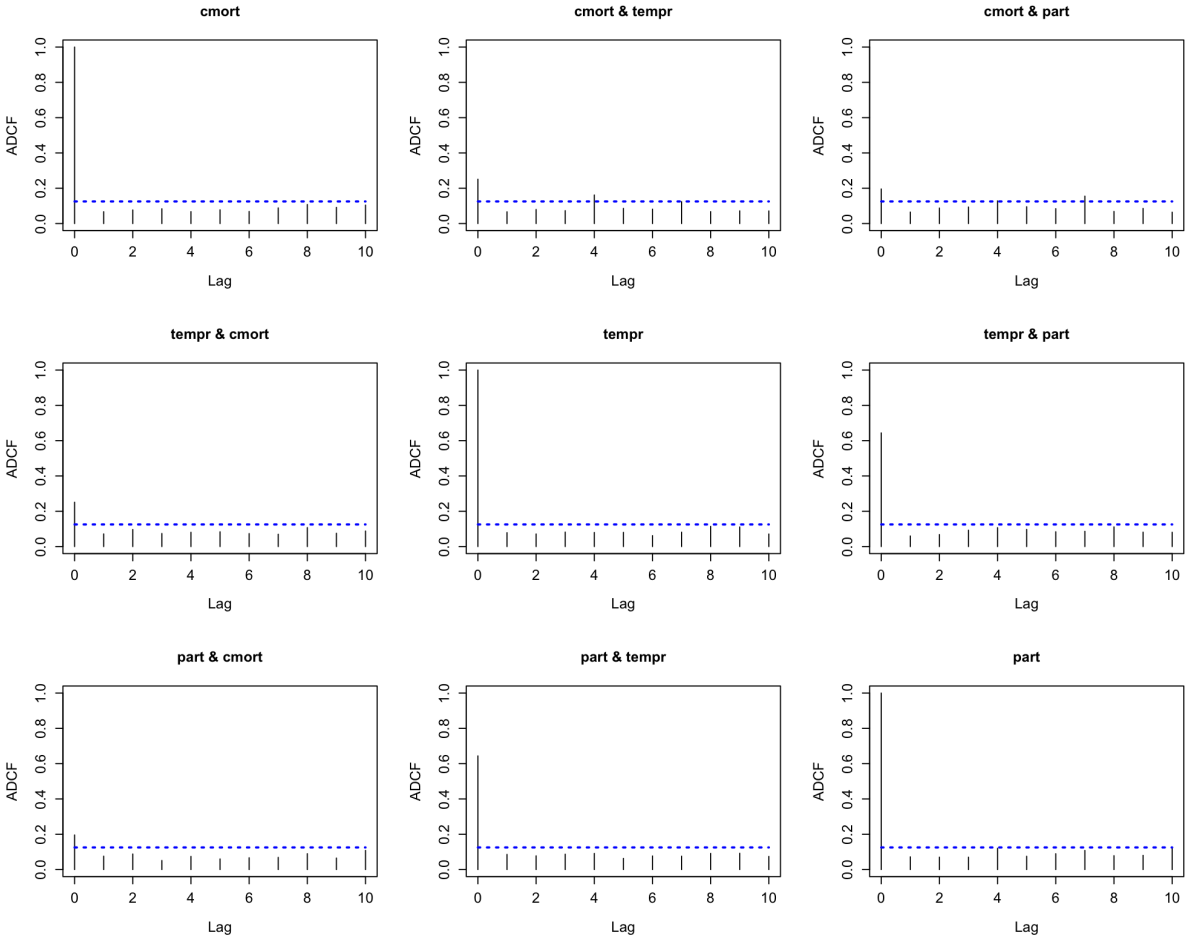


Figure 5.6: Component-wise ADCF for the residuals from the fitted VAR model in Equation (5.9) up to lag 10. The dashed line indicates the wild bootstrap critical value. Most spikes at nonzero lags lie below the critical value, indicating little remaining serial or lagged cross-dependence in the residuals. This suggests that the fitted VAR model captures most of the dynamic dependence in the data.

6. Conclusion

Dependence modelling is central to actuarial work. Pearson's correlation coefficient is often used as the default summary of association, despite extensive shortcomings, and despite the development of strong and viable alternatives in recent years. As illustrated in this paper, correlation may fail to detect nonlinear and non-monotonic dependence, cannot directly address dependence between random vectors, and is inherently pairwise, so it may miss higher-order structures and mutual dependence. In actuarial settings, these limitations are not merely technical: they can affect the validation of modelling assumptions, the interpretation of risk relationships, and even the assessment of aggregate tail risk.

To address these issues, this paper reviews and illustrates several distance-based dependence statistics that assess departure from independence by comparing suitable statistical objects (e.g., copulas) under the observed joint distribution with their counterparts under independence.

Another key message of this paper is that there is no one-size-fits-all tool that is best for all actuarial tasks. When the aim is to quantify and compare dependence strength between two continuous random variables, the Hellinger correlation is particularly attractive because it is bounded, interpretable, and margin-free. When the main goal is instead to detect or test for general dependence, especially in settings involving random vectors (or categorical data), distance covariance provides a more flexible (and potentially more powerful) tool. For problems involving more than two random variables or vectors, joint distance covariance extends this perspective to mutual dependence. Finally, extensions such as the auto-distance correlation function show that the same ideas can also be useful in broader actuarial workflows, for example in time-series model specification and diagnostics.

Overall, we advocate that classical correlation could systematically be advantageously complemented by other dependence tools. Distance-based dependence statistics provide actuaries with an intuitive and practical framework for detecting broader forms of dependence, and for checking independence assumptions more comprehensively.

Acknowledgments

The authors would like to thank Hugh Miller for helpful comments on an earlier version of this paper, which led to significant improvements.

Avanzi and Wong are grateful for financial support provided under Australian Research Council's Linkage (LP130100723, with funding partners Allianz Australia Insurance Ltd, Insurance Australia Group Ltd, and Suncorp Metway Ltd) and Discovery (DP200101859) Projects funding schemes. The views expressed herein are those of the authors and are not necessarily those of the supporting organisations.

Data and code

The data and code for illustrations are provided in the GitHub repository https://github.com/agi-lab/beyond-correlation.

References

Androschuck, T., Gibbs, S., Katrakis, N., Lau, J., Oram, S., Raddall, P., Semchyshyn, L., Stevenson, D., & Waters, J. (2017). Simulation-based capital models: Testing, justifying and communicating choices. a report from the Life Aggregation and Simulation Techniques

Working Party. *British Actuarial Journal*, *22*(2), 257–335. https://doi.org/10.1017/S1357321717000071

Atherino, R., Pizzinga, A., & Fernandes, C. (2010). A Row-Wise Stacking of the Runoff Triangle: State Space Alternatives for IBNR Reserve Prediction. *ASTIN Bulletin: The Journal of the IAA*, *40*(2), 917–946. https://doi.org/10.2143/AST.40.2.2061141

Australian Prudential Regulation Authority. (2023a). Prudential standard GPS 110 capital adequacy [Insurance (prudential standard) determination No. 2 of 2023]. https://www.legislation.gov.au/Details/F2023L00682

Australian Prudential Regulation Authority. (2023b). Prudential standard GPS 114 capital adequacy: Asset risk charge [Insurance (prudential standard) determination No. 4 of 2023]. https://www.legislation.gov.au/Details/F2023L00690

Australian Prudential Regulation Authority. (2023c). Prudential standard GPS 116 capital adequacy: Insurance concentration risk charge [Insurance (prudential standard) determination No. 6 of 2023]. https://www.legislation.gov.au/Details/F2023L00672

Avanzi, B., Boglioni Beaulieu, G., Lafaye de Micheaux, P., Ouimet, F., & Wong, B. (2021). A counterexample to the existence of a general central limit theorem for pairwise independent identically distributed random variables. *Journal of Mathematical Analysis and Applications*, *499*(1), 124982. https://doi.org/10.1016/j.jmaa.2021.124982

Avanzi, B., Taylor, G., & Wong, B. (2016). Correlations between insurance lines of business: An illusion or a real phenomenon? some methodological considerations. *ASTIN Bulletin: The Journal of the IAA*, *46*(2), 225–263. https://doi.org/10.1017/asb.2015.31

Awartani, B. M., & Corradi, V. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: The role of asymmetries. *International Journal of Forecasting*, *21*(1), 167–183. https://doi.org/10.1016/j.ijforecast.2004.08.003

Awiszus, K., Knispel, T., Penner, I., Svindland, G., Voß, A., & Weber, S. (2023). Modeling and pricing cyber insurance: Idiosyncratic, systematic, and systemic risks. *European Actuarial Journal*, *13*(1), 1–53. https://doi.org/10.1007/s13385-023-00341-9

Bernard, C., & Vanduffel, S. (2015). Risk aggregation and diversification [Canadian Institute of Actuaries research paper, dated October 29, 2015].

