Causal Discovery via Simultaneous DAG Recovery Using the Angles Space of Directional Dependence Measures

JD Opdyke, Chief Analytics Officer, DataMineit, LLC

Abstract

Most causal discovery algorithms utilizing a Directed Acyclic Graph (DAG) framework execute sequentially (e.g. constraint-based models) or iteratively (e.g. score-based or functional causal models). In contrast, we develop a new method – Angles-based Directional Dependence (ADD) – that executes over the entire DAG space simultaneously, based on only two matrix estimations. We apply dual orderings on any (positive definite) directional dependence measure for all pairwise relationships, identify statistically significant directional dependence in the angles space (where angles are independent random variables), and then enforce acyclicality to make proper causal interpretations. Potential benefits of the approach include increased coherence, due to simultaneous edge-calling within a positive definite space; increased power, due to the ability to use any directional dependence measure and thus, opportunistically adapt to different or varying data conditions; and increased speed/scalability, due to the need for only two matrix estimations, and two (fast) simulations to define empirical confidence bounds under independence (regardless of the size of the DAG space). We conduct a preliminary empirical study evaluating directional dependence marginally, effectively identifying directed reachability under nonlinear, heavy-tailed data conditions and thus, indicating applications to feature selection in quantitative finance. The promising results justify and encourage a more extensive follow-up study to benchmark against competing algorithms to more fully test the above-mentioned potential benefits of ADD.

Keywords: directional dependence, causality, causal discovery, DAG, directed acyclical graph, causal modeling, quantitative finance, positive definite, feature selection, dependence structure, Pearson's, covariance, correlation, Chatterjee's

Acknowledgments: I thank Alejandro Dominguez Rodriguez, PhD, for his very insightful comments and useful suggestions on earlier drafts of this paper.

Contents

1. Introduction

Directed Acyclic Graphs (DAGs) are a widely used and well-established framework for causal modeling,[1] which includes both causal discovery and causal inference (see Zanga et al., 2025). The former identifies causal relationships and the latter quantifies them. This paper focuses only on causal discovery, seeking to identify causal structure consistent with the DAG representing the 'true' network of causal relationships. Many algorithms designed to recover DAG 'truth' execute sequentially. These include constraint-based methods, such as PC algorithms (see Kalisch, M., & Bühlmann, P. (2007)), FCI algorithms (see Sprites (2001)), and Markov Blanket algorithms (see Aliferis et al. (2010), and Chen et al., (2026)). Most of the other approaches execute iteratively. These include score-based methods, such as NOTEARS (see Zheng et al. (2018)) and its variants (e.g. DYNOTEARS, see Pamfil et al. (2020)), Greedy Equivalence Search (see Chickering, D., (2002) and Hauser and Bühlmann (2012)) and Kernel-based Score Functions (see Huang et al., (2018) and Wang et al., (2024)). They also include functional causal models, such as LiNGAM (see Shimizu et al. (2006)), and its variants such as VarLiNGAM (see Hyvärinen et al. (2010)). In contrast, the method developed herein – Angles-based Directional Dependence (ADD) – estimates the entire DAG structure simultaneously, via dual orderings using any positive definite directional dependence measure. So ADD requires only two estimations of the all-pairwise dependence matrix, one in each direction, and two simulations to establish empirical confidence bounds, under independence, for each cell of the matrix. Inference takes place in the angles space, where angles are independent random variables by their orthogonal construction.[2] Potential benefits of this approach include increased coherence, due to simultaneous edge-calling within a positive definite space;[3] increased power, due to the ability to use any (positive definite) directional dependence measure and thus, opportunistically adapt to different or varying data conditions; and increased speed/scalability, due to the need for only two matrix estimations and two (fast) simulations (regardless of the size of the DAG space). This describes ADD's recovery of directional dependence structure, but making a valid causal interpretation requires satisfying additional assumptions specified below (see A1-A7 in Appendix C).[4]

---

[1] Of course, DAGs are not the only framework for causal discovery. The potential outcomes framework (see Angrist et al. (1996) and Imbens and Rubin (2015)) is the major competitor to Structural Causal Models (the overarching framework utilizing DAGs). Czado (2025) also demonstrates that vine copulas can be remarkably effective in the causal discovery setting. And Rodriguez Dominguez & Yadav (2024) and Rodriguez Dominguez (2026) propose approaches to causal measurement based on system-level and spectral properties, respectively. I address potential and actual limitations of DAGs when used for causal discovery in later sections herein.

[2] See Rapisarda et al. (2007) and Zhang et al. (2015) for detailed descriptions.

[3] The "wobbly world" of causal graphs, often characterized by severe instability and lack of coherence in edge-calling across a DAG, is convincingly tested and demonstrated in Hulse et al. (2025). Similarly, Faltenbacher et al. (2026) carefully document the sources and magnitude of incoherent errors in DAG-based causal discovery. Finally, Padh et al. (2025) describe how a popular class of DAG recovery algorithms (i.e. 'score-based') critically depend on heuristic hyperparameter choices and remain sensitive to threshold choices.

[4] Briefly, these include acyclicity, causal sufficiency, Markov and faithfulness, independent and identically distributed sampling, reliable measurement, and monotone link from (directional) dependence measure to asymmetry.

We first review the long, and more recent history of directional dependence measures, and their more recent use and utility for causal modeling. Next, we do the same for the use of the angles space in measuring and estimating dependence in a wide range of settings. Finally, we combine the two to develop ADD, which utilizes the (positive definite) angles space of (positive definite) directional dependence measures to identify causal relationships within the DAG framework. We conduct a preliminary empirical study to test ADD on highly nonlinear and heavy-tailed data generated by a straightforward, but non-toy DAG. While adjacency recovery requires conditioning, which was not done in this study, the results remain promising, indicating applications to feature selection in quantitative finance and justifying further inquiry. The latter takes the form of a proposed, more extensive study explicitly benchmarking ADD against causal discovery competitors to more fully test its proposed benefits.

2. Background

2.A. Asymmetric, Directional Dependence

We present a brief survey of the relevant literature on directional dependence here, with a focus on readily implemented methods, because this is central to the development of ADD, and these specific methods (covering (1) through (6), and (10) through (12) below) have been extensively tested empirically using ADD. The concept of asymmetric, directional dependence is not new. Recent research developing such measures goes back over a dozen years (see Zheng et al., 2012), but has its direct origins in work done at the end of the nineteenth century (see Yule, 1897, and a related review in Allena and McAleerb (2018)). As discussed in more detail below in the section covering the angles space, ADD's reliance on the Cholesky factor requires the dependence measures it uses to be positive definite. This is a defining feature of dependence measures generally, so for clarity throughout this paper, and with practical, empirical settings in mind, we categorize dependence measures vis-à-vis positive definiteness into three mutually exclusive and exhaustive categories:

C1. Measures that have been analytically proven to be positive definite, and which only exhibit non-positive definiteness empirically under select conditions due to numerical issues. Examples include Pearson's rho, Spearman's rho, and Kendall's tau (see Sabato (2007) for corresponding proofs) as well as the tail dependence matrix (see Embrechts (2016) for a corresponding proof). Note that none of these measures are directional.

C2. Measures for which positive definiteness has not been proven analytically, but which, via extensive empirical simulations and testing under wide-ranging conditions, do not violate positive definiteness any more frequently than those measures from category C1., thus indicating that such violations are likely due only to numerical issues. Examples include the directional measures (1) through (6), and (10) through (12) presented below.

C3. Measures for which positive definiteness has not been proven analytically, and which exhibit non-positive definiteness more often than numerical issues alone would indicate.

The directional dependence measures reviewed below, used with ADD, and used in the empirical study herein all fall into category C2. While empirical findings should never be taken as evidence on the same level as analytical proof, the development and testing of ADD, using all eight of the category C2. measures, required many millions of simulations under wide-ranging data conditions, and these never once yielded a single non-positive definite result: hence their classification in category C2.. Best practices dictate that positive definiteness always should be empirically verified anyway, even for C1. measures, due to the possibility of numerical issues, especially when the factor space is not small (spaces with #factors = 100 were tested), so we do not view the lack of analytic proofs for C2. measures as posing a limiting restriction on their use for these purposes in practice. If the rare edge case yields a non-positive definite matrix due to numerical issues, one can use a 'fixing' algorithm whereby the dependence measure matrix is projected to the nearest positive definite matrix (see the seminal work of Higham (2002) for a widely used example).

Category C3. measures, on the other hand, warrant more caution. While one can always mechanically turn to algorithms like Higham (2002) to enforce positive definiteness, if this occurs due to the construction of the measure and not due solely to numerical issues when performing sampling, this approach would not be advisable. For example, if we observed non-positive definiteness in more than, say, one or two percent of samples (admittedly a subjective threshold), we would avoid this 'matrix-fix' approach (and likely avoid the dependence measure altogether).[5] The main problem is that such an approach would materially distort the sample space of the dependence measure, making inferences based on such an automated, but statistically invalid ruleset potentially highly misleading. The bottom line here is that ADD requires the dependence measures it uses to be positive definite, and implementation of those that fall into category C2., all of which have been tested extensively by ADD, is defensible for practical, scientifically vigilant application in this setting. We review these eight measures below.

A recent example of a directional dependence measure that garnered much attention upon its publication in 2021 is Chatterjee's new correlation coefficient (chcorr). This is largely due to its simplicity and ease of implementation as a measure of non-linear, non-monotonic, regression-based, and cyclical dependence. If observation pairs of X and Y variables are ranked according to X values, with no ties on the X values, so that $\left(\left(X_{(1)},Y_{(1)}\right),\left(X_{(2)},Y_{(2)}\right),\cdots,\left(X_{(n)},Y_{(n)}\right)\right)$ then:

[5] To flip this script on this requirement of positive definiteness, one reasonably could argue that if a dependence measure was analytically shown to be non-positive definite, at least under relevant conditions, and/or its empirical estimates were non-positive definite more often than could be attributable solely to numeric considerations, then researchers and practitioners might want to question the wisdom of using it. Non-positive definiteness also could be a function of unknown or mis-specified data conditions, such as perfect linear dependence unwittingly built into a simulation (although with actual market data, it could be a very useful flag for extreme multicollinearity). Or non-positive definiteness perhaps could be due to a combination of the dependence measure used and the specific data conditions being examined. Either way, the non-positive definite results could be serving as a correct and useful warning to avoid the dependence measure (and/or those simulated data conditions) altogether. In such cases, the requirement of positive definiteness is less a limitation of a method like ADD and more a proper boundary on the right measures and conditions under which such analyses should be conducted in the first place.

(1) $chcorr = \xi_n\left(X,Y\right) := 1 - \frac{3\sum_{i=1}^{n-1}\left|r_{i+1} - r_i\right|}{n^2 - 1}$ where r_i = rank of Y_i

Under ties for some of the X values, break ties uniformly at random, and

(2) $chcorr = \xi_n\left(X,Y\right) := 1 - \frac{n\sum_{i=1}^{n-1}\left|r_{i+1} - r_i\right|}{2\sum_{i=1}^{n} l_i\left(n - l_i\right)}$ where $l_i = \#j$ such that $Y_{(j)} \geq Y_{(i)}$

Chatterjee's new correlation coefficient ranges from zero to one asymptotically (it can exceed these bounds slightly under finite samples), and unlike many commonly used dependence measures, a value of zero indicates independence. Also unlike some other measures, no positive or negative dependence is indicated by a positive or negative sign on the measure value.[6] As an asymmetric dependence measure, the order of X and Y matters: $\xi_n\left(X,Y\right)$ does not necessarily equal $\xi_n\left(Y,X\right)$, by design. In other words, the dependence of Y on X is not assumed to be identical to the dependence of X on Y, respectively: dependence is *directional*.[7] However, note that Chatterjee's can be made to be symmetric by simply taking the maximum of two measures, one in each direction as in (3):

(3) $chcorr\_sym = \max\left[\xi_n\left(X,Y\right), \xi_n\left(Y,X\right)\right]$

Chatterjee's breakthrough has spawned many variants (see Lin & Han (2023), Pascual-Marqui et al. (2024), and especially Gao and Li (2024)). One of these is the "improved Chatterjee's correlation" (ichcorr, or "ICH") derived by Xia et al. (2025)), the motivation of which is to increase power by using inverse distance weightings of all neighboring data values as opposed to just one.

(4) $ichcorr = \xi_n^{IM}\left(X,Y\right) = 1 - \frac{\sum_{i \neq j}^{n}\left|r_i - r_j\right| / \left|i - j\right|}{\frac{n+1}{3}\sum_{i \neq j}^{n}\left|i - j\right|}$

[6] Recall, of course, that widely used non-directional dependence measures such as Pearson's rho, Spearman's rho, and Kendall's tau have (maximum) ranges of –1 to 1, indicate positive/negative dependence with their signs, and values of zero do not necessarily indicate independence. A counter example of the latter, however, is Pearson's under Gaussian data, whereby a value of zero *does* indicate independence.

[7] This holds true for all asymmetric/directional dependence measures, so importantly, the corresponding all-pairwise matrix remains symmetric. In other words, when using, say, Chatterjee (2021), on two factors, say, X3 and X4, the value in cell row 3, column 4 of the matrix is $\xi_n\left(X3,X4\right)$, and the value in cell row 4, column 3 of the matrix is identical, that is, $\xi_n\left(X3,X4\right)$; it is NOT $\xi_n\left(X4,X3\right)$.

Xia et a. (2025) test the power and level of "improved Chatterjee" against other dependence measures, including Chatterjee's, in an empirical study under wide-ranging data conditions. Both Chatterjee's and improved Chatterjee's coefficients exhibit good power under non-monotonic, non-linear, and cyclical dependence, with the latter measure usually winning. Like Chatterjee's, ICH ranges from zero to one, has no sign indicating positive or negative dependence, and indicates independence with a value of zero.[8]

Interestingly, Zhang (2024a) has proposed combining Chatterjee's and Spearman's measures as in (5) in an effort to obtain the best of both worlds: a dependence measure that has reasonable power under cases of non-monotonic, non-linear, and/or cyclical dependence (where Spearman's has little to no power, especially compared to Chatterjee's) as well as reasonable power under monotonic dependence (where Chatterjee's has less power than Spearman's).