Bilodeau, M., & Lafaye de Micheaux, P. (2005). A multivariate empirical characteristic function test of independence with normal marginals. *Journal of Multivariate Analysis*, *95*(2), 345–369. https://doi.org/10.1016/j.jmva.2004.08.011

Boglioni Beaulieu, G. (2023). *Pairwise versus mutual independence: Visualisation, actuarial applications and central limit theorems* [Doctoral dissertation, UNSW Sydney].

Boglioni Beaulieu, G., Lafaye de Micheaux, P., & Ouimet, F. (2021). Counterexamples to the classical central limit theorem for triplewise independent random variables having a common arbitrary margin. *Dependence Modeling*, *9*, 424–438. https://doi.org/10.1515/demo-2021-0120

Bohlmann, G. (1901). Lebensversicherungsmathematik [Artikel I D 4b]. In *Encyklopädie der mathematischen wissenschaften mit einschluss ihrer anwendungen* (pp. 852–917, Vol. 1). B. G. Teubner.

Bohlmann, G. (1909). Die grundbegriffe der wahrscheinlichkeitsrechnung in ihrer anwendung auf die lebensversicherung. *Atti del IV Congresso internazionale dei Matematici (Roma, 6–11 Aprile, 1908)*, *3*, 244–278.

Böttcher, B. (2020). Copula versions of distance multivariance and dhsic via the distributional transform–a general approach to construct invariant dependence measures. *Statistics*, *54*(3), 577–594. https://doi.org/10.1080/02331888.2020.1748029

Böttcher, B., Keller-Ressel, M., & Schilling, R. L. (2019). Distance multivariance: New dependence measures for random vectors. *The Annals of Statistics*, *47*(5), 2757–2789. https://doi.org/10.1214/18-AOS1764

Bradley, C. (1985). The absolute correlation coefficient. *The Mathematical Gazette*, *69*(447), 12–17. https://doi.org/10.2307/3616441

Cai, P., Abdallah, A., & Jeganathan, P. (2025). Recurrent neural networks for multivariate loss reserving and risk capital analysis. *North American Actuarial Journal*, 1–25. https://doi.org/10.1080/10920277.2025.2517149

Central Intelligence Agency. (2020). The world factbook [Source used for the worlddemographics data in the HellCor package].

Chakraborty, S., Yuan, D., & Zhang, X. (2019). *jdcov: Tests for joint independence via joint distance covariance* [GitHub package documentation for shubhadeep4/jdcov]. https://rdrr.io/github/shubhadeep4/jdcov/man/jdcov-package.html

Chakraborty, S., & Zhang, X. (2019). Distance metrics for measuring joint dependence with application to causal inference. *Journal of the American Statistical Association*, *114*(528), 1638–1650. https://doi.org/10.1080/01621459.2018.1513364

Chukhrova, N., & Johannssen, A. (2021). Stochastic claims reserving methods with state space representations: A review. *Risks*, *9*(11), 198. https://doi.org/10.3390/risks9110198

CRO Forum. (2015). Casualty accumulation risk [October 2015]. https://www.thecroforum.org/wp-content/uploads/2015/10/CROF-Casualty-Accumulation-Risk-FINALv11.pdf

Deheuvels, P. (1983). Indépendance multivariée partielle et inégalités de fréchet. In M. C. Demetrescu & M. Iosifescu (Eds.), *Studies in probability and related topics: Papers in honour of Octav Onicescu on his 90th birthday* (pp. 145–155). Editrice Nagard.

Denuit, M., & Dhaene, J. (2003). Simple characterizations of comonotonicity and countermonotonicity by extremal correlations. *Belgian Actuarial Bulletin*, *3*, 22–27.

Denuit, M., Dhaene, J., Goovaerts, M., & Kaas, R. (2006). *Actuarial theory for dependent risks: Measures, orders and models*. John Wiley & Sons. https://doi.org/10.1002/0470016450

Dutang, C., & Charpentier, A. (2026). *CASdatasets: Insurance datasets* [R package version 1.2-1]. https://doi.org/10.57745/P0KHAG

Edelmann, D., Móri, T. F., & Székely, G. J. (2021). On relationships between the pearson and the distance correlation coefficients. *Statistics & Probability Letters*, *169*, 108960. https://doi.org/10.1016/j.spl.2020.108960

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer. https://doi.org/10.1007/978-3-642-33483-2

Embrechts, P., McNeil, A. J., & Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In M. A. H. Dempster (Ed.), *Risk management: Value at risk and beyond* (pp. 176–223). Cambridge University Press. https://doi.org/10.1017/CBO9780511615337.008

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*(4), 987–1007. https://doi.org/10.2307/1912773

European Parliament and the Council of the European Union. (2009). Directive 2009/138/ec of the european parliament and of the council of 25 november 2009 on the taking-up and pursuit of the business of insurance and reinsurance (Solvency II). https://eur-lex.europa.eu/eli/dir/2009/138/oj/eng

Fan, Y., Lafaye de Micheaux, P., Penev, S., & Salopek, D. (2017). Multivariate nonparametric test of independence. *Journal of Multivariate Analysis*, *153*, 189–210. https://doi.org/10.1016/j.jmva.2016.09.014

Feuerverger, A. (1993). A consistent test for bivariate dependence. *International Statistical Review*, *61*(3), 419–433. https://doi.org/10.2307/1403753

Fokianos, K., & Pitsillou, M. (2018). Testing independence for multivariate time series via the auto-distance correlation matrix. *Biometrika*, *105*(2), 337–352. https://doi.org/10.1093/biomet/asx082

Geenens, G., & Lafaye de Micheaux, P. (2020). *HellCor: The hellinger correlation* [R package version 1.3]. https://doi.org/10.32614/CRAN.package.HellCor

Geenens, G., & Lafaye de Micheaux, P. (2022). The Hellinger correlation. *Journal of the American Statistical Association*, *117*(538), 639–653. https://doi.org/10.1080/01621459.2020.1791132

Goldburd, M., Khare, A., Tevet, D., & Guller, D. (2025). *Generalized linear models for insurance rating* (2nd ed.) [2025 revision]. Casualty Actuarial Society. https://www.casact.org/monograph/cas-monograph-no-5

Grari, V., Charpentier, A., & Detyniecki, M. (2022). A fair pricing model via adversarial learning. https://arxiv.org/abs/2202.12008

Hang, W. (2020). *dcov: A fast implementation of distance covariance* [R package version 0.1.1]. https://doi.org/10.32614/CRAN.package.dcov

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems 29*.

International Association of Insurance Supervisors. (2024). Insurance capital standard: Level 1 and level 2 texts [December 2024]. https://www.iais.org/uploads/2024/12/ICS-Level-1-and-Level-2-texts.pdf

Jiang, N. (2020). IFRS 17: Risk adjustment—a numerical example [Society of Actuaries newsletter article]. *The Financial Reporter*, (May), 9–13.

Kafková, S., & Křivánková, L. (2014). Generalized linear models in vehicle insurance. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *62*(2), 383–388. https://doi.org/10.11118/actaun201462020383

Kathari, S., & Tangirala, A. K. (2020). Scalar correlation functions for model structure selection in high-dimensional time-series modelling. *ISA Transactions*, *100*, 275–288. https://doi.org/10.1016/j.isatra.2019.11.033

Kearns, M., Neel, S., Roth, A., & Wu, Z. S. (2018). Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. *Proceedings of the 35th International Conference on Machine Learning*, *80*, 2564–2572.

Kelley, J. L. (2017). *General topology*. Courier Dover Publications.

Kelliher, P. O. J., Acharyya, M., Couper, A., Maguire, E., Nicholas, P., Pang, N., Smerald, C., Stevenson, D., Sullivan, J., & Teggin, P. (2020). Operational risk dependencies. *British Actuarial Journal*, *25*, e5. https://doi.org/10.1017/S1357321720000033

Kelliher, P. (2022). Dependencies and diversification between risks. patrick kelliher fia cera december 2021. *British Actuarial Journal*, *27*, e8. https://doi.org/10.1017/S1357321722000095

Kolmogorov, A. N. (2018). *Foundations of the theory of probability* (Second English Edition). Courier Dover Publications.

Lai, D., & Lu, B. (2017). Autoregressive model for time series as a deterministic dynamic system [June 2017]. *Predictive Analytics and Futurism*, (15), 7–9.

Lakens, D., & Caldwell, A. (2025). *TOSTER: Two one-sided tests (tost) equivalence testing* [R package version 0.8.6]. https://doi.org/10.32614/CRAN.package.TOSTER

Lee, H. M., Antonio, K., Avanzi, B., Marchi, L., & Zhou, R. (2025). Machine learning with multi-type protected attributes: Intersectional fairness through regularisation. https://arxiv.org/abs/2509.08163

Lee, R. (2000). The lee-carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, *4*(1), 80–91. https://doi.org/10.1080/10920277.2000.10595882

Leyder, S., Raymaekers, J., & Rousseeuw, P. J. (2026). Robust distance covariance. *International Statistical Review*, *94*(1). https://doi.org/10.1111/insr.70005

Li, H., Zhou, R., & Ji, M. (2023). Nonlinear modeling of mortality data and its implications for longevity bond pricing. *Risks*, *11*(12), 207. https://doi.org/10.3390/risks11120207

Macdonald, A., Gallop, A., Miller, K., Richards, S., Shah, R., & Willets, R. (2007). Cmi working paper 25 (revised): Stochastic projection methodologies: Lee-carter model features, example results and implications [Revised version issued November 2007].

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin: The Journal of the IAA*, *23*(2), 213–225. https://doi.org/10.2143/AST.23.2.2005092

McLeod, A. I., & Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, *4*(4), 269–273. https://doi.org/10.1111/j.1467-9892.1983.tb00373.x

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Molnar, C. (2020). *Interpretable machine learning*. https://christophm.github.io/interpretable-ml-book/

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. John Wiley Sons. https://doi.org/10.1002/9780470316559

National Association of Insurance Commissioners. (2025). NAIC model review manual [Adopted by Casualty Actuarial and Statistical (C) Task Force on November 4, 2025]. https://content.naic.org/sites/default/files/inline-files/NAIC%20Model%20Review%20Manual__%20adopted%20by%20CASTF%2011.04.25.pdf

Ng, W. Y., Wang, L. R., Liu, S., & Fan, X. (2025). Causal SHAP: Feature attribution with dependency awareness through causal discovery. *2025 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Peng, L., & Wang, R. (2014). Interval estimation for bivariate t-copulas via kendall's tau. *Variance*, *8*(1), 43–54.

Peters, G. W., Yan, H., & Chan, J. (2021). Statistical features of persistence and long memory in mortality data. *Annals of Actuarial Science*, *15*(2), 291–317. https://doi.org/10.1017/S1748499521000129

Puccetti, G., & Scarsini, M. (2010). Multivariate comonotonicity. *Journal of Multivariate Analysis*, *101*(1), 291–304. https://doi.org/10.1016/j.jmva.2009.08.003

Rényi, A. (1959). On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungaricae*, *10*(3–4), 441–451. https://doi.org/10.1007/BF02024507

Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: Simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia Medica*, *31*(1), 27–53. https://doi.org/10.11613/BM.2021.010502

Shaw, R. A., Smith, A. D., & Spivak, G. S. (2011). Measurement and modelling of dependencies in economic capital. *British Actuarial Journal*, *16*(3), 601–699. https://doi.org/10.1017/S1357321711000249

Shaw, R., & Spivak, G. (2009). Correlations and dependencies in economic capital models. *2009 Risk and Investment Conference, Institute and Faculty of Actuaries*.

Shi, P. (2014). A copula regression for modeling multivariate loss triangles and quantifying reserving variability. *ASTIN Bulletin: The Journal of the IAA*, *44*(1), 85–102. https://doi.org/10.1017/asb.2013.23

Shi, P., & Shi, K. (2023). Non-life insurance risk classification using categorical embedding. *North American Actuarial Journal*, *27*(3), 579–601. https://doi.org/10.1080/10920277.2022.2123361

Shumway, R. H., Azari, A. S., & Pawitan, Y. (1988). Modeling mortality fluctuations in los angeles as functions of pollution and weather effects. *Environmental Research*, *45*(2), 224–241. https://doi.org/10.1016/S0013-9351(88)80049-5

Sun, H., Xu, M., & Zhao, P. (2021). Modeling malicious hacking data breach risks. *North American Actuarial Journal*, *25*(4), 484–502. https://doi.org/10.1080/10920277.2020.1752255

Székely, G. J., & Rizzo, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics*, *3*(4), 1236–1265. https://doi.org/10.1214/09-AOAS312

Székely, G. J., & Rizzo, M. L. (2012). On the uniqueness of distance covariance. *Statistics & Probability Letters*, *82*(12), 2278–2282. https://doi.org/10.1016/j.spl.2012.08.007

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, *35*(6), 2769–2794. https://doi.org/10.1214/009053607000000505

Tsagris, M., Pitsillou, M., & Fokianos, K. (2026). *dCovTS: Distance covariance and correlation for time series analysis* [R package version 1.5]. https://doi.org/10.32614/CRAN.package.dCovTS

Usman, F., & Chan, J. S. K. (2022). New loss reserve models with persistence effects to forecast trapezoidal losses in run-off triangles. *ASTIN Bulletin: The Journal of the IAA*, *52*(3), 877–920. https://doi.org/10.1017/asb.2022.17

van den Heuvel, E., & Zhan, Z. (2022). Myths about linear and monotonic associations: Pearson's r, spearman's ρ, and kendall's τ. *The American Statistician*, *76*(1), 44–52. https://doi.org/10.1080/00031305.2021.2004922

Venter, G. G. (2002). Tails of copulas. *Proceedings of the Casualty Actuarial Society*, *89*(171), 68–113.

Xin, X., & Huang, F. (2024). Antidiscrimination insurance pricing: Regulations, fairness criteria, and models. *North American Actuarial Journal*, *28*(2), 285–319. https://doi.org/10.1080/10920277.2023.2190528

Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis*, *33*(3), 438–457. https://doi.org/10.1111/j.1467-9892.2011.00780.x

A. Distance in mathematics

It is important to note that distance-based dependence statistics are not necessarily a *distance* in the mathematical sense. In mathematics, a distance on a set S is a function $d : S \times S \to [0, \infty)$ such that for all $x, y, z \in S$ it satisfies the following axioms (Kelley, 2017):

$$d(x, y) \geq 0 \qquad \text{(Non-negativity)} \tag{A.1}$$
$$d(x, y) = 0 \iff x = y \qquad \text{(Identity of indiscernibles)} \tag{A.2}$$
$$d(x, y) = d(y, x) \qquad \text{(Symmetry)} \tag{A.3}$$
$$d(x, z) \leq d(x, y) + d(y, z) \qquad \text{(Triangle inequality).} \tag{A.4}$$

However, distance-based dependence statistics do not generally satisfy these properties when interpreted as distance metrics on random variables or random vectors. In particular, the identity of indiscernibles may fail. For example, for distance covariance, if $\boldsymbol{X} \in \mathcal{X}$, $\text{dCov}^2(\boldsymbol{X}, \boldsymbol{X}) = 0$ only if $\boldsymbol{X} = \mathbb{E}[\boldsymbol{X}]$ almost surely (Theorem 3(iv), Székely and Rizzo, 2009); otherwise, we have $\text{dCov}(\boldsymbol{X}, \boldsymbol{X}) > 0$. Thus, distance covariance does not satisfy the identity of indiscernibles axiom.

B. Details of simulated examples

B.1. Computation of the VaR and TVaR for the one-year loss deviation example

To illustrate how zero pairwise correlation can mask material dependence in capital aggregation, consider a one-year risk model with three business lines. For each business line $i = 1, 2, 3$, let X_i denote the deviation of the realised one-year loss from its expected value. Hence, $X_i > 0$ represents an adverse outcome, meaning that the realised loss is higher than expected, while $X_i < 0$ represents a favourable outcome, meaning that the realised loss is lower than expected.