(5) $$zcorrsp = I_{n\_sp}\left(X,Y\right) = \max\left\{\left|sr_{X,Y}\right|, \sqrt{5/2}\,\xi_n\left(X,Y\right)\right\}$$

Zhang's (2024a) combined correlation is signless, and ranges from zero to one, where zero indicates independence. This dependence measure also is directional/asymmetric due to its inclusion of Chatterjee's coefficient. Zhang (2024b) later derived the symmetric version of this test as (6):

(6) $$zcorrsp\_sym = \max\left\{\left|sr_{X,Y}\right|, \sqrt{5/2}\,\xi_n\left(X,Y\right), \sqrt{5/2}\,\xi_n\left(Y,X\right)\right\}$$

For completeness, Spearman's rho is defined below:

(7) $$sr_{X,Y} = \frac{\sum_{i=1}^{n}\left(R_{X_i} - \frac{1}{n}\sum_{j=1}^{n} R_{X_j}\right)\left(R_{Y_i} - \frac{1}{n}\sum_{j=1}^{n} R_{Y_j}\right)\Big/\left(n-1\right)}{\sqrt{\sum_{i=1}^{n}\left(R_{X_i} - \frac{1}{n}\sum_{j=1}^{n} R_{X_j}\right)^2\Big/\left(n-1\right)}\sqrt{\sum_{i=1}^{n}\left(R_{Y_i} - \frac{1}{n}\sum_{j=1}^{n} R_{Y_j}\right)^2\Big/\left(n-1\right)}}$$

where R indicates the rank ordering of the subscripted variable. In the event of ties in the ranks, widespread convention is to use the averaged rank (see Zar, 1999). If there are no ties in the data, (7) can be shortened to

(8) $sr_{X,Y} = 1 - \frac{6\sum_{i=1}^{n}\left(R_{X_i} - R_{Y_i}\right)^2}{n^3 - n}$, although this version also can be adjusted for ties as in (9):

[8] See Appendix A for code implementing ICH that efficiently increases speed by an order of magnitude.

(9) $$sr_{X,Y} = \frac{\left(\left[n^3-n\right]/6\right) - \sum_{i=1}^{n}\left(R_{X_i}-R_{Y_i}\right)^2 - \sum_{i=1}^{mx}\left(t_{X_i}^3-t_{X_i}\right)/12 - \sum_{i=1}^{my}\left(t_{Y_i}^3-t_{Y_i}\right)/12}{\sqrt{\left[\left(\left[n^3-n\right]/6\right)-2\sum_{i=1}^{m}\left(t_{X_i}^3-t_{X_i}\right)\right]\cdot\left[\left(\left[n^3-n\right]/6\right)-2\sum_{i=1}^{m}\left(t_{Y_i}^3-t_{Y_i}\right)\right]}}$$ where mx and my=# tied groups in x and y, respectively, and t_{X_i} and t_{Y_i} are the number of ties in each tied group.[9]

Another measure similar to Chatterjee's is the Differential Distance Correlation (DDC) of Liu and Shang (2025). DDC's values also are signless and range from zero to one, with zero indicating independence. Notably, similar to the widely used Szekely's distance correlation (see Szekely et al. (2007)), DDC can be multidimensional, but when X is univariate so that DDC can be used in an all-pairwise matrix, it is defined as (10) below:

(10) $$DDC_n\left(X\mid Y\right) = 1 - \frac{1}{(n-1)}\sum_{i=1}^{n-1}\left\|X_{(i)}-X_{(i+1)}\right\| \Big/ \left[\binom{n}{2}^{-1}\sum_{i=1}^{n}\left(2i-n-1\right)X^{(i)}\right]$$

where $\left\{\left(X_{(i)},Y_{(i)}\right)\right\}_{i=1}^{n}$ are ordered to satisfy $Y_{(i)}\le\cdots\le Y_{(n)}$, and $X^{(i)}$ are ordered to satisfy $X_{(i)}\le\cdots\le X_{(n)}$. Liu and Shang (2025) show in an empirical study that DDC has power similar to that of Chatterjee's measure, but with slightly more power under damped oscillator data. And like Chatterjee's measure, DDC is directional, so in the general case, $DDC_n\left(X\mid Y\right)\ne DDC_n\left(Y\mid X\right)$. Also like Chatterjee's measure, a symmetric, non-directional version may be obtained via (11)[10]:

(11) $$DDC_{n-sym}\left(X,Y\right) = \max\left[DDC_n\left(X\mid Y\right), DDC_n\left(Y\mid X\right)\right]$$

Finally, asymmetric, directional dependence measures also can be applied only to the tails of X and Y, and for financial settings, it is important to note that correlation breakdowns often are associated specifically with (asymmetric) tail dependence, as described by Pramanik (2024): "Extensive evidence has been gathered showcasing the prevalence of heavy-tailed distributions and asymmetric tail interdependence within equity and foreign exchange markets, particularly during times of crisis." And from Ito and Yoshiba (2025): "We provide new evidence that lower tail dependence coefficients increased compared to upper ones for all pairs in the COVID-19 crash…" One straightforward example of an asymmetric tail dependence measure is that of Deidda et al, (2023) which is essentially Kendall's tau applied conditionally, only when the percentile, *q*, of X (or Y) is exceeded:

---

[9] Averaging the ranks of ties in (7) is computationally preferred to (9) (see Zar (1999)).

[10] Based on email correspondence with author Yixiao Liu, July 9, 2025.

(12) $\hat{\tau}_{X,Y}(q) = \left(2(k-2)!/k!\right) \sum_{1 \le i \le j \le n} \text{sgn}\left(X_i - X_j\right) \text{sgn}\left(Y_i - Y_j\right) I\left(X_i, X_j > X_{(n-k)}\right)$

where $q = 1 - k/n$, and $k \le n$ is the number of exceedences used in the tail, and $I(\;)$ is the indicator function (one when true, zero otherwise) ensuring that only the *k* largest observations of X are used. Note again that generally, $\hat{\tau}_{X,Y}(q) \neq \hat{\tau}_{Y,X}(q)$, that is, this tail dependence measure is directional, and the effect of X's tail on Y's tail is not assumed to be the same as that of Y's tail on X's tail.

Other directional, asymmetric dependence measures include the dynamic asymmetric tail dependence measure of Ito and Yoshiba (2025), the QAD measure of Junker et al. (2021), the generalized correlation of Zheng et al. (2012), the extremal directional dependence measure of Garcin and Nicolas (2026), the multivariate directional tail-weighted dependence measure of Li and Joe (2024), and the measures of Vinod (2022) and others described in Jondeau (2016).[11]

2.B. Relationship between Measures of Association and Causality

This brief review of some of the more widely used and readily implemented directional dependence measures is relevant here because of their useability and utility within causal models, including ADD. But this utility holds not only for directional measures of association, but also for some of the oldest and most widely used non-directional measures of association, such as the covariance matrix. For example, Rodriguez Dominguez (2023, 2025a), together with subsequent work on causal PDE-control models for dynamic portfolio optimization (see Rodriguez Dominguez (2025c)), demonstrate that association-based structures, such as covariance and sensitivity measures, can be embedded within causal frameworks through structured representations of common drivers. Similarly, Cai et al. (2025) utilize the Cholesky factorization of the covariance matrix as the foundation of their causal algorithm, and this turns out to be a not uncommon approach (see also, for example, Li et al. (2025)). Additionally, Pascual-Marqui et al. (2024) ingeniously combine Chatterjee's and Szekely's measures to effectively perform directional, causal regressions. The extant literature contains many similar examples (see also Blömbaum et al. (2019) and MacKinnon & Lamp (2022)), making the point that some of the best applied causal modeling frameworks are those that in no small part intelligently utilize the most widely used and well established association-based dependence measures,[12] especially those that are asymmetric / directional.

---

[11] For completeness we cite these measures, but do not focus on their implementation in this setting as they are more computationally involved and less readily implemented compared to (1) through (6), (10), (11), and (12). On a related note, Metsämuuronen (2022) intriguingly demonstrates that under certain conditions, such as when categorical and ordinal data are being analyzed and the number of categories between the two variables differs dramatically, even Pearson's correlation can be unambiguously directional.

[12] Consistent with this theme, note also that the selection of causal drivers in Rodriguez Dominguez (2023, 2025a) follows the Reichenbach Common Cause Principle (Reichenbach, 1956), a foundational idea in probabilistic causality which emphasizes the identification of common causes through observed ***correlations***.

This should come as no surprise because, as noted above, while the development of directional dependence measures is not new, going back to the late nineteenth century (see Yule (1897)), neither is causal modeling, going back at least to Wright (1921). While gratuitous resistance to more recent and seminal work on causality (see Pearl (2000, 2009, and 2018)) undoubtedly plagued its introduction, we caution against the other end of the spectrum, that is, portraying causal modeling as brand new, separate from, and/or superior to association-based modeling. This can provide a convenient strawman for argumentative debates in the form of the tired mantra "correlation is not causation," but for rigorous, applied research it draws a false, bright line that seriously limits the advancement of causal modeling, precluding even from preliminary consideration the development of some of the most effective current (and future) approaches![13]

Making strides toward the effective application of causal frameworks, especially in quantitative finance, will be incremental and based in nontrivial ways on over a century of rigorous association-based research and methodology development, even in the face of paradigm shifts; claims to the contrary should be met with appropriate scientific skepticism. For a compelling empirical study refuting some of the overly strong promotional claims made regarding the application of causal models to, for example, quantitative finance, see Rodriguez Dominguez (2025b). This work convincingly demonstrates that while 'proof' of causality is desirable, it is not per se *necessary* for many objectives, such as portfolio optimization.

So rather than making overreaching and/or unsubstantiated claims regarding causality *versus* association-based methods, it is far more accurate and appropriate and *useful* to say that measures of association, both directional and non-directional, often serve as effective foundational supports for causal methods. Correlation may not be causation, and in multivariate settings the longtime challenge has been, and continues to be, avoiding the attribution of causality to relationships where only spurious association exists. However, correlation and other measures of association undoubtedly are often very strong *indicators* of causation, and the responsible use of association-based dependence measures elevates and underpins some of the best causal models. Ignoring over a century of rigorous development of association-based dependence measures when tackling difficult causal problems is, in fact, the limiting path, throwing the proverbial baby out with the bathwater. Causal modelers should not fear or eschew association-based measures of dependence, but rather, carefully and responsibly and rigorously embrace them. This is the path ADD takes to accurate and powerful causal discovery and DAG recovery, as described below.

[13] For example, one book promoting the use of causal models in quantitative finance mistakenly characterizes association-based dependence, which covers an enormous swath of all statistical work in the past century, as being strictly non-directional: "Third, unlike association, causality is directional" p.5, Lopiz de Prado (2023). Such oversights ignore the long-established, rich, and useful literature on (association-based) directional dependence measures described above, thus limiting innovation in applied causal research.

2.C. The Angles Space

To describe the development of ADD, we must first review in this section the angles space and the related literature, and then in the next section, its specific use with directional dependence measures, like ADD.

Reliance on spherical angles and hypersphere parameterizations is increasingly common in quantitative finance (see for some examples Golts & Jones, 2009; Li, Q., 2018; Hellton, 2020; Zhang, 2022; and Saxena et al., 2023), in large part due to its scale invariance. It has even been used to define entire financial markets (see Kim and Lee, 2016),[14] as well as in other settings to solve, with greater robustness and efficiency, complex multivariate problems (see Zhang and Yang (2025)). This more recent utilization is based on a long history, going back to the origin of modern statistics. The use of spherical angles for the analysis specifically of Pearson's correlation matrix goes back at least to Fisher (1915, 1928), but Joarder & Ali (1992) and Rapisarda et al. (2007) provide geometrically motivated, thorough, and clear descriptions of its derivations. The bivariate version of this is simply the widely known "cosine similarity" shown in (13) below, where Pearson's correlation is defined by the cosine of the angle between two mean-centered variables, which equals the inner product divided by the product of the two vectors' (Euclidean) norms.

(13) $$\cos\left(\hat{\theta}\right)=\frac{\text{inner product}}{\text{product of norms}}=\frac{\langle \mathbf{X},\mathbf{Y}\rangle}{\|\mathbf{X}\|\|\mathbf{Y}\|}=\frac{\sum_{i=1}^{n}\left(X_i-\frac{1}{n}\sum_{j=1}^{n}X_j\right)\left(Y_i-\frac{1}{n}\sum_{j=1}^{n}Y_j\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\frac{1}{n}\sum_{j=1}^{n}X_j\right)^2}\sqrt{\sum_{i=1}^{N}\left(X_i-\frac{1}{n}\sum_{j=1}^{n}Y_j\right)^2}}=\frac{\hat{Cov}_{X,Y}}{s_X s_Y}=r_{X,Y},\ 0\le\hat{\theta}\le\pi$$

It turns out that the matrix analogue for cosine similarity remains valid for any positive definite matrix, not just Pearson's rho, so it applies to any positive definite dependence measure, as is well established in the literature (see Joarder and Ali (1992), Pinheiro and Bates (1996), Rapisarda et al. (2007), Pouramadi and Wang (2015), and Cordoba et al. (2018)). Translation between the (positive definite) dependence measure matrix and the angles matrix is shown below, both formulaically in (14)-(16) and in SAS/IML computer code in Table A. This is, in fact, a multivariate bijection, so each (positive definite) dependence measure matrix corresponds uniquely, one-to-one, to a matrix of angles. The steps for translating between dependence measure values and angles, in both directions, are shown in A.-C. below.

A. estimate the correlation (dependence measure) matrix from sample data
B. obtain the Cholesky factorization of this matrix
C. apply inverse trigonometric and trigonometric functions to B. to obtain the corresponding matrix of spherical angles

and in reverse:

[14] Note that even for purposes of *estimating* dependence measure values, spherical angles can be one of the best approaches. As determined in Pinheiro and Bates (1996): "Of the five parameterizations considered here, the spherical parameterization presents the best combination of performance and statistical interpretability of individual parameters."

C. start with a matrix of spherical angles
B. apply trigonometric functions to C. to obtain the Cholesky factorization
A. multiply B. by its transpose to obtain the corresponding matrix of the dependence measure
(see Rapisarda et al. (2007), and Pourahmadi & Wang (2015), but note a typo in the formula in Pourahmadi & Wang (2015), for the first 3 steps)

Central to this translation mechanism is obtaining the Cholesky factor of the dependence measure matrix, which is usually a built-in function in most statistical and mathematical software. The relevant formulae are included below in (14) for completeness.