We define S_i as the direction of the deviation for business line i, where they are independent and identically distributed. Here, $S_i = 1$ corresponds to an adverse deviation and $S_i = -1$ to a favourable deviation, with

$$\mathbb{P}(S_i = 1) = \mathbb{P}(S_i = -1) = \tfrac{1}{2}, \quad i = 1, 2, 3.$$

We then define a common shock multiplier R, independent of (S_1, S_2, S_3), by

$$R = \begin{cases} 1, & \text{with probability } 1 - 0.045 \text{ (normal year)}, \\ 20, & \text{with probability } 0.045 \text{ (volatile year)}. \end{cases} \tag{B.1}$$

The variable R represents the severity of the overall operating environment in that year. In a normal year, deviations from expectation are small, whereas in a volatile year the same favourable or adverse direction is magnified across all business lines. The one-year loss deviation for business line i is then defined by

$$X_i = RS_i, \qquad i = 1, 2, 3.$$

The aggregate one-year loss deviation is

$$L = X_1 + X_2 + X_3.$$

The probability mass function of L is given by

$$\begin{aligned}
\mathbb{P}(L = 1) = \mathbb{P}(L = -1) &= (1 - 0.045) \cdot \frac{3}{8} = 0.358125, \\
\mathbb{P}(L = 20) = \mathbb{P}(L = -20) &= 0.045 \cdot \frac{3}{8} = 0.016875, \\
\mathbb{P}(L = 3) = \mathbb{P}(L = -3) &= (1 - 0.045) \cdot \frac{1}{8} = 0.119375, \\
\mathbb{P}(L = 60) = \mathbb{P}(L = -60) &= 0.045 \cdot \frac{1}{8} = 0.005625.
\end{aligned} \tag{B.2}$$

Hence, the value-at-risk (VaR) at level $\alpha = 0.995$ is

$$\mathsf{VaR}_{0.995}(L) = \mathsf{inf}\{\ell \mid F_L(\ell) \geq 0.995\} = 60. \tag{B.3}$$

In addition, the tail value-at-risk (TVaR) is

$$\mathsf{TVaR}_{0.995}(L) = \mathbb{E}\left[L \mid L \geq \mathsf{VaR}_{0.995}(L)\right] = 60. \tag{B.4}$$

Here, the one-year loss deviations X_1, X_2, and X_3 are not independent, since they share the common shock multiplier R. However, the pairwise correlations between the one-year loss deviations are all zero:

$$\mathsf{Corr}(X_i, X_j) = \frac{\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]}{\sqrt{\mathsf{Var}(X_i)\,\mathsf{Var}(X_j)}} = \frac{\mathbb{E}[R^2]\mathbb{E}[S_i S_j]}{\mathbb{E}[R^2]} = 0, \qquad \forall\, i \neq j,$$

If one incorrectly assumes independence on this basis, while keeping the same marginals for X_1, X_2, and X_3, then for $i = 1, 2, 3$ we have

$$\begin{aligned}
\mathbb{P}(X_i = \pm 1) &= 0.4775, \\
\mathbb{P}(X_i = \pm 20) &= 0.0225.
\end{aligned} \tag{B.5}$$

Let L' denote the aggregate one-year loss deviation under this incorrect assumption of independence between the three business lines. Then Table B.1 displays the probability mass function of L'.

ℓ	$\mathbb{P}(L' = \ell)$
± 60	0.00001139
± 41	0.00072520
± 39	0.00072520
± 22	0.01539042
± 20	0.03081502
± 18	0.01539042
± 3	0.10887298
± 1	0.32806936

Table B.1: Probability mass function of $L' = X_1 + X_2 + X_3$ under the incorrect assumption that X_1, X_2, and X_3 are independent.

Given this, we have $\mathsf{VaR}_{0.995}(L') = 22$ and

$$\mathsf{TVaR}_{0.995}(L') = \frac{22\mathbb{P}(L' = 22) + 39\mathbb{P}(L' = 39) + 41\mathbb{P}(L' = 41) + 60\mathbb{P}(L' = 60)}{\mathbb{P}(L' \geq 22)} \approx 23.5749. \tag{B.6}$$

This is substantially lower than the correct value $\text{TVaR}_{0.995}(L) = 60$. The example therefore illustrates that treating zero correlation as independence can materially understate aggregate tail risk.

B.2. Data generation for the Multi-line car insurance and medical insurance example

Consider a road accident that gives rise to losses in two lines of business.

- **Motor line** $\boldsymbol{X} = (X_1, X_2, X_3)$, which consists of three components: vehicle repair cost, bodily injury liability, and claims handling cost, denoted by X_1, X_2, and X_3, respectively. For these random variables, we introduce the following latent variables:
  - U_1: latent index of physical collision intensity;
  - U_2: latent index of injury intensity.

  We also introduce an indicator $R \sim \text{Bernoulli}(0.5)$, which indicates whether the claim is severity-driven or review-driven. If the claim is severity-driven, the claims handling cost X_3 should increase as the medical cost increases. In contrast, if it is review-driven, then a smaller medical cost may still lead to a larger claims handling cost, for example because the claim is difficult to verify.
- **Medical line** Y, which is the medical claim cost. It depends on a common latent variable reflecting the combined severity of the accident and the injury.

Let

$$U_1, U_2 \overset{\text{i.i.d.}}{\sim} \text{Unif}(0,1). \tag{B.7}$$

Define the latent variable

$$C = (U_1 + U_2) \bmod 1, \tag{B.8}$$

which combines information on physical collision intensity and injury intensity. Next, define

$$W = \begin{cases} C, & \text{if } R = 1, \\ 1 - C, & \text{if } R = 0. \end{cases} \tag{B.9}$$

The motor-line components are generated as

$$X_1 = F_1^{-1}(U_1), \qquad F_1 \sim \Gamma(2.5, 2000), \tag{B.10}$$

$$X_2 = F_2^{-1}(U_2), \qquad F_2 \sim \Gamma(3.0, 3000), \tag{B.11}$$

$$X_3 = F_3^{-1}(W), \qquad F_3 \sim \Gamma(2.0, 1200). \tag{B.12}$$

The medical-line component is generated as

$$Y = G_1^{-1}(C), \qquad G_1 \sim \Gamma(2.2, 2500). \tag{B.13}$$

This is generated by the following code:

```r
set.seed(123)
n <- 5000

# latent variables
U1 <- runif(n)
U2 <- runif(n)
U3 <- runif(n)
```

```r
C <- (U1 + U2) %% 1

# original motor components
X1 <- qgamma(U1, shape = 2.5, scale = 2000)
X2 <- qgamma(U2, shape = 3.0, scale = 3000)
# realistic medical components
Y1 <- qgamma(C, shape = 2.2, scale = 2500)

# regime indicator:
# R = 1 severity-driven pathway
# R = 0 dispute-driven pathway
p <- 0.50
R <- rbinom(n, size = 1, prob = p)

W <- ifelse(R == 1, C, 1 - C)

# investigation/dispute cost
X3 <- qgamma(W, shape = 2.0, scale = 1200)

dat <- data.frame(X1, X2, X3, Y1, Y2)
```

C. Definition of perfect dependence

In actuarial science, the notion of perfect dependence is often associated with *comonotonicity* (Denuit et al., 2006; Embrechts et al., 2002). For two univariate random variables X and Y, comonotonicity holds if and only if there exists a uniform random variable U on $(0, 1)$ such that

$$X = F_X^{-1}(U), \qquad Y = F_Y^{-1}(U), \tag{C.1}$$

where F_X and F_Y denote the CDFs of X and Y, respectively. Intuitively, this means that X and Y always move in the same direction; in particular, if they represent risks, then an increase in one risk is accompanied by an increase in the other. Two comonotonic random variables always have the same rank under the same realised value of U, and hence Kendall's τ and Spearman's ρ are both equal to 1 (Denuit & Dhaene, 2003).

When one considers two random vectors of the same dimension, $\boldsymbol{X} = (X_1, \ldots, X_p) \in \mathbb{R}^p$ and $\boldsymbol{Y} = (Y, \ldots, Y_p) \in \mathbb{R}^p$, the concept of comonotonicity becomes more complicated, since vectors are not totally ordered; that is, not every pair of vectors can be ranked as larger or smaller in a component-wise sense. To address this issue, Puccetti and Scarsini (2010) introduced several notions of comonotonicity for random vectors. One such notion is *strong comonotonicity*, which holds if there exists a uniform random variable U on $(0, 1)$ such that

$$(\boldsymbol{X}, \boldsymbol{Y}) = \left(\left(F_{X_1}^{-1}(U), \ldots, F_{X_d}^{-1}(U) \right), \left(F_Y^{-1}(U), \ldots, F_{Y_d}^{-1}(U) \right) \right). \tag{C.2}$$

In other words, every component of the concatenated vector $(\boldsymbol{X}, \boldsymbol{Y})$ moves in the same direction.