(14) A (dependence measure) matrix *R* will be real, symmetric positive-definite,[15] so the unique matrix *B* that satisfies $R = BB^T$ where *B* is a lower triangular matrix (with real and positive diagonal entries), and B^T is its transpose, is the Cholesky factorization of *R*. Formulaically, *B*'s entries are as follows:

$$B_{j,j} = (\pm)\sqrt{R_{j,j} - \sum_{k=1}^{j-1} B_{j,k}^2}\,, \qquad B_{i,j} = \frac{1}{B_{j,j}}\left(R_{i,j} - \sum_{k=1}^{j-1} B_{i,k}B_{j,k}\right) \text{ for } i > j$$

The Cholesky factor, which is only defined under positive definiteness, can be viewed as a matrix analog to the square root of a scalar, because similar to a square root, the product of it and its transpose yields the original matrix. Importantly, the Cholesky factor places us on the UNIT hyper-(hemi)sphere (where scale does not matter) because the sum of the squares of its rows always equals one. Next, we recursively apply inverse trigonometric and trigonometric functions to each cell of the Cholesky factor to obtain each cell's angle value per (15); or in reverse, we obtain a dependence matrix value from trigonometric functions applied to each cell's angle value per (16) (see both Joarder & Ali (1992) and Rapisarda et al. (2007), for meticulous derivations of these formulas). Again, note that this relationship is a multivariate bijection: a one-to-one, invertible, strictly monotone correspondence,[16] with a unique dependence measure matrix yielding a unique angles matrix, and vice versa.

(15) For *R*, a p x p correlation matrix, $R = \begin{bmatrix} 1 & r_{1,2} & r_{1,3} & \cdots & r_{1,p} \\ r_{2,1} & 1 & r_{2,3} & \cdots & r_{2,p} \\ r_{3,1} & r_{3,2} & 1 & \cdots & r_{3,p} \\ r_{4,1} & r_{4,2} & r_{4,3} & \cdots & r_{4,p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{p,1} & r_{p,2} & r_{p,3} & \cdots & 1 \end{bmatrix}$,

$R = BB^t$ where B is the Cholesky factor of R and

[15] Semi-positive definiteness includes the case of eigenvalues exactly equal to zero, which I largely ignore herein as a border case relevant mainly for textbook examples, since variables (whether financial returns, or factors) would have to exhibit exact linear dependence for an eigenvalue to be exactly zero.

[16] Note that angle values increase over [0, $\boldsymbol{\pi}$] as dependence measure values decrease over [-1, 1]; or for dependence measures with ranges of [0, 1], the angle values range only over [0, $\boldsymbol{\pi}/2$] and the relationship remains monotonic decreasing.

$$B=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \cos(\theta_{2,1}) & \sin(\theta_{2,1}) & 0 & \cdots & 0 \\ \cos(\theta_{3,1}) & \cos(\theta_{3,2})\sin(\theta_{3,1}) & \sin(\theta_{3,2})\sin(\theta_{3,1}) & \cdots & 0 \\ \cos(\theta_{4,1}) & \cos(\theta_{4,2})\sin(\theta_{4,1}) & \cos(\theta_{4,3})\sin(\theta_{4,2})\sin(\theta_{4,1}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \cos(\theta_{p,1}) & \cos(\theta_{p,2})\sin(\theta_{p,1}) & \cos(\theta_{p,3})\sin(\theta_{p,2})\sin(\theta_{p,1}) & \cdots & \prod_{k=1}^{n-1}\sin(\theta_{p,k}) \end{bmatrix} \text{ for } i>j \text{ angles } \theta_{i,j}\in(0,\pi).$$

To obtain an individual angle $\theta_{i,j}$, we have:

$$\text{For } i>1:\ \ \theta_{i,1} = \ \arccos(b_{i,1}) \text{ for } j=1;\ \text{ and } \theta_{i,j} = \ \arccos\left(b_{i,j} \Big/ \prod_{k=1}^{j-1}\sin(\theta_{i,k}) \right) \text{ for } j>1$$

(16) To obtain an individual correlation, $r_{i,j}$, we have, simply from $R=BB^T$:

$$r_{i,j}=\cos(\theta_{i,1})\cos(\theta_{j,1})+\sum_{k=2}^{i-1}\left[\cos(\theta_{i,k})\cos(\theta_{j,k})\prod_{l=1}^{k-1}\sin(\theta_{i,l})\sin(\theta_{j,l})\right]+\cos(\theta_{j,i})\prod_{l=1}^{i-1}\sin(\theta_{i,l})\sin(\theta_{j,l}) \text{ for } 1\le i<j\le n$$

As stated in Zhang et al. (2015), "Hence a model for the angles … is equivalent to a model for the correlation matrix." SAS/IML code translating dependence measure values to angle values and angle values to dependence measure values is shown in Table A below. For empirical examples of this translation, see Appendix A, which also includes graphs of the finite sample probability density functions (pdf's) of the angles under various data conditions, along with corresponding spectral distributions.

The above all is straightforward and well-established in the literature, and provides a universal[17] measure of dependence in the cell-level angles that possesses a critically important quality for this setting: namely, the angles are multivariate independent, by the nature of their orthogonal construction (see Pourahmadi and Wang (2016); Ghosh et al. (2021); Rapisarda et al. (2007); Tsay and Pourahmadi (2017); and Zhang et al. (2015)).[18] This allows a causal discovery model, like ADD, to assess the potential existence of each pairwise relationship represented in a DAG independent of all the others. And dual orderings, using directional dependence measures, will allow ADD to establish the ***direction*** of any such relationships.

Note that the cell-level pdf of each of the angles obviously will differ from those of the dependence measure, and the relationship between the two densities is nontrivial in the general case, requiring the determinant of the Jacobian matrix. Translating from the distribution of the angles (which has no

---

[17] Angles are 'universal' here in the sense that they can be calculated based on any positive definite dependence measure matrix.

[18] This independence is well established in the literature. Zhang et al. (2015) (supplementary material) and Rapisarda et al. (2007) use a geometric interpretation of the correlation matrix, based on (orthogonal) Givens rotations, to explain in detail the relationship between correlations and angles as well as why the angles distributions are multivariate independent.

TABLE A:

Correlations to Angles

```
* INPUT rand_R is a valid correlation matrix;

cholfact = T(root(rand_R, "NoError"));

rand_corr_angles = J(nrows,nrows,0);
  do j=1 to nrows;
    do i=j to nrows;
      if i=j then rand_corr_angles[i,j]=.;
      else do;
        cumprod_sin = 1;
        if j=1 then rand_corr_angles[i,j]=arcos(cholfact[i,j]);
        else do;
          do kk=1 to (j-1);
            cumprod_sin = cumprod_sin*sin(rand_corr_angles[i,kk]);
          end;
          rand_corr_angles[i,j]=arcos(cholfact[i,j]/cumprod_sin);
        end;
      end;
    end;
  end;

* OUTPUT rand_corr_angles is the corresponding matrix of angles;
```

SAS/IML code (v9.4)

Angles to Correlations

```
* INPUT rand_angles is a valid matrix of correlation angles;

Bs=J(nrows, nrows, 0);
do j=1 to nrows;
  do i=j to nrows;
    if j>1 then do;
      if i>j then do;
        sinprod=1;
        do gg=1 to (j-1);
          sinprod = sinprod*sin(rand_angles[i,gg]);
        end;
        Bs[i,j]=cos(rand_angles[i,j])*sinprod;
      end;
      else do;
        sinprod=1;
        do gg=1 to (i-1);
          sinprod = sinprod*sin(rand_angles[i,gg]);
        end;
        Bs[i,j]=sinprod;
      end;
    end;
    else do;
      if i>1 then Bs[i,j]=cos(rand_angles[i,j]);
      else Bs[i,j]=1;
    end;
  end;
end;
rand_R = Bs*T(Bs);

* OUTPUT rand_R is the corresponding correlation matrix;
```

universal, finite-sample analytical form in the extant literature)[19] to that of the dependence measure will lose multivariate independence between the cells, which is what we are exploiting in the angles to make causal discoveries. So the densities of the *angles* is what we focus on, making statistical inference in the angle space by using critical values of the empirical cumulative distribution function under independence, when conducting dual orderings.

3. Methodology

3.A. Rationale: Dual Orderings Using the Angles Space of Directional Dependence Measures

We now rely on the above descriptions of directional dependence measures and the (positive definite) angles space to outline the steps implementing ADD. We note, again, that ADD is designed specifically for causal discovery: it is not a method designed to provide effect sizes, as do regression-based

[19] Note that the fully analytic, finite sample distribution of the angles has been derived by Opdyke (2022, 2024a, and 2026) for the specific case of Pearson's rho under the Gaussian identity matrix. See www.DataMineit.com/DMI_publications.htm for an interactive, downloadable excel workbook implementing this.

approaches (see for example Pascual-Marqui et al. (2024); Blömbaum et al. (2019); and MacKinnon & Lamp (2022)), nor does it provide counterfactual outcomes. Rather, ADD identifies existing causal structure aimed at DAG recovery.

The approach here is to use, like some of the other causal frameworks listed above, the directional dependence measures described above in section 2. Take, for example, Chatterjee's improved correlation ("ICH") coefficient (see Xia et al. (2025)), which ADD can use twice via dual orderings, once with the data variables in one column order, and again with the data variables in the reverse column order. So if we have a treatment variable (X) and a dependent variable (Y) and relevant covariates (V1, V2, V3), we column-sort the data in one order (e.g. an ordering of X, V1, V2, V3, Y), and then column-sort the data in the reverse order (Y, V3, V2, V1, X), and apply ICH to each ordering; the two resulting ICH matrices will together capture all potential associations, in both directions, of all the variables/factors.[20] And all the cells of these two estimated dependence matrices will fully map to the relevant variable categories that make up a DAG (e.g. the confounders, colliders, mediators, independent variables, causes of X, consequences of X, causes of Y, and consequences of Y). For example, for a particular pair of factors, X and Y (i.e. the X-Y cell of the matrix, which is equivalent to an edge in a graph), two findings from this approach – one of a statistically significant, directional effect of X on Y (X→Y), but NO statistically significant, directional effect of Y on X (Y→X) – provide evidence of a directional effect of X on Y (X→Y). When assumptions A1-A7 are satisfied (see Appendix C), under conditioning such evidence is consistent with adjacency, and under marginal dependence they identify reachability. For readability purposes, I leave the formal definitions of A1-A7 in Appendix C,[21] and present the rationale for this approach here in the text as follows.

Consider a parent–child pair X→Y with $Y = f\left(X,\, U_Y\right),\ U_Y \perp X$, and NOT Y→X. Under additive-noise mechanisms, $Y = f\left(X\right) + U_Y$, independence between X and U_Y implies that the conditional variability of Y given X is dominated by U_Y, whereas the inverse mapping $X = g\left(Y\right) + \varepsilon$ generally violates independence (i.e. Y and the induced noise are statistically dependent). Independence between the cause and its exogenous noise term is the identifying restriction that generates directional asymmetry: only the true causal direction preserves independence, while the inverse mapping generally will violate it. This violation is precisely what ADD's directional measures exploit when evaluating the two possible directions of a pairwise relationship. In other words, because the one direction preserves independence between the input and the associated exogenous noise, while the reverse generally does not, the dual-order calculations allow the expected asymmetry to manifest directly in the empirical directional

[20] The same rationale applies to feature selection generally, which in quantitative finance often is referred to as feature engineering.

[21] Again, the major assumptions include acyclicity, causal sufficiency, Markov and faithfulness, independent and identically distributed sampling, reliable measurement, and monotone link from (directional) dependence measure to asymmetry.

measures. So we therefore have that directional functionals *M* that reward predictability under independent noise yield $E\left[M(X \to Y)\right] > E\left[M(Y \to X)\right]$.

Now for linear-Gaussian cases, this asymmetry collapses as both directions are equivalent. But for nonlinear and/or non-Gaussian and/or heteroskedastic mechanisms, the asymmetry is restored. And ADD leverages this by: (i) computing *M* in both directions (via dual orderings), (ii) performing joint finite-sample statistical tests in angle space under positive definite constraints, and (iii) calling an edge only when one direction is significant and the other direction is not. I implement these steps, using ICH, in the preliminary empirical study described below.

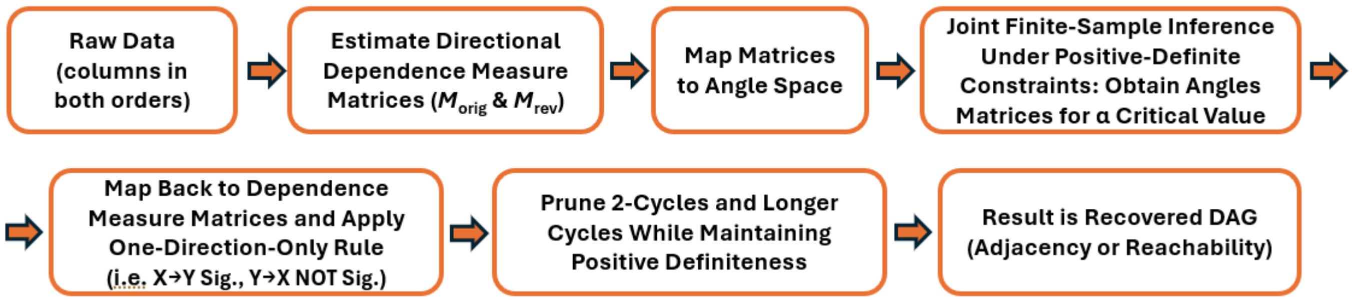
4. Preliminary Empirical Study

The objective of ADD and this study is to identify causal structure consistent with the 'ground truth' DAG, assuming it is rightly specified.[22] I demonstrate ADD's effectiveness for this purpose in the empirical study presented below, and importantly note again that, as explained in detail in Appendix C, under assumptions A1-A7, and when directional dependence is evaluated conditionally on DAG space V \ {X, Y}, the one-direction-only ADD rule is pointwise sound for adjacency; when directional dependence is evaluated marginally, as is done in this empirical study, ADD identifies directed reachability rather than direct parenthood. Promising results from this preliminary study will justify conditioning and more expansive testing in a larger, follow-up study described below. For the reader's reference during this step-by-step explanation, I place here in Graph 1 ADD's entire DAG recovery pipeline.

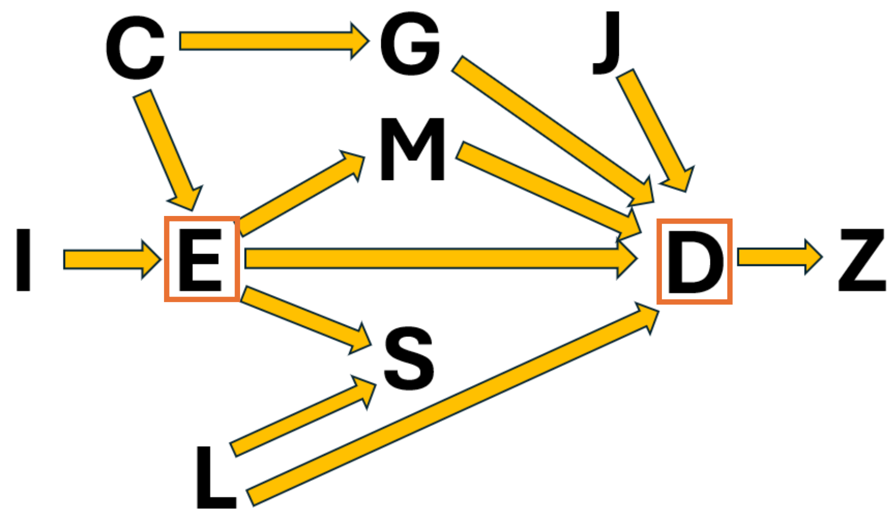
We start with the straightforward, but non-toy DAG presented in Digitale et al. (2022) (see Graph 2 below), which contains confounders, colliders, and mediators, as well as both causes and consequences of the two primary factors of interest, E and D. There are multiple causal paths linking E and D (E→D, E→M→D) as well as multiple non-causal paths linking E and D (E←C→G→D, E→S←L→D). Rightly identifying the directional relationships shown via the arrows in the DAG – and ONLY these relationships – will allow researchers to make unbiased estimation of the effect of E on D (by controlling for other effects on D, but only where appropriate). So accurate DAG recovery is the necessary precursor to this estimation. Note that some of the roles these variables play include C as a 'confounder,' M as a 'mediator,' S as a 'collider,' Z as an effect of D, I as an instrumental variable, i.e. a cause of E but not (directly) of D, J as an effect modifier, etc. For a complete and thorough description of the DAG I refer readers to the freely downloadable source, Digitale et al. (2022).

[22] MacKinnon and Lamp (2022) appropriately advise that "The correct causal model is an exacting qualification, requiring a program of research with precise definition of causal effects, specification of assumptions, and sensitivity analysis for how violating assumptions affects results. Statistical analysis is useful for demonstrating associations between variables that are consistent or inconsistent with a causal model."

Graph 1: DAG Recovery Pipeline



Graph 2: Directed Acyclic Graph of Digitale et al. (2022)



The dimension of the corresponding all-pairwise matrix, where each cell represents an edge of the graph, is p = 10 factors, making p(p-1)/2 = 45 pairwise cells. ADD's objective here is to rightly identify the directional effect in each of the 12 cells where one exists, and to rightly identify no effect in the 33 cells that have no effect (see Graph 3). I generate data under the null hypothesis of independence, to obtain the relevant (one-sided) confidence intervals, using multivariate uniform-distributed variables (only independence matters here; distributional changes make virtually no empirical difference in the distributions of the angles under the null, as expected). I generate data based on the alternate hypothesis (the DAG) via the code in Table C below.[23] All parametric distributions are heavy-tailed

[23] One of the goals here was to test the power to identify the individual effects of multiple parent variables, with different frequencies and functional forms, on a single child variable. The extensive use of coefficients in Table C allowed for such testing (e.g. the individual effects of J, E, M, G, and L on D) by preventing differences in scale from compounding by orders of magnitude. Performance under standardized scale is consistent with an absence of the marginal variance problem identified by Reisach et al. (2021). Unlike other causal discovery algorithms, ADD does not appear to have this problem, as we might expect due at least in part to the Cholesky factor placing angles on the UNIT hyper-hemisphere. However, this will be explicitly tested in a follow-up empirical study.

Graph 3: Directional Effects of 12 Pairwise Relationships
(Effect Direction is Column to Row, Original Order)

ORIGINAL ORDER

	I	C	E	G	L	S	M	J	D	Z
I										
C	X									
E	←	←								
G	X	←	X							
L	X	X	X	X						
S	X	X	←	X	←					
M	X	X	←	X	X	X				
J	X	X	X	X	X	X	X			
D	X	X	←	←	←	X	←	←		
Z	X	X	X	X	X	X	X	X	←	

REVERSE ORDER

	Z	D	J	M	S	L	G	E	C	I
Z										
D	→									
J	X	→								
M	X	→	X							
S	X	X	X	X						
L	X	→	X	X	→					
G	X	→	X	X	X	X				
E	X	→	X	→	→	X	X			
C	X	X	X	X	X	X	→	→		
I	X	X	X	X	X	X	X	→	X	

(student's t-distributed),[24] all error terms are asymmetric (lognormal distributed), and all functional relationships are highly nonlinear (most are sinusoidal, with some polynomial terms). As under the null, every simulation has n = 100 observations (which represents only about five months of trading days).[25] This yields, over 10,000 simulations of this DAG, the average Pearson's, Spearman's, and Kendall's

[24] Notably, angles-based dependence measures have been shown to be very robust to heavy-tailed data (see Zhang and Yang (2025)).

[25] Using reasonably small samples is conservative here, making it more difficult for ADD to rightly identify directional effects. Strong results under these conditions indicate good power of the estimator. Note that other studies treating DAG recovery use sample sizes as large as n=1,000 (see for example https://hub.crunchdao.com/competitions/causality-discovery), and even n=4100 (see Xue et al. (2025)).

TABLE C: SAS/IML Code Generating Data Under the Alternate Hypotheses (DAG)

```
call randgen(R_ERR_LOGN, "LOGNORMAL",0,sqrt(log( (1+sqrt(5))/2 )) );
R_ERR_LOGN = R_ERR_LOGN - exp( log( (1+sqrt(5))/2 )/2 );

call randgen(R_ERR_LOGN2, "LOGNORMAL",0,sqrt(log( (1+sqrt(5))/2 )) );
R_ERR_LOGN2 = R_ERR_LOGN2 - exp( log( (1+sqrt(5))/2 )/2 );

call randgen(R_ERR_LOGN3, "LOGNORMAL",0,sqrt(log( (1+sqrt(5))/2 )) );
R_ERR_LOGN3 = R_ERR_LOGN3 - exp( log( (1+sqrt(5))/2 )/2 );

call randgen(R_ERR_LOGN4, "LOGNORMAL",0,sqrt(log( (1+sqrt(5))/2 )) );
R_ERR_LOGN4 = R_ERR_LOGN4 - exp( log( (1+sqrt(5))/2 )/2 );

call randgen(R_ERR_LOGN5, "LOGNORMAL",0,sqrt(log( (1+sqrt(5))/2 )) );
R_ERR_LOGN5 = R_ERR_LOGN5 - exp( log( (1+sqrt(5))/2 )/2 );

call randgen(R_ERR_LOGN6, "LOGNORMAL",0,sqrt(log( (1+sqrt(5))/2 )) );
R_ERR_LOGN6 = R_ERR_LOGN6 - exp( log( (1+sqrt(5))/2 )/2 );

call randgen(CCC, "T", 10);

call randgen(III, "T",10);

EEE = cos(2*constant('PI')*CCC) + cos(2*constant('PI')*III) + R_ERR_LOGN/8;

GGG = -sin(2*constant('PI')*0.95*CCC) + R_ERR_LOGN2;

call randgen(LLL, "T",10);

SSS = -cos(2*constant('PI')*LLL) - sin(2*constant('PI')*(EEE)) + R_ERR_LOGN3;

MMM = -0.7*cos(2*constant('PI')*(0.95*EEE)) + R_ERR_LOGN4/3;
call randgen(JJJ, "T",10);

DDD = -2.3*cos(2*constant('PI')*(0.6*JJJ)) + 1.35*cos(2*constant('PI')*(0.25*EEE)) - 3.8*MMM##2 +
2.1*cos(2*constant('PI')*(0.65*GGG)) + 2.3*sin(2*constant('PI')*LLL) + R_ERR_LOGN5/8;

ZZZ = cos(2*constant('PI')*(DDD)) + R_ERR_LOGN6/6;
```

TABLE D: Average Pearson's Rho, Spearman's Rho, and Kendall's Tau Matrices
10,000 Simulations Under the Alternate Hypothesis of the DAG

Pearson's Rho

	I	C	E	G	L	S	M	J	D	Z
I	1									
C	0.001	1								
E	0.001	0.000	1							
G	-0.001	0.001	-0.001	1						
L	-0.001	0.001	0.001	-0.001	1					
S	0.000	0.000	0.050	0.000	-0.001	1				
M	0.001	0.000	0.009	-0.002	0.000	-0.001	1			
J	0.001	-0.001	0.000	0.000	0.000	0.000	0.000	1		
D	-0.001	0.000	0.001	-0.041	0.000	0.001	-0.168	-0.001	1	
Z	0.000	0.001	0.001	-0.001	-0.001	0.001	0.000	-0.001	0.000	1

Spearman's Rho

	I	C	E	G	L	S	M	J	D	Z
I	1									
C	0.001	1								
E	0.001	0.000	1							
G	0.000	0.001	-0.001	1						
L	-0.001	0.001	0.001	-0.001	1					
S	0.000	0.000	0.036	0.000	-0.001	1				
M	0.001	0.000	0.008	-0.001	0.001	-0.001	1			
J	0.001	0.000	0.000	0.000	0.000	0.001	0.000	1		
D	0.000	0.000	0.002	-0.053	-0.001	0.001	-0.039	-0.002	1	
Z	0.000	0.001	0.000	-0.002	-0.001	0.000	0.000	0.000	0.000	1

Kendall's Tau

	I	C	E	G	L	S	M	J	D	Z
I	1									
C	0.000	1								
E	0.000	0.000	1							
G	0.000	0.000	-0.001	1						
L	0.000	0.001	0.000	-0.001	1					
S	0.000	0.000	0.024	0.000	-0.001	1				
M	0.001	0.000	0.005	-0.001	0.000	-0.001	1			
J	0.001	0.000	0.000	0.000	0.000	0.001	0.000	1		
D	0.000	0.000	0.001	-0.036	-0.001	0.001	-0.026	-0.002	1	
Z	0.000	0.001	0.000	-0.001	-0.001	0.000	0.000	0.000	0.000	1

matrices shown in Table D. These show essentially no non-directional associations.[26] I choose a directional dependence measure for ADD: the improved Chatterjee correlation coefficient ("ICH") of Xia

[26] One of the cells shows a slight effect, but only under Pearson's (which is more sensitive to outliers), and this is almost certainly due to the different functional form for that factor, which is one (a quadratic term) that is more sensitive to bi-directional effects (when using ICH) than those characterizing most of the DAG (i.e. sinusoidal).

et al. (2025).[27] Because this is a directional measure whose values range from zero (independence) to one (perfect dependence), I need specifically the upper ($1 - \alpha$) confidence bound[28] under data generated under independence, to test whether one can reject the null hypothesis of independence when the data are generated under the alternate hypothesis (the DAG).[29] We set $\alpha = 0.05$, but conservatively adjust it using a Bonferroni adjustment ($\alpha\_adj = \alpha/2$) because I am conducting two tests on the same data for each edge / cell of the matrix: one with the columns of factor data in the "original" order, and one with the columns of factor data in the reverse order (see Graph 3; the order of the columns in the data match the column order(s) in the matrices).[30]

For each cell, there are four possible outcomes of the two tests:
A. significant effect in the original direction, no significant effect in the reverse direction
B. no significant effect in the original direction, significant effect in the reverse direction
C. no significant effect in either direction
D. significant effects in both directions

Under independence, the probabilities associated with A.-D. then become:
A. Pr(A.) = $\alpha\_adj * (1 - \alpha\_adj) = 0.025*0.975 = 0.024375$
B. Pr(B.) = $\alpha\_adj * (1 - \alpha\_adj) = 0.025*0.975 = 0.024375$
C. Pr(C.) = $(1 - \alpha\_adj) * (1 - \alpha\_adj) = 0.975*0.975 = 0.950625$
D. Pr(D.) = $(\alpha\_adj) * (\alpha\_adj) = 0.025*0.025 = 0.000625$

[27] ICH is noted in Xia et al. (2025) to have good power under these data conditions, so it is an appropriate choice for a preliminary study whose design is meant to test and justify further inquiry (we propose below an extensive follow-up study under wide-ranging data conditions). But this study poses nontrivial challenges to testing the efficacy of ADD in other ways, such as using much smaller sample sizes compared to other studies cited herein. The choice of dependence measure also brings up a potentially large advantage of ADD over competitors, that is, its ability to opportunistically use any (positive definite) directional dependence measure as befits the data conditions of different settings. This flexibility can provide increases in statistical power that other methods based on fixed measures will lack. In fact, the data conditions of this study are more consistent with a setting we want to focus on – quantitative finance – than others tested in the literature (e.g. circular data), so it is reasonable, and a benefit of ADD, to be able to select an appropriate dependence measure that acknowledges and exploits the type of data we are faced with in a given setting.

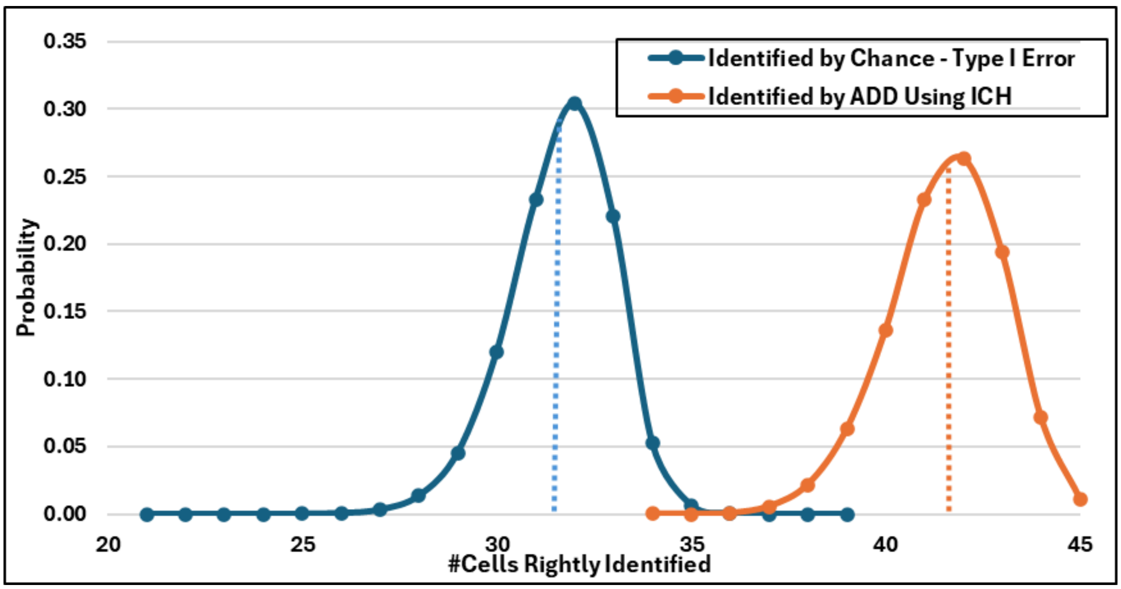
[28] The upper confidence bound is ($1 - \alpha$) rather than ($1 - \alpha/2$) because this is a one-sided hypothesis test.