In some actuarial discussions, perfect dependence is also described in terms of exact predictability, such that one variable can be determined exactly when the other is known (R. A. Shaw et al., 2011; R. Shaw & Spivak, 2009). This notion is noted as *symmetric predictability* in Geenens and Lafaye de Micheaux (2022). Formally, X and Y are symmetrically predictable if there exist measurable functions ψ_1 and ψ_2 such that

$$X = \psi_1(Y), \qquad Y = \psi_2(X). \tag{C.3}$$

In this paper, we adopt the weaker notion of perfect dependence defined in Geenens and Lafaye de Micheaux (2022). Under this notion, comonotonicity implies perfect dependence, but perfect dependence does not necessarily imply comonotonicity. Specifically, we say that X and Y are perfectly dependent if there exist a uniform random variable U on $(0, 1)$ and a measurable function $\psi : [0, 1] \to \mathbb{R}^2$ such that

$$(X, Y) = \psi(U). \tag{C.4}$$

Intuitively, two risks X and Y are perfectly dependent when they are fully driven by a single underlying risk factor, so that once this common factor is realised, both variables are determined exactly. Equivalently, the pair (X, Y) contains only one source of uncertainty, rather than separate sources.

D. Hypothesis tests for correlation coefficients

In this section, we consider two random variables X and Y, and let $\{(x_i, y_i)\}_{i=1}^n$ denote an observed sample from (X, Y) of size n. Below, we describe the procedures for testing whether the corresponding population correlation coefficient is equal to zero for each type of correlation considered.

D.1. Pearson's r

For testing the hypothesis that the Pearson correlation coefficient between two random variables X and Y is zero, we consider

$$\begin{aligned} H_0 &: r_{\text{pop}} = 0, \\ H_1 &: r_{\text{pop}} \neq 0, \end{aligned} \tag{D.1}$$

where r_{pop} denotes the population Pearson correlation coefficient.

We perform this test using a bootstrap procedure. Let $\{(x_i, y_i)\}_{i=1}^n$ denote the observed sample of (X, Y) of size n. A bootstrap sample is generated by resampling these observed pairs with replacement, giving

$$\{(x_{i_j^*}, y_{i_j^*})\}_{j=1}^n, \tag{D.2}$$

where $i_1^*, \ldots, i_n^*$ are sampled independently from $\{1, \ldots, n\}$ with replacement. For each bootstrap sample, we compute the Pearson correlation coefficient. Repeating this procedure R times yields bootstrap correlations $r_1^*, \ldots, r_R^*$. For the two-sided test with null value 0, the p-value is computed as

$$p = 2 \min\{\hat{p}, 1 - \hat{p}\}, \tag{D.3}$$

where

$$\hat{p} = \frac{\#(r_b^* < 0) + 0.5 \times \#(r_b^* = 0)}{R}. \tag{D.4}$$

This is implemented in the function `boot_cor_test` in the `R` package `TOSTER` (Lakens & Caldwell, 2025).

D.2. Kendall's τ

We consider the hypothesis test

$$\begin{aligned} H_0 &: \tau = 0, \\ H_1 &: \tau \neq 0, \end{aligned} \tag{D.5}$$

where τ denotes the population Kendall rank correlation coefficient.

For the asymptotic version of the test, we first compute the sample Kendall rank correlation coefficient $\hat{\tau}$. Then we define

$$T_0 = \frac{n(n-1)}{2}. \tag{D.6}$$

If there are ties in x, let $t_1, t_2, \ldots$ denote the corresponding tie block sizes, that is, the numbers of observations sharing the same rank. Similarly, if there are ties in y, let $u_1, u_2, \ldots$ denote the corresponding tie block sizes. Set

$$T_1 = \sum_i \frac{t_i(t_i - 1)}{2}, \qquad T_2 = \sum_j \frac{u_j(u_j - 1)}{2}. \tag{D.7}$$

Then

$$S = \hat{\tau} \sqrt{(T_0 - T_1)(T_0 - T_2)}. \tag{D.8}$$

The variance of S is estimated by

$$\mathsf{Var}(S) = \frac{v_0 - v_t - v_u}{18} + \frac{v_1}{2n(n-1)} + \frac{v_2}{9n(n-1)(n-2)}, \tag{D.9}$$

where

$$v_0 = n(n-1)(2n+5), \tag{D.10}$$

$$v_t = \sum_i t_i(t_i-1)(2t_i+5), \qquad v_u = \sum_j u_j(u_j-1)(2u_j+5), \tag{D.11}$$

$$v_1 = \left(\sum_i t_i(t_i-1)\right)\left(\sum_j u_j(u_j-1)\right), \tag{D.12}$$

and

$$v_2 = \left(\sum_i t_i(t_i-1)(t_i-2)\right)\left(\sum_j u_j(u_j-1)(u_j-2)\right). \tag{D.13}$$

The asymptotic test statistic is then

$$z = \frac{S}{\sqrt{\mathsf{Var}(S)}}. \tag{D.14}$$

For the two-sided test, the p-value is computed as

$$p = 2\,\mathsf{min}\{\Phi(z),\, 1-\Phi(z)\}, \tag{D.15}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. In `R`, this asymptotic test can be conducted using `cor.test` with the option `method = "kendall"`.

D.3. Spearman's ρ

We consider the hypothesis test

$$\begin{aligned} H_0 &: \rho_S = 0, \\ H_1 &: \rho_S \neq 0, \end{aligned} \tag{D.16}$$

where ρ_S denotes the population Spearman rank correlation coefficient. Let $\hat{\rho}_S$ be the sample Spearman rank correlation coefficient. Let $R_i = \mathsf{rank}(x_i)$ and $Q_i = \mathsf{rank}(y_i)$ for $i = 1, \ldots, n$. The corresponding rank-difference statistic is

$$S = \sum_{i=1}^{n} (R_i - Q_i)^2. \tag{D.17}$$

Define

$$d = \frac{n(n^2-1)}{6}. \tag{D.18}$$

Then

$$r = 1 - \frac{S}{d}, \tag{D.19}$$

and the test statistic is given by

$$T = \frac{r}{\sqrt{(1-r^2)/(n-2)}}. \tag{D.20}$$

Under the asymptotic approximation, the p-value for the two-sided test is computed from the t distribution with $n-2$ degrees of freedom as

$$p = 2 \min \left\{ F_{t_{n-2}}(T),\, 1 - F_{t_{n-2}}(T) \right\}, \tag{D.21}$$

where $F_{t_{n-2}}$ denotes the cumulative distribution function of the t distribution with $n-2$ degrees of freedom. In `R`, this can be obtained using `cor.test` with `method = "spearman"`.

E. Details for Hellinger correlation

E.1. Empirical estimator of the Hellinger correlation

For a sample $\{(x_i, y_i)\}_{i=1}^n$ from two continuous random variables (X, Y), the Hellinger correlation is estimated through the associated copula. Let

$$\widehat{U}_i = (\widehat{U}_{i1}, \widehat{U}_{i2}) = \big(\widehat{F}_{n,X}(x_i), \widehat{F}_{n,Y}(y_i)\big), \qquad i = 1, \dots, n, \tag{E.1}$$

denote the pseudo-observations, where

$$\widehat{F}_{n,X}(x) = \frac{1}{n+1} \sum_{j=1}^{n} \mathbf{1}(x_j \le x), \qquad \widehat{F}_{n,Y}(y) = \frac{1}{n+1} \sum_{j=1}^{n} \mathbf{1}(y_j \le y). \tag{E.2}$$

For each pseudo-observation $\widehat{U}_i$, define its nearest-neighbour distance by

$$\widehat{R}_i = \min_{j \neq i} \|\widehat{U}_j - \widehat{U}_i\|_2, \tag{E.3}$$

where $\|\cdot\|$ denotes the Euclidean norm, that is, for a vector $\boldsymbol{z} = (z_1, z_2)^\top \in \mathbb{R}^2$,

$$\|\boldsymbol{z}\| = \sqrt{z_1^2 + z_2^2}. \tag{E.4}$$

The basic estimator of

$$B = \int_{[0,1]^2} \sqrt{c(u, v)}\, du\, dv \tag{E.5}$$

is

$$\widetilde{B} = \frac{2\sqrt{n-1}}{n} \sum_{i=1}^{n} \widehat{R}_i, \tag{E.6}$$

where c is the copula density of (X, Y). Since this basic estimator may be biased in finite samples, Geenens and Lafaye de Micheaux (2022) propose a normalised version $\widetilde{B}_{KL}$, the details of which are omitted here; see the original paper for the full construction. The empirical Hellinger correlation is then obtained by applying the transformation

$$\widehat{\eta}_{KL} = \frac{2}{\widetilde{B}_{KL}^2} \left[\widetilde{B}_{KL}^4 + \sqrt{4 - 3\widetilde{B}_{KL}^4} - 2 \right]^{1/2}. \tag{E.7}$$

From a computational perspective, the estimator remains practical for moderate sample sizes. The main computational cost comes from the nearest-neighbour calculations and the cross-validation step used to select the tuning parameters. In practice, these steps are implemented automatically in the `R` package `HellCor`.

E.2. Monte Carlo simulation for hypothesis testing

The idea of a Monte Carlo test is to approximate the null distribution of a test statistic by repeatedly generating samples under the null hypothesis and recomputing the statistic on each simulated sample.

For testing whether two continuous random variables are independent using the Hellinger correlation, we consider the hypothesis

$$\begin{aligned} H_0 &: \eta = 0, \\ H_1 &: \eta \neq 0. \end{aligned} \tag{E.8}$$

Under H_0, the two random variables are independent, so the associated copula is the independence copula. Hence, the null distribution of the empirical Hellinger correlation can be approximated by repeatedly generating independent bivariate uniform samples and computing the corresponding empirical Hellinger correlation on each sample.

Specifically, suppose that we generate M Monte Carlo samples under H_0 and compute the corresponding empirical Hellinger correlations $\hat{\eta}_1, \ldots, \hat{\eta}_M$. Let $\hat{\eta}$ denote the empirical Hellinger correlation computed from the original observations. The Monte Carlo p-value is then approximated by

$$p = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}(\hat{\eta}_m > \hat{\eta}). \tag{E.9}$$

That is, the p-value is the proportion of simulated statistics under the null hypothesis that exceed the observed value of the test statistic. This can be computed using the `HellCor` function with the option `pval.comp = TRUE`. From a computational perspective, this testing procedure is more demanding than computing the empirical Hellinger correlation alone, since the statistic must be recomputed repeatedly on many Monte Carlo samples under the null hypothesis. In the current `HellCor` implementation, the p-value is obtained by Monte Carlo simulation, and will become computationally intensive for large sample sizes or when many samples are generated.

F. Details for distance covariance

F.1. Definition of distance covariance in detail

As stated in Equation (3.7), the distance covariance between two random vectors $\boldsymbol{X} \in \mathbb{R}^p$ and $\boldsymbol{Y} \in \mathbb{R}^q$ is given by

$$\|\varphi_{\boldsymbol{XY}}(\boldsymbol{s},\boldsymbol{t}) - \varphi_{\boldsymbol{X}}(\boldsymbol{s})\varphi_{\boldsymbol{Y}}(\boldsymbol{t})\|_w^2 = \int_{\mathbb{R}^{p+q}} |\varphi_{\boldsymbol{XY}}(\boldsymbol{s},\boldsymbol{t}) - \varphi_{\boldsymbol{X}}(\boldsymbol{s})\varphi_{\boldsymbol{Y}}(\boldsymbol{t})|^2 \, w(\boldsymbol{s},\boldsymbol{t}) \, d\boldsymbol{s} \, d\boldsymbol{t}. \tag{F.1}$$

Here, the characteristic function of $\boldsymbol{X}$ is given by

$$\varphi_{\boldsymbol{X}}(\boldsymbol{s}) = \mathbb{E}\left[e^{i\boldsymbol{s}^\top \boldsymbol{X}}\right], \tag{F.2}$$

where $i = \sqrt{-1}$ is the imaginary unit, i.e. one of the solutions to the equation $x^2 = -1$. The characteristic function of $\boldsymbol{Y}$ is defined analogously. In addition, the joint characteristic function of $(\boldsymbol{X}, \boldsymbol{Y})$ is given by

$$\varphi_{\boldsymbol{XY}}(\boldsymbol{s},\boldsymbol{t}) = \mathbb{E}\left[e^{i\boldsymbol{s}^\top \boldsymbol{X} + i\boldsymbol{t}^\top \boldsymbol{Y}}\right]. \tag{F.3}$$

Furthermore, notice that characteristic functions can be complex-valued, and hence $\varphi_{\boldsymbol{XY}}(\boldsymbol{s},\boldsymbol{t}) - \varphi_{\boldsymbol{X}}(\boldsymbol{s})\varphi_{\boldsymbol{Y}}(\boldsymbol{t})$ can also be complex-valued. Let $f(x)$ be a complex-valued function. Then

$$|f(x)|^2 = \overline{f(x)} \cdot f(x), \tag{F.4}$$

where $\overline{f(x)}$ is the complex conjugate of $f(x)$, that is, if $z = a + b\sqrt{-1} \in \mathbb{C}$, then the complex conjugate is $\overline{z} = a - b\sqrt{-1}$.

The weight function is defined as

$$w(\boldsymbol{s},\boldsymbol{t}) = (c_p c_q \|\boldsymbol{s}\|^{p+1} \|\boldsymbol{t}\|^{q+1})^{-1}, \tag{F.5a}$$

$$c_d = \frac{\pi^{(1+d)/2}}{\Gamma((1+d)/2)}, \tag{F.5b}$$

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \tag{F.5c}$$

where for a vector $\boldsymbol{z} = (z_1, \ldots, z_d)$,

$$\|\boldsymbol{z}\| = \sqrt{\sum_{k=1}^{d} z_k^2}. \tag{F.6}$$

As discussed in Section 3.2.1, there are many possible choices of weight function. In this paper, we adopt the one used by Székely et al. (2007), since it leads to a neat estimator that does not require direct computation or estimation of the characteristic functions, as we discuss below.

F.2. Empirical estimator of the distance covariance

Let $\{(\boldsymbol{x}_j, \boldsymbol{y}_j)\}_{j=1}^n$ be an i.i.d. sample from the joint distribution of $(\boldsymbol{X}, \boldsymbol{Y})$, where $\boldsymbol{x}_j \in \mathbb{R}^p$ and $\boldsymbol{y}_j \in \mathbb{R}^q$ for $j = 1, \ldots, n$. To compute the empirical estimator of distance covariance, we first compute the $\mathcal{U}$-centered distance matrices, denoted by $\widetilde{U}_{\boldsymbol{x}}$ and $\widetilde{U}_{\boldsymbol{y}}$, respectively. The $\mathcal{U}$-centered matrix

$\widetilde{U}_{\boldsymbol{x}}$ is defined by

$$\widetilde{U}_{\boldsymbol{x}}(i,j) = \begin{cases} \|\boldsymbol{x}_i - \boldsymbol{x}_j\| - \dfrac{1}{n-2}\displaystyle\sum_{l=1}^{n}\|\boldsymbol{x}_i - \boldsymbol{x}_l\| \\ \qquad - \dfrac{1}{n-2}\displaystyle\sum_{k=1}^{n}\|\boldsymbol{x}_k - \boldsymbol{x}_j\| & i \neq j, \\ \qquad + \dfrac{1}{(n-1)(n-2)}\displaystyle\sum_{k=1}^{n}\sum_{l=1}^{n}\|\boldsymbol{x}_k - \boldsymbol{x}_l\|, \\ 0, & i = j. \end{cases} \tag{F.7}$$

and $\widetilde{U}_{\boldsymbol{y}}$ is defined analogously.

The unbiased estimator of the *squared* distance covariance is then given by

$$\widetilde{dCov}^2(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{n(n-3)}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{U}_{\boldsymbol{x}}(i,j)\widetilde{U}_{\boldsymbol{y}}(i,j). \tag{F.8}$$

This estimator is unbiased for the population quantity $dCov^2(\boldsymbol{X},\boldsymbol{Y})$, provided that $n > 3$. From a computational perspective, the main burden in estimating distance covariance lies in constructing the pairwise distance matrices for $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ and $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n$, which requires $O(n^2)$ pairwise distance calculations and $O(n^2)$ memory. Once these matrices are obtained, the $\mathcal{U}$-centering step and the summation in Equation (F.8) can be implemented in $O(n^2)$ time. Hence, the estimator is generally practical for moderate sample sizes, but may become computationally demanding for large datasets.

F.3. Permutation test for distance covariance

Consider two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ with observed sample $\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}_{i=1}^n$ of size n, and let $\widetilde{dCov}^2(\boldsymbol{X},\boldsymbol{Y})$ denote the estimated squared distance covariance. We wish to test the hypothesis

$$\begin{aligned} H_0 &: \boldsymbol{X} \perp \boldsymbol{Y}, \\ H_1 &: \boldsymbol{X} \not\perp \boldsymbol{Y}. \end{aligned} \tag{F.9}$$

We conduct the permutation test by randomly permuting the sample of $\boldsymbol{Y}$ while keeping the sample of $\boldsymbol{X}$ fixed. Let π be a random permutation of the index set $\{1,\ldots,n\}$, that is, a random reordering of the indices. For example, when $n = 4$, valid permutations include $(2,4,3,1)$ and $(4,1,2,3)$. Using such a permutation, we form the permuted sample

$$\{(\boldsymbol{x}_i,\boldsymbol{y}_{\pi(i)})\}_{i=1}^n \tag{F.10}$$

and compute the corresponding squared distance covariance, denoted by $\widetilde{dCov}_\pi^2(\boldsymbol{X},\boldsymbol{Y})$.