[29] Recall that ADD's inference is based on the angles corresponding to each cell of the matrix, so this is actually the lower α bound because of the monotonic decreasing relationship between angles and dependence measure values: angles approach zero as dependence measure values approach one (or perfect dependence), and angles approach $\pi/2$ as (directional) dependence measure values approach zero (or independence). For non-directional measures, like Pearson's or Kendall's or Spearman's, angles approach π as dependence measure values approach negative one, or rather, the lower bound of the matrix, as negative one is a maximal lower bound that is not always attainable.

[30] Regardless of the relationship between these two tests (they are unlikely to be independent), the conservativeness of the Bonferroni adjustment, along with the small number of tests per cell (only two), mitigates any concern about inflated false positives due to multiplicity (see Bonferroni, 1936). The effects of multiplicity across cells (for which the tests are independent, by design, because the angles are independent random variables, as described in earlier sections) are ignored for purposes of this exercise.

The 2-cycles from D. are immediately pruned to preserve the acyclicality of the DAG by retaining only the stronger of the two effects (i.e. the one with the smaller p-value). Next, longer cycles are pruned based on several criteria including the number of called edges (cells) per node (factor/variable), the sum of the p-values associated with all the nodes with the same number of edges, and within each of these groups, the individual p-values of the specific edges (see Appendix C for details). All pruning is performed under the constraint of positive definiteness. Once acyclicality is enforced, what we are interested in is the random variable that is the number of the 45 pairwise cells, which are edges of the DAG, that are rightly identified, where rightly identified means the 12 cells with directional relationships are rightly chosen as outcome A., and those 33 cells with no directional relationships are rightly chosen as outcome C. Under the null hypothesis of independence, this distribution is shown in Graph 4 below, and the expected value is approximately 31.66 right identifications.[31]

Graph 4: Distribution of #Rightly Identified Directional Effects/Non-Effects: All 45 Pairs (Fully Exploratory Analysis, No Priors)



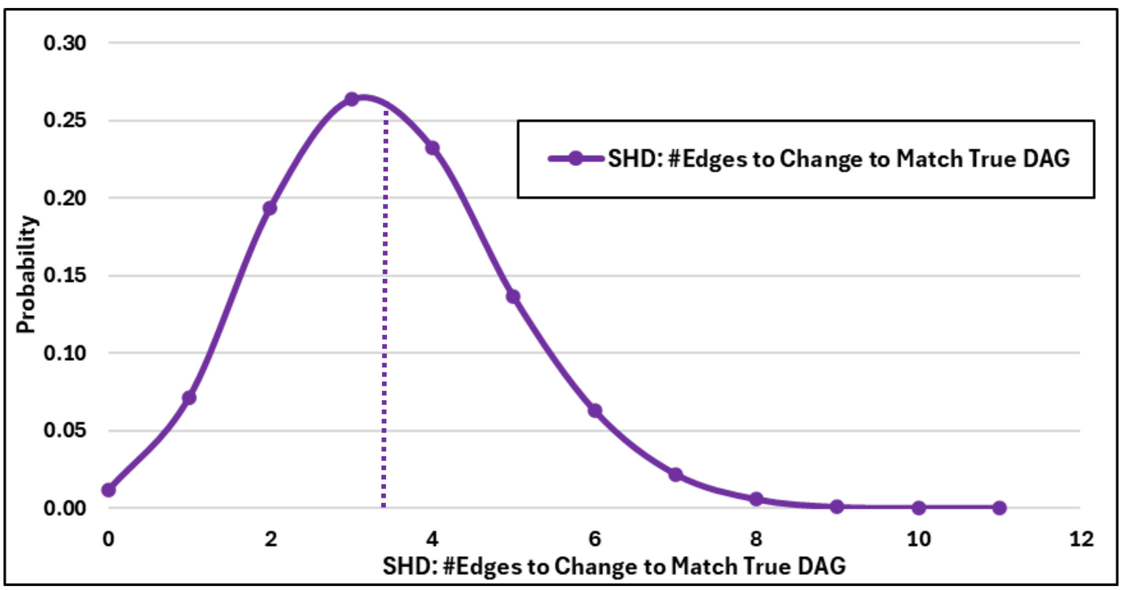
ADD is applied as described above, and shown in the two matrices in Graph 3. The columns/factors of data first are sorted in order so that the 12 relationship cells all are in the 'right' direction according to the DAG. Graph 3 in the Original Order shows the directional effect going from column to row (so, for example, the cell with row E and column I indicates the effect of factor I on factor E). ICH is applied to the data with the factors in this order, and the resulting matrix of ICH values is saved. Then the column order is reversed, ICH is again applied, the resulting matrix of ICH values is saved, and then both the rows and columns of this second matrix are reversed so that the two matrices can be compared easily, cell for cell. If, for a given cell, the result of the first test is statistically significant, but the result of the second test is

[31] I obtain this distribution via a simple simulation that applies random uniform variables generated with support [0, 1] to rightly identify each of the 12 directional relationships with Pr(A.) and each of the 33 non-relationships with Pr(C.) yielding, after 100 million simulations, Graph 4, with an expected value of approximately 31.66 right identifications.

NOT statistically significant, then the result is classified as outcome A.: a directional effect in the right direction. For the 33 'non-relationship' cells, a 'right' outcome is a finding of no statistically significant effect in either direction, i.e. outcome C. This is repeated 10,000 times, and for each simulation, acyclicality is enforced, and then the number of times each cell is rightly classified, either as one with a directional effect or one with no effect, is summed and shown in Graph 4 above. The expected value of this distribution is 41.56, compared to that under the null which is 31.66.[32] The difference between the two distributions, which unarguably is material, also is highly statistically significant, with a p-value<0.00001.[33] Note that this #right metric simply is the inverse of the Structural Hamming Distance (SHD), that is, the number of edge modifications that need to be changed for an estimated DAG to match the true DAG. The corresponding SHDs across the 10,000 simulations, with a mean of 3.44, are shown in Graph 4a. I combine Graphs 4 and 4a in Graph 4b to put them on the same scale.

In addition to SHD, I present the confusion matrix, otherwise known as the error matrix, below in Table E, along with various common model performance metrics based on it. These values all indicate strong predictive power of the ADD model under these conditions.

Graph 4a: Distribution of SHD: #Edges Changed to Match True DAG



[32] Note that acyclicality is not enforced for the distribution under the null, making this a conservative comparison. In other words, the null distribution is likely to be slightly smaller, on average, if edges are pruned when cycles occasionally are formed at random. This is true of the null distributions shown in Graphs 4b, 5, and 6 as well.

[33] This is a one-sided p-value based on a permutation test of 100,000 samples with sample sizes of n1=n2=45 (see Opdyke (2003, 2011, and 2013) for details on permutation tests). These sample sizes are used because they reflect the outcomes of 45 Bernoulli variables, that is, for 45 cells/edges, 45 outcomes of zero or one that sum to the #right in a given sample.

Graph 4b: Distribution of #Rightly Identified Directional Cells/Edges vs. SHD

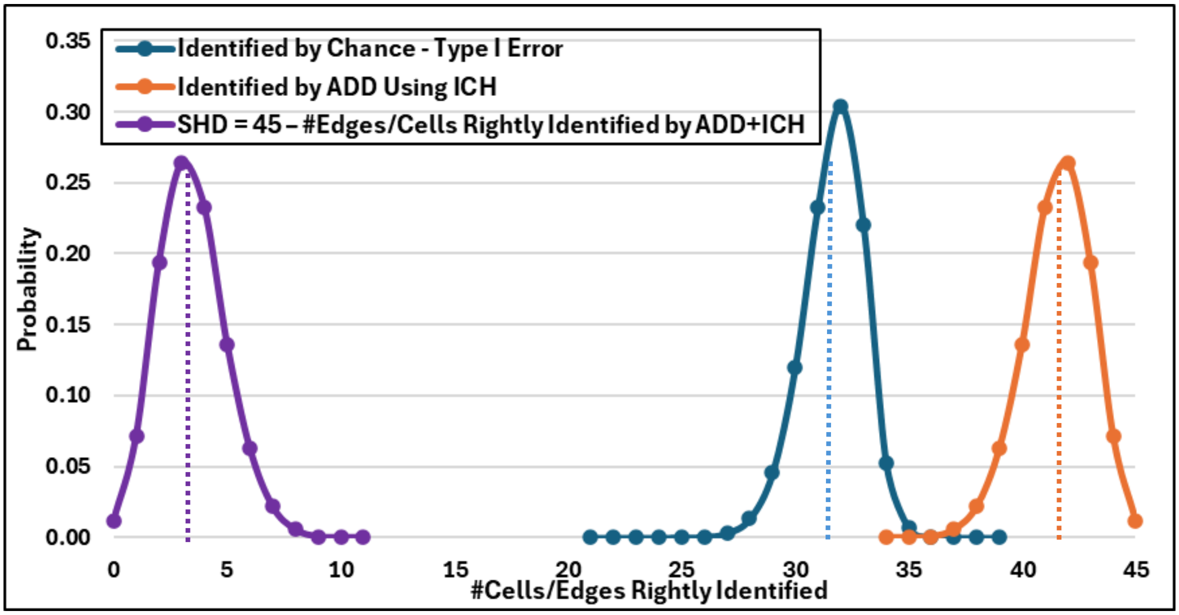


TABLE E: Error Matrix, and Associated Performance Metrics

Error Matrix - Numbers

	Predicted: Right	Predicted: Not Right	TOTAL
Right Direction	94,694	25,306	120,000
Not Right Direction	9,127	320,873	330,000

Error Matrix - Rates

	Predicted: Right	Predicted: Not Right	TOTAL
Right Direction	78.9%	21.1%	100%
Not Right Direction	2.8%	97.2%	100%

Precision =	0.91	MCC =	0.80	Jaccard =	0.73
Recall =	0.79	F1 =	0.85	Prevalence Threshold=	0.16
Accuracy =	0.92	Youden =	0.76	Balanced Accuracy =	0.88

I also present below in Table F the percentage of right classifications, across all simulations, for each specific cell/edge. Cells with red font indicate lower relative scores on the %right metric. The red on the effect of M on D is almost certainly due to the functional form being a quadratic term as opposed to a sinusoidal term, where the former is more sensitive to bi-directional results (when using ICH) and thus, yields fewer rightly identified one-direction effects; also, the effect of E on D is (relatively) muted due to the large number of parent nodes to D. This type of table can be very informative regarding the specific strengths and weaknesses of an estimator and its framework under different DAGs and different data conditions.

If we examine only the 12 cells with directional relationships, the results remain strong. Under the null hypothesis of independence, the distribution of the number of right identifications becomes simply a binomial distribution with probability = Pr(A.), because there are no cells with 'no relationship' included. This is shown in Graph 5, compared to ADD's distribution of the number of right identifications for these 12 cells. The expected value of the former is approximately 0.29, and of the latter, 9.47. This very material difference also is highly statistically significant, even under smaller sample sizes, with a p- value < 0.00001.[34] ADD's %right value is 78.9% here (note that this is just recall, the percent of "positives" rightly identified), while that under the null is about 2.4%, which is close to Pr(A.), as expected.

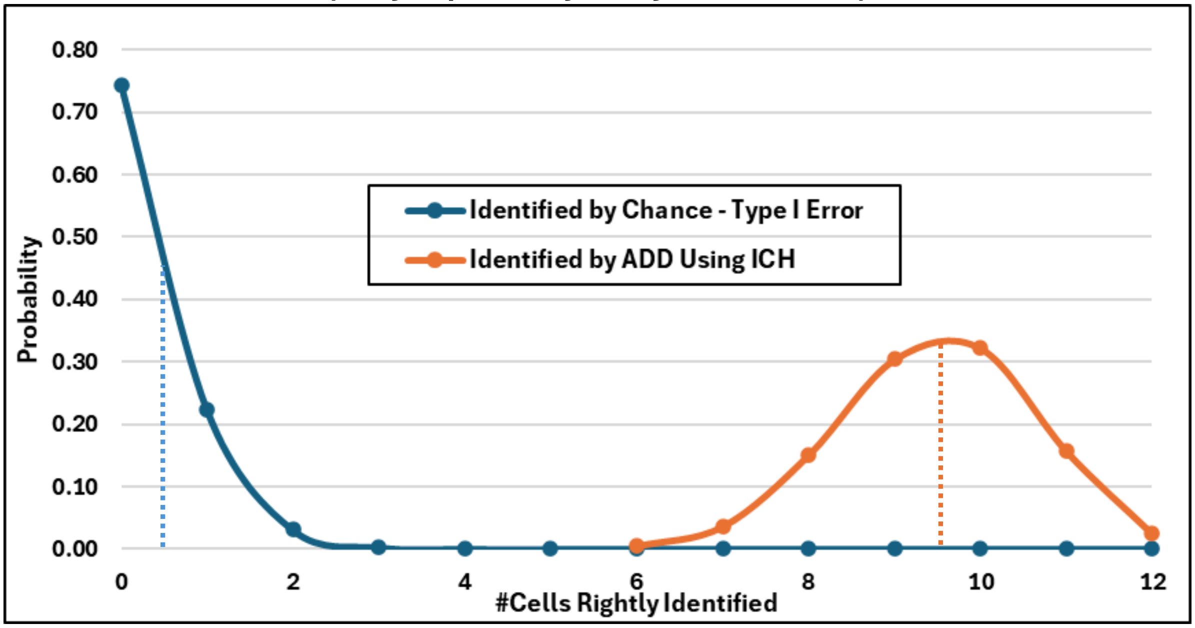
TABLE F: %Rightly Identified in Each Cell
A. 12 Cells with Directional Relationships, B. 33 Cells with No Relationships