Repeating this procedure B times yields

$$\widetilde{dCov}_{\pi_1}^2(\boldsymbol{X},\boldsymbol{Y}),\ldots,\widetilde{dCov}_{\pi_B}^2(\boldsymbol{X},\boldsymbol{Y}), \tag{F.11}$$

where each π_b is an independent random permutation of $\{1,\ldots,n\}$. The permutation p-value is then computed as

$$p = \frac{1 + \sum_{b=1}^{B}\mathbf{1}\Big(\widetilde{dCov}_{\pi_b}^2(\boldsymbol{X},\boldsymbol{Y}) \geq \widetilde{dCov}^2(\boldsymbol{X},\boldsymbol{Y})\Big)}{B+1}. \tag{F.12}$$

The permutation test is more demanding than point estimation alone. The main burden comes from two sources. First, computing the distance covariance for a single sample requires forming pairwise distances, which is quadratic in the sample size. Second, the permutation test recomputes the same statistic over many randomly permuted samples. Hence, if B permutations are used, the overall computational cost is roughly $O(Bn^2)$, with $O(n^2)$ memory required if the full distance matrices are stored. As a result, the procedure is generally feasible for moderate sample sizes, but can become computationally intensive for large datasets or when a large number of permutations is required.

G. Details for joint distance covariance

G.1. Weight function in joint distance covariance

In Equation (4.2), the integral relies on the weight function $w(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_d)$ for d random vectors $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d)$. The weight function is defined as

$$w(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_d) = \left(c_{p_1} \cdots c_{p_d} \|\boldsymbol{t}_1\|^{p_1+1} \cdots \|\boldsymbol{t}_k\|^{p_d+1} \right)^{-1}, \tag{G.1}$$

where p_i is the dimension of $\boldsymbol{X}_i$ for $i = 1, \ldots, d$, and the constant c_{p_i} is identically defined as in Equation (F.5).

G.2. Empirical estimator of the distance covariance

The unbiased estimator of the *squared* joint distance covariance can be written as

$$\widetilde{JdCov}^{\,2}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) = \frac{1}{n(n-3)} \sum_{k=1}^{n} \sum_{l=1}^{n} \prod_{i=1}^{d} \left(\widetilde{U}_{\boldsymbol{x}_i}(k,l) + 1 \right) - \frac{n}{n-3}, \tag{G.2}$$

where $\{(\boldsymbol{x}_{1,j}, \ldots, \boldsymbol{x}_{d,j})\}_{j=1}^{n}$ is the i.i.d. sample from $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d)$ with $n > 3$. Here, $\widetilde{U}_{\boldsymbol{x}_i}$ denotes the $\mathcal{U}$-centered distance matrix for the sample from the i-th random vector, defined analogously to the $\mathcal{U}$-centered matrix used in Equation (F.8).

G.3. Permutation test for joint distance covariance

Consider d random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_d$ with estimated joint distance covariance $\widetilde{JdCov}^{\,2}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d)$. We wish to test the hypothesis

$$\begin{aligned} &H_0 : \boldsymbol{X}_1 \perp \boldsymbol{X}_2 \perp \cdots \perp \boldsymbol{X}_d, \\ &H_1 : \boldsymbol{X}_1, \ldots, \boldsymbol{X}_d \text{ are not mutually independent.} \end{aligned} \tag{G.3}$$

This can be done similarly using a permutation test. For the sample

$$\{(\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,d})\}_{i=1}^{n}, \tag{G.4}$$

we construct a permuted sample by randomly permuting $\boldsymbol{x}_{i,2}, \ldots, \boldsymbol{x}_{i,d}$ while keeping $\boldsymbol{x}_{i,1}$ fixed, in the same way as in Appendix F.3. More precisely, for each permutation replicate $b = 1, \ldots, B$, let $\pi_2^{(b)}, \ldots, \pi_d^{(b)}$ be independent random permutations of the index set $\{1, \ldots, n\}$. The corresponding permuted sample is then

$$\left\{ \left(\boldsymbol{x}_{i,1}, \boldsymbol{x}_{\pi_2^{(b)}(i),2}, \boldsymbol{x}_{\pi_3^{(b)}(i),3}, \ldots, \boldsymbol{x}_{\pi_d^{(b)}(i),d} \right) \right\}_{i=1}^{n}. \tag{G.5}$$

For each such permuted sample, we estimate the squared joint distance covariance, yielding

$$\widetilde{JdCov}^{\,2}_{\pi^{(1)}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d), \ldots, \widetilde{JdCov}^{\,2}_{\pi^{(B)}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d). \tag{G.6}$$

The p-value is then computed as

$$p = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\left(\widetilde{JdCov}^{\,2}_{\pi^{(b)}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) \geq \widetilde{JdCov}^{\,2}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d) \right)}{B+1}. \tag{G.7}$$

G.4. Illustration of zero 3rd-order distance covariance with non-zero 2nd-order distance covariance

Consider three random variables X_1, X_2, and X_3 such that

$$\begin{aligned}\mathbb{P}\big((X_1, X_2) = (0,0)\big) = \mathbb{P}\big((X_1, X_2) = (1,1)\big) = \frac{3}{8},\\ \mathbb{P}\big((X_1, X_2) = (0,1)\big) = \mathbb{P}\big((X_1, X_2) = (1,0)\big) = \frac{1}{8},\end{aligned} \tag{G.8}$$

and let $X_3 \sim \text{Bernoulli}\big(\frac{1}{2}\big)$ be independent of (X_1, X_2). Here, X_1 and X_2 are not independent, since

$$\mathbb{P}(X_1 = 1, X_2 = 1) = \frac{3}{8} \neq \frac{1}{4} = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1). \tag{G.9}$$

The 3rd-order distance covariance of (X_1, X_2, X_3) is given by

$$dCov^2(X_1, X_2, X_3) = \int \left| \mathbb{E}\left[\prod_{j=1}^{3} \left(\varphi_{X_j}(t_j) - e^{it_j X_j}\right) \right] \right|^2 w(t_1, t_2, t_3)\, dt_1 dt_2 dt_3. \tag{G.10}$$

Since X_3 is independent of (X_1, X_2), we have

$$\mathbb{E}\left[\prod_{j=1}^{3} \left(\varphi_{X_j}(t_j) - e^{it_j X_j}\right) \right] = \mathbb{E}\left[\left(\varphi_{X_1}(t_1) - e^{it_1 X_1}\right) \left(\varphi_{X_2}(t_2) - e^{it_2 X_2}\right) \right] \mathbb{E}\left[\varphi_{X_3}(t_3) - e^{it_3 X_3} \right]. \tag{G.11}$$

But

$$\mathbb{E}\left[\varphi_{X_3}(t_3) - e^{it_3 X_3} \right] = \varphi_{X_3}(t_3) - \varphi_{X_3}(t_3) = 0, \tag{G.12}$$

so the integrand in Equation (G.10) is identically zero, and hence

$$dCov^2(X_1, X_2, X_3) = 0. \tag{G.13}$$

In contrast, we have $dCov^2(X_1, X_2) > 0$, since X_1 and X_2 are not independent. Therefore, the 3rd-order distance covariance alone does not capture 2nd-order pairwise dependence. This illustrates why lower- and higher-order terms need to be combined in order to characterise mutual independence.

G.5. Details of `pg15training`

Table G.1: Details of variables in `pg15training` used for model training

Variable	Type and Range	Pre-processing	Notes
Response Variable			
`nclaims`	Count (0–7)	None	Renamed from `Numtppd`, counts third-party material claims.
Driver's attributes			
`Female`	Categorical (Male, Female)	One-hot encoding (Male=0, Female=1)	Renamed from `Gender`. Protected attribute representing the policyholder's gender. Excluded from model input.
`Occupation`	Categorical (5 classes)	One-hot encoding	Includes Employed, Unemployed, Housewife, Self-employed, Retired.
`Region`	Categorical (10 classes)	One-hot encoding	Protected attribute representing the policyholder's living region. Renamed from `Group2` with 10 classes (region L to region U), excluded from input.
`Age`	Integer (18–75)	Min-max scaling	–
`Bonus`	Integer (-50–150)	Min-max scaling	Represents no-claim discount (bonus-malus).
`Exposure`	Continuous (91–365 days)	Converted to years	Renamed from `Exppdays`, used as model offset.
`PolDur`	Continuous (0–15)	Min-max scaling	Policy duration in years.
`Density`	Continuous (14.38–297.39)	Min-max scaling	Population density per km^2.
Vehicle's attributes			
`CarCat`	Categorical (Small, Medium, Large)	Ordinal encoding (Small=0, Medium=1, Large=2)	Renamed from `Category`.
`CarType`	Categorical (A–F)	One-hot encoding	Renamed from `Type`.
`CarGroup`	Categorical (1–20)	One-hot encoding	Renamed from `Group1`.
`Value`	Continuous (1,000–49,995)	Min-max scaling	Vehicle's market value in Euros.