	I	C	E	G	L	S	M	J	D	Z
I										
C										
E	99.6%	99.5%								
G		99.8%								
L										
S			82.7%		76.9%					
M			100.0%							
J										
D			34.7%	68.9%	59.5%		51.4%	75.4%		
Z									98.5%	

	I	C	E	G	L	S	M	J	D	Z
I										
C	95.1%									
E										
G	96.2%		97.6%							
L	95.3%	95.8%	96.3%	97.0%						
S	97.1%	97.8%		97.4%						
M	97.7%	97.7%		96.6%	97.4%	98.3%				
J	95.6%	96.2%	96.8%	97.3%	97.5%	98.1%	97.9%			
D	97.2%	97.1%				98.0%				
Z	96.6%	97.6%	97.8%	98.1%	98.0%	98.0%	98.6%	99.0%		

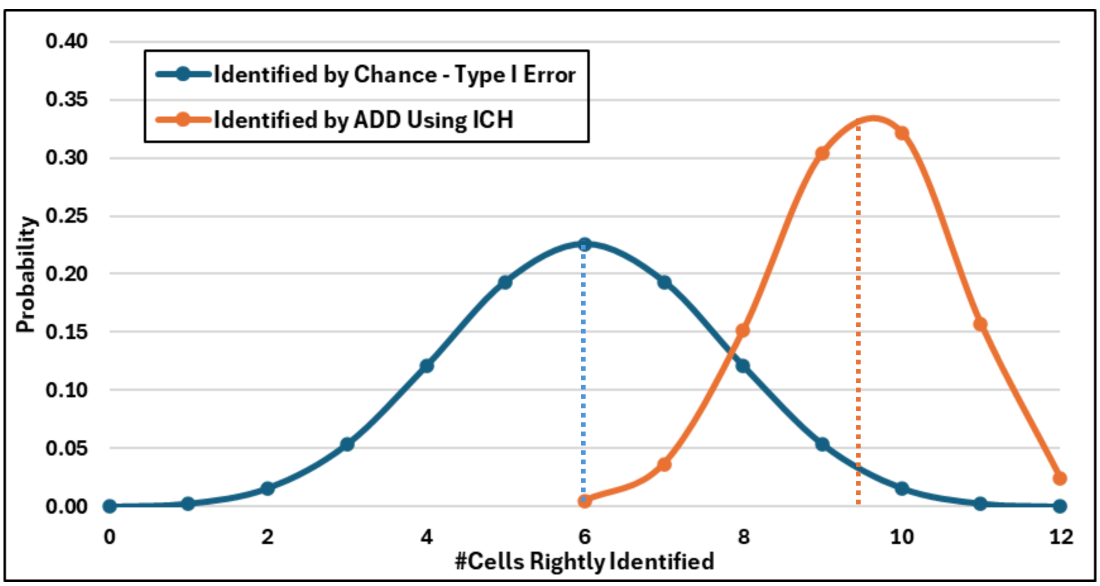
[34] For this permutation test, sample sizes n1 = n2 = 12, and 100,000 simulations were run.

Graph 5: Distribution of #Rightly Identified Directional Effects: 12 Pairs with Effects (Fully Exploratory Analysis, No Priors)



Even if we gave a researcher an unfair advantage over ADD and he or she somehow KNEW that the 12 cells were associated with an effect (i.e. perfect priors), without knowing the direction of the effect, the distribution under the null would be a coin toss, with an expected value for the number of right selections of 6 out of the 12. Even under these conditions (see Graph 6), ADD's effect is material, and highly statistically significant, with a p-value of 0.00068.[35] All of these results are summarized below in Table G.

Graph 6: Distribution of #Rightly Identified Directional Effects: 12 Pairs with Effects (Perfect Priors on which 12 Pairs Have SOME Effect, but not the Direction)



[35] For this permutation test, sample sizes n1 = n2 = 12, and 100,000 simulations were run.

This is a preliminary and relatively narrow empirical study, but its results are promising for ADD's general application to DAG recovery for causal modeling. While ADD is not designed to provide effect sizes or counterfactual outcomes, it appears to be able to effectively identify causal structure by recovering directional structure, consistent with the DAG in the sense of reachability, with reasonable power. First, the 78.9%right metric value (which is the recall for the subgroup of cells with an actual directional effect) compares very favorably with outcomes of similar recent studies.[36] Also, the sample sizes used herein arguably are not large, thus providing a reasonable challenge to ADD vis-à-vis its statistical power.[37]

TABLE G: ADD's DAG Recovery vs. No Model

All 45 Pairs, Fully Exploratory Analysis, No Priors

no model-type I error*		ADD-ICH		
# right	% right	# right	% right	p-value*
31.66	**70.4%**	**41.56**	**92.3%**	**<0.00001**

* Binomial vector: p12=adj_α*(1-adj_α), p33=p=[1 - adj_α]^2

* p-value: permutation test, nsims=100k, n1=n2=45

12 Pairs with Effects, Fully Exploratory Analysis, No Priors

no model-type I error**		ADD-ICH		
# right	% right	# right	% right	p-value**
0.29	**2.4%**	**9.47**	**78.9%**	**<0.00001**

** Binomial distribution, p12=adj_α*(1-adj_α)

** p-value: permutation test, nsims=100k, n1=n2=12

12 Pairs with Effects, Perfect Priors on Right Pairs but not Direction of Effect

no model-type I error~		ADD-ICH		
# right	% right	# right	% right	p-value~
6.00	**50.0%**	**9.47**	**78.9%**	**0.00068**

~ Binomial distribution, p12=0.5

~p-value: permutation test, nsims=100k, n1=n2=12

33 Pairs with No Effect, Fully Exploratory Analysis, No Priors

no model-type I error~~		ADD-ICH	
# right	% right	# right	% right
31.37	**95.1%**	**32.09**	**97.2%**

~~ Binomial distribution, p33=[1 - adj_α]^2

[36] See the recall of 76.7% for the winning DAG recovery method in an online contest with over 2,000 participants and 5,000 submissions (see https://crunchdao.com/case-studies/adia-lab-causality).

[37] The above-noted contest used sample sizes (n=1,000) ten times as large as those used herein (see https://hub.crunchdao.com/competitions/causality-discovery). Sample sizes in other studies (see Xue et al. (2025)) are larger still (n=4100), at forty-one times as large as our study.

Importantly, ADD maintains the flexibility to use any (positive definite) directional dependence measure for this problem of DAG recovery, as befits the data conditions of a particular setting, which in itself gives it advantages over many competitors. While ADD does inherit the strengths and weaknesses of the (directional) dependence measure used, and these will vary notably based on the data types being evaluated, careful scrutiny of the data can turn this potential drawback into a strong comparative advantage, permitting the opportunistic selection of dependence measures to increase power over less flexible causal algorithms. This also arguably makes ADD more robust in that it remains useable and applicable in the face of notably changing conditions (as long as those changes are measurable). Additionally, ADD's simultaneous estimation of all pairwise relationships, which properly and automatically enforces positive definiteness, may provide greater coherence in its edge calls compared to other causal algorithms, especially those that call edges sequentially. This has been a serious obstacle plaguing many of the widely used DAG recovery algorithms to date (see Hulse et al. (2025) and Faltenbacher et al. (2026)), so this is a nontrivial potential advantage of ADD. Finally, the fact that ADD requires only two matrix estimations, and two fast simulations to establish confidence intervals under independence, means that it scales extremely efficiently with the size of the DAG. All of these actual and potential advantages will be thoroughly explored and tested in a more extensive, follow-up empirical study that compares ADD to competitor algorithms. The study will include different sample sizes, different data types (e.g. both stationary and non-stationary), varying signal-to-noise ratios, different directional dependence measures, different DAGs, and different pruning algorithms, so it will be quite comprehensive. But first implementing this more narrow, yet reasonably challenging study was an efficient, useful, and large first step on this path of empirical testing and validation.

5. Conclusions and Next Steps

In ADD, we develop an original approach to causal discovery via DAG recovery: ADD simultaneously identifies directional dependence using dual orderings in the angles space of directional dependence measures. Our preliminary empirical study provides promising results in terms of accurate and powerful DAG recovery (in the sense of reachability), under challenging data conditions similar to those encountered in quantitative finance settings (i.e. nonlinearity and heavy-tailedness). This indicates ADD's potential utility for feature selection in this setting. Additional potential advantages include increased coherence, due to simultaneous edge-calling within a positive definite space; increased power, due to the flexible capability to use any directional dependence measure and thus, opportunistically adapt to different or varying data conditions; and increased speed/scalability, due to the need for only two matrix estimations, and two fast simulations to define empirical confidence bounds under independence, regardless of the size of the DAG space. A more extensive follow-up study will test these by directly benchmarking performance against competing causal discovery algorithms.

We end on a somewhat cautionary note, however, recognizing that this dive into causality via DAG recovery all begs the bigger and valid question of whether DAGs can be used reliably within "self-referencing open systems like capital markets" to begin with (Polakow et al., 2023). This applies to other settings as well, because DAGs have strong limitations generally, as cited herein, cautioned by Dawid (2009) and MacKinnon and Lamp (2021), and demonstrated in recent research by Hulse et al. (2025), Padh et al. (2025), Kaiser and Sipos (2022), Faltenbacher et al. (2026), Gong et al. (2024), and De Lara (2024).[38] But progress continues to be made here. Specifically, Reisach et al. (2026) persuasively contend that a major limitation of "nontemporal" causal DAGs is that they fail to explicitly incorporate the notion of time into the DAG, thus arguably obstructing justification of the acyclicity assumption; they address this by creating *composite* causal variables that refer to quantities at one or multiple time points. Similarly, the work of Cüppers et al. (2024) explicitly enable dynamic delays in their causal framework using a Minimum Description Length to add edges in their discovery algorithm in topological order. And the recent work of Rodriguez Dominguez (2023, 2025a, 2025c) addresses the challenging issue of temporality by embedding association-based dependence measures directly within time-dependent causal frameworks. With a focus on this latter work, we will explore integration between ADD and these methodologies to directly address temporal issues in the follow-up study.

6. References

Aliferis, C., Statnikov, A., Tsamardinos, I., Mani, S., and Koutsoukos, X., (2010), "Local Causal and Markov Blanket Induction for Causal Discovery and Feature Selection for Classification Part I: Algorithms and Empirical Evaluation," 11(7):171–234.

Allena, D., and McAleerb, M., (2018), "'Generalized Measures of Correlation for Asymmetry, Nonlinearity, and Beyond': Comment," ICAE – Instituto Complutense de Análisis Económico, Working Paper 1823.

Angrist, J., Imbens, G., and Rubin, D., (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91(434), 444-455.

Blömbaum, P., Janzing, D., Washio, T., Shimizu, S., and Schölkopf, B., (2019), "Analysis of Cause-Effect Inference by Comparing Regression Errors," *Peer Journal of Computational Science*, 5:e169. doi: 10.7717/peerj-cs.169.

Bonferroni, C. E., (1936), "Teoria statistica delle classi e calcolo delle probabilità," *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*.

Cai, Y., Li, X., Sun, M., and Li, P., (2025), "Recovering Linear Causal Models with Latent Variables via Cholesky Factorization of Covariance Matrix," In: Le Thi, H.A., Le, H.M., Nguyen, Q.T. (eds) Advances in Data Science and Optimization of Complex Systems, ICAMCS 2024, Lecture Notes in Networks and Systems, Vol. 1311, Springer, Cham.

[38] From Polakow et al. (2023): "The clarion call for causal reduction in the study of capital markets is intensifying. However, in self-referencing and open systems such as capital markets, the idea of unidirectional causation (if applicable) may be limiting at best, and unstable or fallacious at worst." From Gong et al. (2024): "… potential outcomes (PO) and structural causal models (SCMs) stand as the predominant frameworks. However, these frameworks face notable challenges in practically modeling counterfactuals … we identify an inherent model capacity limitation, termed as the 'degenerative counterfactual problem', emerging from the consistency rule that is the cornerstone of both frameworks." And from De Lara (2024): "Most of the literature on causality considers the structural framework of Pearl and the potential-outcomes framework of Neyman and Rubin to be formally equivalent, and therefore interchangeably uses the do-notation and the potential-outcome subscript notation to write counterfactual outcomes. In this paper, we … prove that structural counterfactual outcomes and potential outcomes do not coincide in general – not even in law." See Opdyke (2024b) for a more complete review of this literature.

Chatterjee, S., (2021), “A New Coefficient of Correlation,” *Journal of the American Statistical Association*, Vol 116(536), 2009-2022.
Chen, Zhongli, Yang, Y., Li, C., Yang, J., Li, M., Chen, Zhiwen, Wang, M., and Gui, X., (2026), “A constrain–select Markov blanket discovery algorithm for causal feature selection,” *Neurocomputing*, 683, 133517.
Chickering, D., (2002), “Optimal Structure Identification With Greedy Search,” Journal of Machine Learning Research, 3, 507-554.
Cordoba, I., Varando, G., Bielza, C., and Larranaga, P., (2018), “A fast Metropolis-Hastings method for generating random correlation matrices,” *IDEAL*, pp. 117-124, part of Lec Notes in Comp Sci., Vol 11314.
Cüppers, J., Xu, S., Musa, A., and Vreeken, J., (2024), “Causal Discovery from Event Sequences by Local Cause-Effect Attribution,” *38th Conference on Neural Information Processing Systems* (NeurIPS’24).
Czado, C., (2025), “Vine Copula Based Structural Equation Models,” *Computational Statistics and Data Analysis*, pp.453-477.
Dawid, A., (2009), “Beware of the DAG!,” *JMLR: Workshop and Conference Proceedings*, 6: 59-86.
De Lara, L., (2023), “On the (in)compatibility between potential outcomes and structural causal models and its signification in counterfactual inference,” arXiv:2309.05997v3 [math.ST].
Deidda, C., Engelke, S., and De Michele, C., (2023), “Asymmetric Dependence in Hydrological Extremes,” *Water Resources Research*, Vol. 59, Issue 12.
Digitale, J., Martin, J., and Glymour, M., (2023), “Tutorial on Directed Acyclic Graphs,” *Journal of Clinical Epidemiology*, 142, 264-267.
Embrechts, P., Hofert, M., and Wang, R., (2016), “Bernoulli and Tail-Dependence Compatibility,” *The Annals of Applied Probability*, Vol. 26(3), 1636-1658.
Faltenbacher, S., Wahl, J., Herman, R., and Runge, J., (2026), “How PC-based Methods Err: Towards Better Reporting of Assumption Violations and Small Sample Errors,” arXiv:2502.14719v2 [stat.ML].
Fisher, R. A., (1915), “Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population,” *Biometrika*, 10, 507-521.
Fisher, R. A., (1928), “The General Sampling Distribution of the Multiple Correlation Coefficient,” Rothamsted Experimental Station, Harpenden, Herts.
Gao, M., Li, Q., (2024), “A Family of Chatterjee’s Correlation Coefficients and Their Properties,” arXiv:2403.17670v1 [stat.ME]
Garcin, M., and Nicolas, M., (2026), “Directional Dependence of Extreme Events,” working paper, arXiv:2604.03215v1 [stat.ME].
Ghosh, R., Mallick, B., and Pourahmadi, M., (2021) “Bayesian Estimation of Correlation Matrices of Longitudinal Data,” *Bayesian Analysis*, 16, Number 3, pp. 1039–1058.
Golts, M., and Jones, G., “A Sharper Angle on Optimization,” ssrn.com, 1483412.
Gong, H., Lu, C., and Zang, Y., (2024), “Distribution-consistency Structural Causal Models” arXiv:2401.15911v2 [cs.AI]
Hauser, A., and Bühlmann, P., (2012), “Characterization and Greedy Learning of Interventional Markov Equivalence Classes of Directed Acyclic Graphs,” *Journal of Machine Learning Research*, 13, 2409-2464.
Hellton, K., (2020), “Penalized Angular Regression for Personalized Predictions,” arXiv:2001.09834v2 [stat.ME].
Higham, N., (2002), “Computing the Nearest Correlation Matrix – A Problem from Finance,” *IMA Journal of Numerical Analysis*, 22, 329-343.