H. Details for (multivariate) auto-distance correlation function

H.1. Empirical estimator of the auto-distance correlation function

Let $\{\boldsymbol{X}_t\}_{t=1}^n$ be a time series, where $\boldsymbol{X}_t \in \mathbb{R}^p$ for $t = 1, \ldots, n$. For a fixed lag $h \geq 0$, consider the paired sample

$$\{(\boldsymbol{X}_t, \boldsymbol{X}_{t+h})\}_{t=1}^{n-h}. \tag{H.1}$$

To estimate the auto-distance covariance at lag h, we first define the pairwise distance matrices

$$a_{ij} = \|\boldsymbol{X}_i - \boldsymbol{X}_j\|, \qquad b_{ij} = \|\boldsymbol{X}_{i+h} - \boldsymbol{X}_{j+h}\|, \qquad i, j = 1, \ldots, n-h. \tag{H.2}$$

We then form the corresponding doubly centred distance matrices, similar to the estimation of distance covariance in Equation (F.7):

$$A_{ij} = a_{ij} - \bar{a}_{i\cdot} - \bar{a}_{\cdot j} + \bar{a}_{\cdot\cdot}, \tag{H.3}$$

and

$$B_{ij} = b_{ij} - \bar{b}_{i\cdot} - \bar{b}_{\cdot j} + \bar{b}_{\cdot\cdot}, \tag{H.4}$$

where

$$\bar{a}_{i\cdot} = \frac{1}{n-h}\sum_{l=1}^{n-h} a_{il}, \qquad \bar{a}_{\cdot j} = \frac{1}{n-h}\sum_{k=1}^{n-h} a_{kj}, \qquad \bar{a}_{\cdot\cdot} = \frac{1}{(n-h)^2}\sum_{k=1}^{n-h}\sum_{l=1}^{n-h} a_{kl}, \tag{H.5}$$

and similarly for $\bar{b}_{i\cdot}$, $\bar{b}_{\cdot j}$, and $\bar{b}_{\cdot\cdot}$. The sample auto-distance covariance at lag h is then given by

$$\hat{\mathcal{V}}^2_{\boldsymbol{X}}(h) = \frac{1}{(n-h)^2}\sum_{i=1}^{n-h}\sum_{j=1}^{n-h} A_{ij}B_{ij}. \tag{H.6}$$

The sample auto-distance correlation function is then defined as

$$\hat{\mathcal{R}}^2_{\boldsymbol{X}}(h) = \sqrt{\frac{\hat{\mathcal{V}}^2_{\boldsymbol{X}}(h)}{\hat{\mathcal{V}}^2_{\boldsymbol{X}}(0)}}. \tag{H.7}$$

H.2. Empirical estimator of the multivariate auto-distance correlation function

Let $\{\boldsymbol{X}_t\}_{t=1}^n$ be a d-variate time series, where

$$\boldsymbol{X}_t = (X_{t;1}, \ldots, X_{t;d})^\top. \tag{H.8}$$

For a fixed lag $h \geq 0$ and for each pair of components $(r, m) \in \{1, \ldots, d\}^2$, consider the paired sample

$$\{(X_{t;r}, X_{t+h;m})\}_{t=1}^{n-h}. \tag{H.9}$$

Define the pairwise distance matrices

$$a^{(r)}_{ij} = |X_{i;r} - X_{j;r}|, \qquad b^{(m)}_{ij} = |X_{i+h;m} - X_{j+h;m}|, \qquad i, j = 1, \ldots, n-h. \tag{H.10}$$

We then form the corresponding doubly centred distance matrices

$$A^{(r)}_{ij} = a^{(r)}_{ij} - \bar{a}^{(r)}_{i\cdot} - \bar{a}^{(r)}_{\cdot j} + \bar{a}^{(r)}_{\cdot\cdot}, \tag{H.11}$$

and

$$B^{(m)}_{ij} = b^{(m)}_{ij} - \bar{b}^{(m)}_{i\cdot} - \bar{b}^{(m)}_{\cdot j} + \bar{b}^{(m)}_{\cdot\cdot}, \tag{H.12}$$

where

$$\bar{a}_{i\cdot}^{(r)} = \frac{1}{n-h}\sum_{l=1}^{n-h} a_{il}^{(r)}, \qquad \bar{a}_{\cdot j}^{(r)} = \frac{1}{n-h}\sum_{k=1}^{n-h} a_{kj}^{(r)}, \qquad \bar{a}_{\cdot\cdot}^{(r)} = \frac{1}{(n-h)^2}\sum_{k=1}^{n-h}\sum_{l=1}^{n-h} a_{kl}^{(r)}, \tag{H.13}$$

and similarly for $\bar{b}_{i\cdot}^{(m)}$, $\bar{b}_{\cdot j}^{(m)}$, and $\bar{b}_{\cdot\cdot}^{(m)}$.

The component-wise sample auto-distance covariance at lag h is then

$$\hat{\mathcal{V}}_{rm}^2(h) = \frac{1}{(n-h)^2}\sum_{i=1}^{n-h}\sum_{j=1}^{n-h} A_{ij}^{(r)} B_{ij}^{(m)}. \tag{H.14}$$

Collecting these entries for all $r, m = 1, \ldots, d$ gives the sample auto-distance covariance matrix

$$\hat{\mathcal{V}}^{(2)}(h) = \big(\hat{\mathcal{V}}_{rm}^2(h)\big)_{r,m=1}^{d}. \tag{H.15}$$

The corresponding sample auto-distance correlation matrix is obtained by normalising each entry as

$$\hat{\mathcal{R}}_{rm}^2(h) = \frac{\hat{\mathcal{V}}_{rm}^2(h)}{\big\{\hat{\mathcal{V}}_{rr}^2(0)\hat{\mathcal{V}}_{mm}^2(0)\big\}^{1/2}}, \tag{H.16}$$

provided that $\hat{\mathcal{V}}_{rr}(0)\hat{\mathcal{V}}_{mm}(0) \neq 0$, and

$$\hat{\mathcal{R}}^{(2)}(h) = \big(\hat{\mathcal{R}}_{rm}^2(h)\big)_{r,m=1}^{d}. \tag{H.17}$$

H.3. Details of the GARCH(1,1) model for S&P-500 log monthly return

We fitted a standard GARCH$(1,1)$ model with a constant conditional mean and standardised Student-t shocks to the monthly log returns of the S&P500 index. The model is specified as

$$\begin{aligned} r_t &= \mu + \epsilon_t, \\ \epsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \tag{H.18}$$

where z_t is i.i.d. standardised Student-t with shape parameter ν. Using the fitted model, the estimated specification is

$$\begin{aligned} r_t &= 0.008671 + \epsilon_t, \\ \epsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= 0.000097 + 0.132827\,\epsilon_{t-1}^2 + 0.831858\,\sigma_{t-1}^2, \\ z_t &\sim t_\nu(0,1), \qquad \nu = 6.693353. \end{aligned} \tag{H.19}$$

H.4. Details of the VAR(2) model for LA mortality data

The fitted VAR(2) model is

$$\Delta_{52}\boldsymbol{X}_t = \boldsymbol{c} + A_1\Delta_{52}\boldsymbol{X}_{t-1} + A_2\Delta_{52}\boldsymbol{X}_{t-2} + \boldsymbol{\varepsilon}_t, \tag{H.20}$$

where

$$\boldsymbol{X}_t = \begin{pmatrix} \texttt{cmort}_t \\ \texttt{tempr}_t \\ \texttt{part}_t \end{pmatrix}, \qquad \boldsymbol{c} = \begin{pmatrix} -0.547469 \\ -0.153940 \\ -0.723195 \end{pmatrix},$$

$$A_1 = \begin{pmatrix} 0.259392 & -0.135324 & -0.078177 \\ -0.068930 & 0.042400 & -0.021730 \\ 0.022020 & -0.178370 & 0.000446 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0.321994 & 0.006543 & -0.026449 \\ 0.014650 & 0.078810 & -0.037450 \\ -0.193402 & -0.107707 & 0.161109 \end{pmatrix}.$$