Huang, B., Zhang, K., Lin, Y., Schölkopf, B., and Glymour, C., (2018), “Generalized Score Functions for Causal Discovery,” KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 1551-1560.

Hulse, J., Eisty, N., and Menzies, T., (2025), “Shaky Structures: The Wobbly World of Causal Graphs in Software Analytics,” *Empirical Software Engineering*, 30:142.

Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P., (2010), “Estimation of a Structural Vector Autoregression Model using Non-Gaussianity,” *Journal of Machine Learning Research*, 11: 1709-1731.

Imbens, G., and Rubin, D., (2015), <u>Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction</u>, Cambridge University Press.

Ito, K., and Yoshiba, T., (2025), “Dynamic Asymmetric Tail Dependence Structure Among Multi-asset Classes for Portfolio Management: Dynamic Skew-*t* Copula Approach,” *International Review of Economics and Finance*, 97:103724.

Joarder, A., and Ali, M., (1992), “Distribution of the Correlation Matrix for a Class of Elliptical Models,” Communications in Statistics – Theory and Methods, 21(7), 1953-1964.

Jondeau, E., (2016), “Asymmetry in Tail Dependence of Equity Portfolios,” *Computational Statistics & Data Analysis*, Vol 100, pp351-368.

Junker. R., Griessenberger, F., and Trutschnig, W., (2021), “Estimating scale-invariant directed dependence of bivariate distributions,” *Computational Statistics & Data Analysis*, Volume 153.

Kaiser, M., and Sipos, M., (2022), “Unsuitability of NOTEARS for Causal Discovery when Dealing with Dimensional Quantities,” *Neural Processing Letters*, 54: 1587-1595.

Kalisch, M., & Bühlmann, P. (2007). "Estimating high-dimensional directed acyclic graphs with the PC-algorithm." *Journal of Machine Learning Research*, 8(22):613−636.

Kim, W., and Lee, Y., (2016), “A Uniformly Distributed Random Portfolio,” *Quantitative Finance*, Vol. 16, No. 2, pp.297-307.

Li, J., Chu, B., Scheller, I., Gagneur, J., and Maathuis, M., (2025), “Root Cause Discovery via Permutations and Cholesky Decomposition,” arXiv:2410.12151v5 [stat.ME].

Li, Q., (2018), “Covariance Modelling with Hypersphere Decomposition Method and Modified Hypersphere Decomposition Method,” Doctoral Thesis, University of Manchester.

Li, X., and Joe, H., (2024), “Multivariate Directional Tail-weighted Dependence Measures,” *Journal of Multivariate Analysis*, Vol 203.

Lin, Z., and Han, F., (2023), “On Boosting the Power of Chatterjee’s Rank Correlation,” *Biometrika*,110(2), 283–299.

Liu, Y., and Shang, P., (2025), “Differential Distance Correlation and Its Applications,” arXiv:2507.00524 [stat.ME].

Lopiz de Prado, M., (2023), <u>Causal Factor Investing</u>, Cambridge University Press.

MacKinnon, D., and Lamp, S., (2021), “A Unification of Mediator, Confounder, and Collider Effects,” *Prev Sci.*, 22(8), 1185-1193.

Marchenko, A., Pastur, L., (1967), "Distribution of eigenvalues for some sets of random matrices,” *Matematicheskii Sbornik*, N.S. 72 (114:4): 507–536.

Metsämuuronen, J., (2022), “Reminder of the Directional Nature of the Product-Moment Correlation Coefficient,” *Academia Letters*, Article 5313.

Opdyke, JD, (2003), “Fast Permutation Tests That Maximize Power Under Conventional Monte Carlo Sampling for Pairwise and Multiple Comparisons,” *Journal of Modern Applied Statistical Methods*, 2(1), 27-49.

Opdyke, JD, (2011), “Permutation Tests (and Sampling Without Replacement) Orders of Magnitude Faster Using SAS®,” *InterStat*, January, 2011.

Opdyke, JD, (2013), “Bootstraps, Permutation Tests, and Sampling Orders of Magnitude Faster Using SAS®,” *Computational Statistics: WIREs Interdisciplinary Reviews*, 5(5), 391-405.
Opdyke, JD, (2022), “Beating the Correlation Breakdown: Robust Inference and Flexible Scenarios and Stress Testing for Financial Portfolios,” QuantMindsEdge: Alpha and Quant Investing: New Research: Applying Machine Learning Techniques to Alpha Generation Models, June 6 (available at www.DataMineit.com).
Opdyke, JD, (2024a), Keynote Presentation: “Beating the Correlation Breakdown, for Pearson’s and Beyond: Robust Inference and Flexible Scenarios and Stress Testing for Financial Portfolios,” QuantStrats11, NYC, March 12, 2024 (available at www.DataMineit.com).
Opdyke, JD, (2024b), RoundTable Writeup: “Association-based vs Causal Research: the Hype, the Contrasts, and the Stronger-than-expected Complementary Overlaps,” QuantStrats11, NYC, March 12, 2024 (available at www.DataMineit.com).
Opdyke, JD, (2026), invited speaker, *PSTAT Seminar*, "The Highly Versatile Angles Space of Positive Definite Dependence Measures: Causal Discovery, Inference, Sampling, and Generalized Entropy," University of California at Santa Barbara, Department of Statistics and Applied Probability, 1/28/26.
Padh, K., Zhufeng, L., Casolo, C., and Kilbertus, N., (2025), “Your Assumed DAG is Wrong and Here’s How to Deal With It,” Proceedings of Machine Learning Research, 275:1-30, arXiv:2502.17030v1 [stat.ML]
Pamfil, R., Sriwattanaworachai, N., Desai, S., Pilgerstorfer, P., Beaumont, P., Georgatzis, K., and Aragam, B., (2020), “DYNOTEARS: Structure Learning from Time-Series Data,” Proceedings of the 23rd Intl Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Vol108.
Pascual-Marqui, R., Kochi, K., and Kinoshita, T., (2024), “Distance-based Chatterjee Correlation: A New Generalized Robust Measure of Directed Association for Multivariate Real and Complex-valued Data,” arXiv:2406.16458 [stat.ME].
Pearl, J., (2009), “Causal inference in statistics: An overview,” *Statistics Surveys*, 3:96–146, 2009.
Pearl, J., (2000), Causality: Models, Reasoning, and Inference, Cambridge University Press.
Pearl, J., and MacKenzie, D., (2018), The Book of Why: The New Science of Cause and Effect, New York: Basic Books.
Pinheiro, J., and Bates, D., (1996), “Unconstrained parametrizations for variance-covariance matrices,” *Statistics and Computing*, Volume 6, 289–296.
Polakow, D., Gebbie, T., and Flint, E., (2023), “Epistemic Limits of Empirical Finance: Causal Reductionism and Self-Reference,” arXiv:2311.16570v2 [q-fin.GN].
Pourahmadi, M., Wang, X., (2015), “Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor,” *Stat. and Probability Letters*, 106, (C), 5-12.
Pramanik, P., (2024), “Measuring Asymmetric Tails Under Copula Distributions,” *European Journal of Statistics*, 4:7.
Rapisarda, F., Brigo, D., and Mercurio, F., (2007), “Parameterizing Correlations: A Geometric Interpretation,” *IMA Journal of Management Mathematics*, 18(1), 55–73.
Reichenbach, H., (1956). *The Direction of Time*, University of California Press.
Reisach, A., Seiler, C., and Weichwald, S., (2021), “Beware of the Simulated DAG! Causal Discovery Benchmarks May Be Easy To Game,” *NeurIPS’21*, arXiv:2102.13647v3 [stat.ML].
Reisach, A., Suárez, A., Weichwald, S., and Chambaz, A., (2026), “The Case for Time in Causal DAGs,” working paper, arXiv:2501.19311v3 [stat.ME].
Rodriguez Dominguez, A., (2023), “Portfolio optimization based on neural networks sensitivities from ass ets dynamics respect common drivers,” *Machine Learning with Applications*, 11(15), 100447.
Rodriguez Dominguez, A., (2025a), “Causal Portfolio Optimization: Principles and Sensitivity-Based Solutions,” arXiv:2504.05743v2 [q-fin.PM].

Rodriguez Dominguez, A., (2025b), “Is Causality Necessary for Efficient Portfolios? A Computational Perspective on Predictive Validity and Model Misspecification,” arXiv:2507.23138v3 [q-fin.PM].
Rodriguez Dominguez, A., (2026), “Order-Constrained Spectral Causality in Multivariate Time Series,” https://ssrn.com/abstract=6035875 or http://dx.doi.org/10.2139/ssrn.6035875 .
Rodriguez Dominguez, A., and Yadav, O., (2024), “Measuring causality with the variability of the largest eigenvalue,” *Data Science in Finance and Economics*.
Sabato, S., Yom-Tov, E., Tsherniak, A., Rosset, S., (2007), “Analyzing systemlogs: A new view of what’s important,” *Proceedings, 2nd Workshop of Computing Systems ML*, pp.1–7.
Saxena, S., Bhat, C., and Pinjari, A., (2023), “Separation-based parameterization strategies for estimation of restricted covariance matrices in multivariate model systems,” *Journal of Choice Modelling*, Vol. 47, 100411.
Shimizu, S., Hoyer, P., Hyvärinen, A., and Kerminen, A., (2006), “A Linear Non-Gaussian Acyclic Model for Causal Discovery,” *Journal of Machine Learning Research*, 7:2003-2030.
Sprites, P., (2001), “An Anytime Algorithm for Causal Inference,” *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics*, PMLR R3:278-285, 2001.
Szekely, G., Rizzo, M., and Bakirov, N., (2007), “Measuring and Testing Dependence by Correlation of Distances,” *The Annals of Statistics*, 35(6), pp2769-2794.
Tsay, R., and Pourahmadi, M., (2017), “Modelling structured correlation matrices,” *Biometrika*, 104(1), 237-242.
Vinod, H., (2022), “Asymmetric Dependence Measurement and Testing,” arXiv:2211.16645v1 [stat.ME].
Wang, W., Huang, B., Liu, F., You, X., Liu, T., Zhang, K., and Gong, M., (2024), “Optimal Kernel Choice for Score Function-based Causal Discovery,” *Proceedings of the 41st International Conference on Machine Learning*, PMLR, 235:50650-50668.
Wright, S., (1921), “Correlation and Causation,” *Journal of Agricultural Research*, 20(7), 557-585.
Xia, L., Cao, R., Du, J., and Chen, X., (2025), “The Improved Correlation Coefficient of Chatterjee,” *Journal of Nonparametric Statistics*, pp1-17.
Xue, A., Rao, J., Sarkararaman, S., and Pimentel, H., (2025), “dotears: Scalable and Consistent Directed Acyclic Graph Estimation Using Observational and Interventional Data,” *iScience*, 28, 111673.
Yule, G.U., (1897), “On the Significance of Bravais Formulæ for Regression, in the Case of Skew Correlation,” *Proceedings of The Royal Society London*, 477-489.
Zanga, A., Ozkirimlib, E., and Stella, F., (2026), “A Survey on Causal Discovery: Theory and Practice,” arXiv:2305.10032v2 [cs.AI] 26 August.
Zar, J., (1999), Biostatistical Analysis, 4th Ed., Prentice Hall, New Jersey.
Zhang, Y., (2022), “Angle Based Dependence Measures in Metric Space,” arXiv:2206.01459v1 [stat.ME].
Zhang, Y., and Yang, S., (2025), “Kernel Angle Dependence Measures in Metric Spaces,” *Journal of Computational Statistics*, 34(1), 280-289.
Zhang, W., Leng, C., and Tang, Y., (2015), “A Joint Modeling Approach for Longitudinal Studies,” *Journal of the Royal Stat. Society, Series B*, 77(1), 219-238.
Zhang, Q., (2024a), “On relationships between Chatterjee’s and Spearman’s correlation coefficients,” *Communications in Statistics – Theory and Methods*, 54(1), 259-279.
Zhang, Q., (2024b), “On the Extensions of the Chatterjee-Spearman Test” arXiv:2406.16859v1 [stat.ME]
Zheng, S., Shi, N.-Z., and Zhang, Z. (2012), “Generalized Measures of Correlation for Asymmetry, Nonlinearity, and Beyond,” *Journal of the American Statistical Association*, 107, 1239–1252.
Zheng, X., Aragam, B., Ravikumar, P., and Xing, E., (2018), “DAGs with NO TEARS: Continuous Optimization for Structure Learning,” working paper, arXiv:1803.01422v2 [stat.ML].

7. APPENDIX A1: Translations Between Dependence Measure Matrices & Angles Matrices

Scenario 1 (S1): Identity Matrix

Dependence Measure Matrix

1	0	0	0	0
0	**1**	0	0	0
0	0	**1**	0	0
0	0	0	**1**	0
0	0	0	0	**1**

Angles Matrix

π/2				
π/2	π/2			
π/2	π/2	π/2		
π/2	π/2	π/2	π/2	

Scenario 2 (S2): Block Matrix

Dependence Measure Matrix

1	-0.2	-0.2	0.3	0.3
-0.2	**1**	-0.2	0.3	0.3
-0.2	-0.2	**1**	0.3	0.3
0.3	0.3	0.3	**1**	0.6
0.3	0.3	0.3	0.6	**1**

Angles Matrix

1.772				
1.772	1.823			
1.266	1.175	1.002		
1.266	1.175	1.002	1.295	

Cell # Matrix for Angles PDFs below in Appendix A3:

Cell # Matrix for PDFs

7				
8	**4**			
9	**5**	**2**		
10	**6**	**3**	**1**	

APPENDIX A Comments:

1. As expected, note that the identity (independence) matrix yields symmetric angles pdfs centered on π/2, as expected. Also as expected, the corresponding empirical spectral distribution aligns with that of Marchenko Pastur (see Marchenko and Pastur (1967)).

2. As expected, the spectral distribution of Pearson's rho is more variable than that of Kendall's tau, as it is generally more sensitive to outlying observations.

3. Not that the characteristics of the spectral distributions compared to those of the angles distributions, i.e. more asymmetry, multi-modality, and long, essentially unbounded tails, make their estimation more challenging and less robust than that of the individual angles distributions. See Opdyke (2022, 2026) for more on this topic.

4. Note the larger scale of the y-axes for the improved Chatterjee pdfs.

7. APPENDIX A2: Dependence Measure Spectral Distributions, Five Data Generating Mechanisms

Scenarios (n=252, Marchenko Pastur (Independence) Added as Referential Baseline)

S1: Measure = Pearson's Rho; Dependence Structure = Identity Matrix; DGM = Multivariate Gaussian

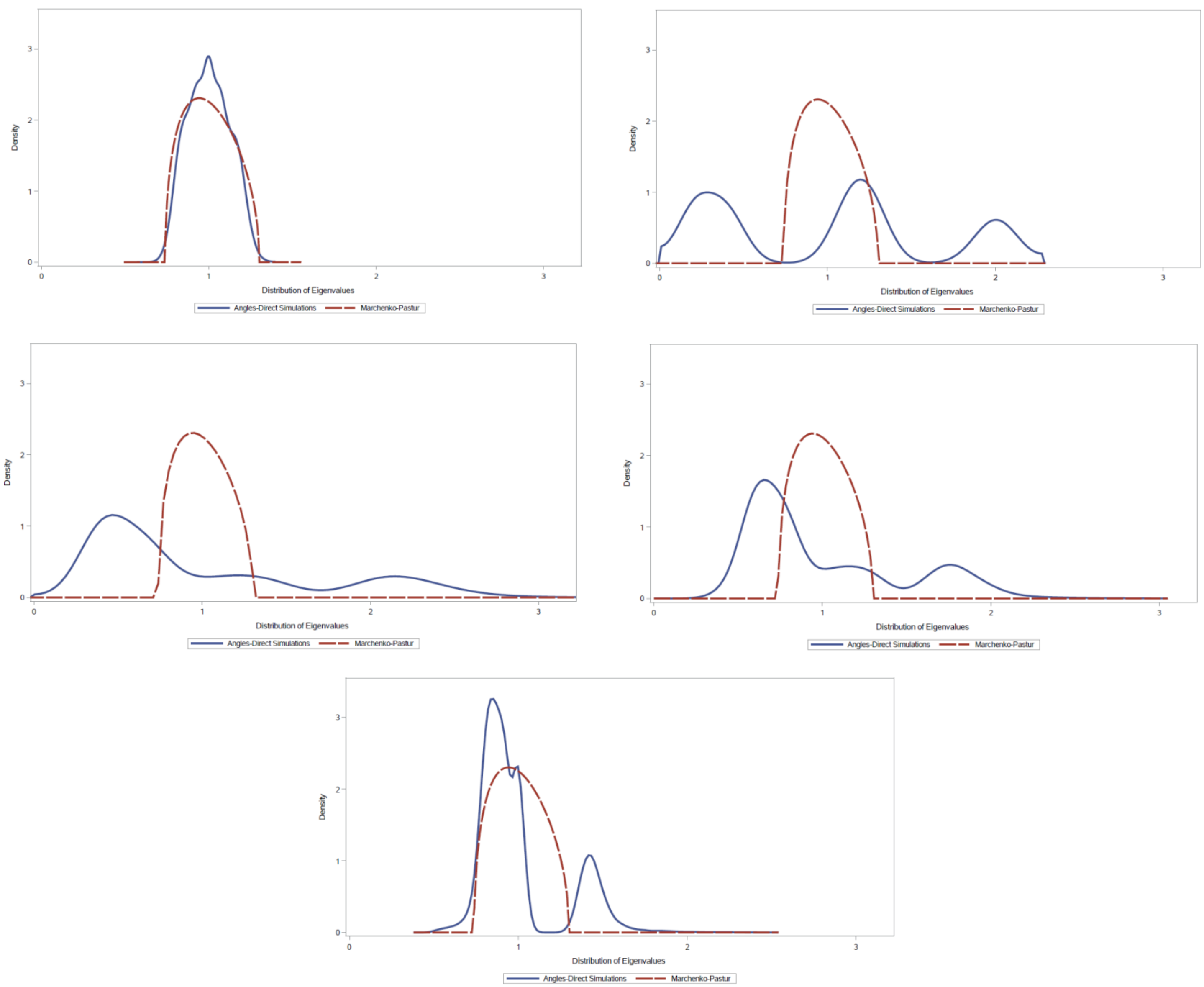
S2: Measure = Pearson's Rho; Dependence Structure = Block Matrix; DGM = Multivariate Gaussian

S2A: Measure = Pearson's Rho; Dependence Structure = Block Matrix;
DGM = Marginal Distributions with Different Asymmetry, Heavy-Tailedness, Serial-Correlation, and Non-stationarity (via Church's (2012) asymmetric student's t copula with varying degrees of freedom: serial correlation and non-stationarity imposed ex post, with subsequent rescaling of the first two moments)

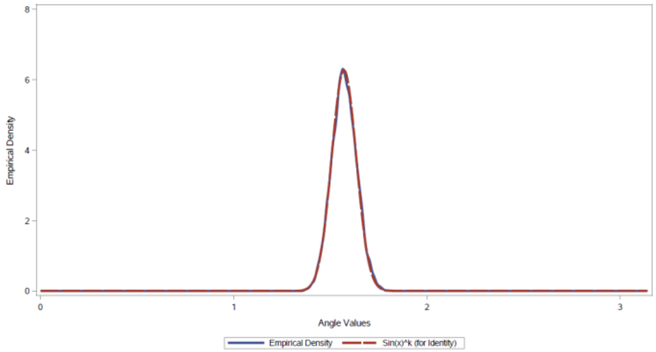
S2B: Measure = Kendall's tau; Dependence Structure = Block Matrix; DGM = Same as S2A

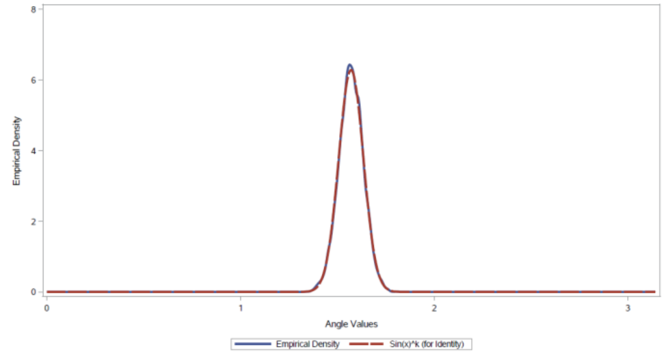
S2C: Measure = Improved Chatterjee's; Dependence Structure = Block Matrix; DGM = Same as S2A

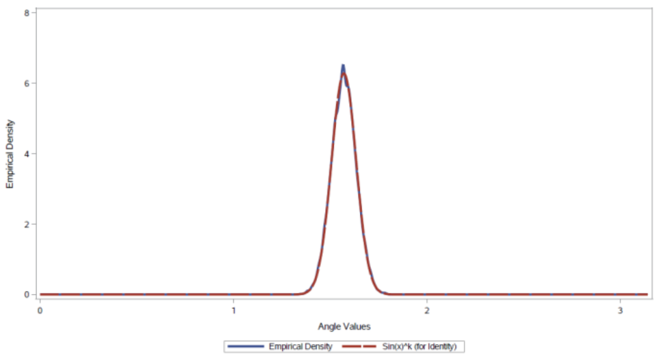
Spectral Distributions: S1, S2, S2A, S2B, S2C

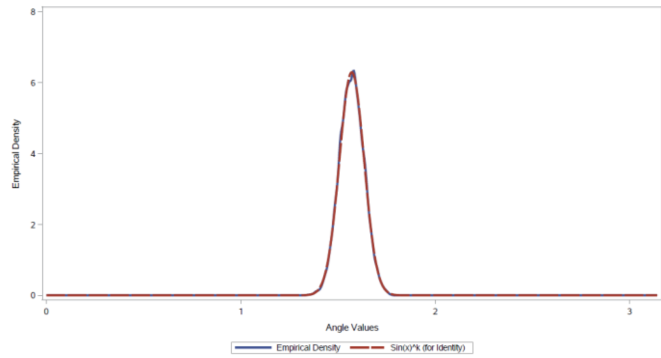


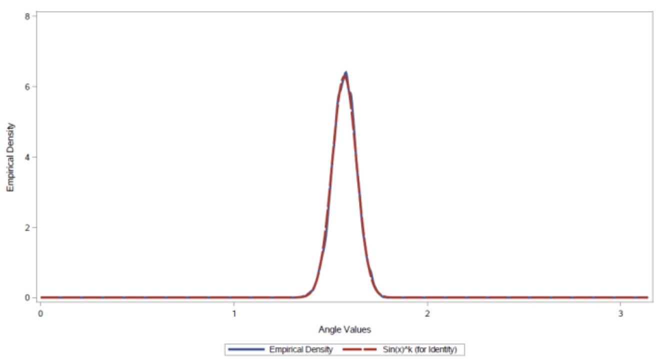
7. APPENDIX A3: Scenario S1 Angles PDFs: n=252, NSims=10k, Cell # Per Appendix A1 (1, 2, ..., 10)

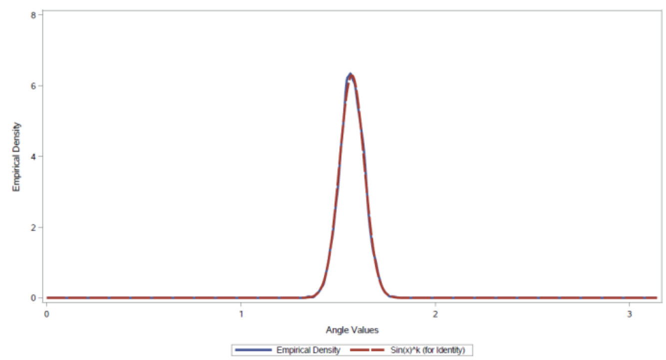


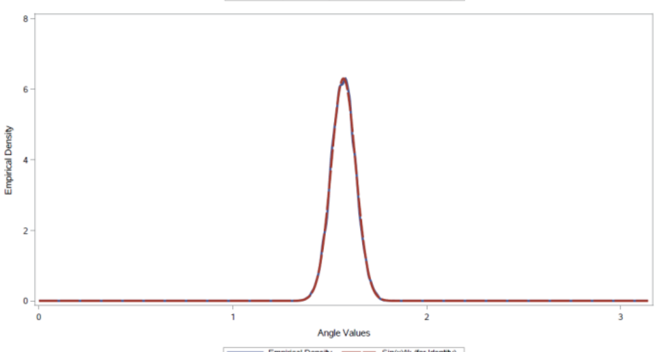


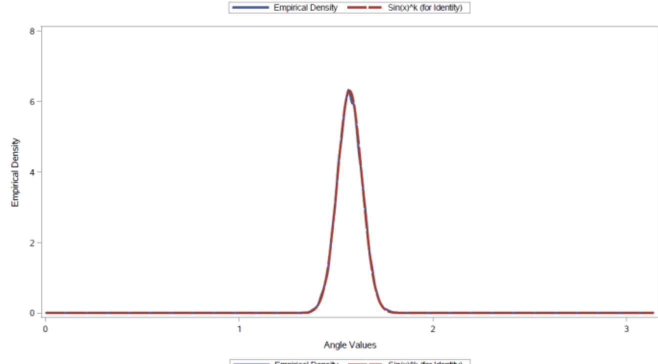


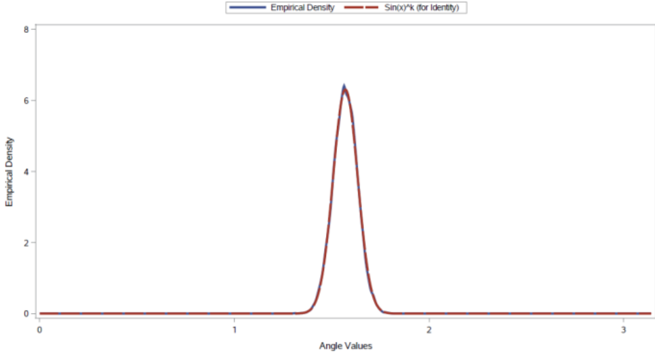


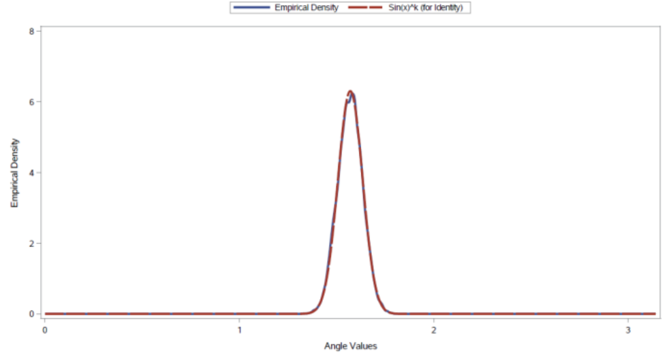




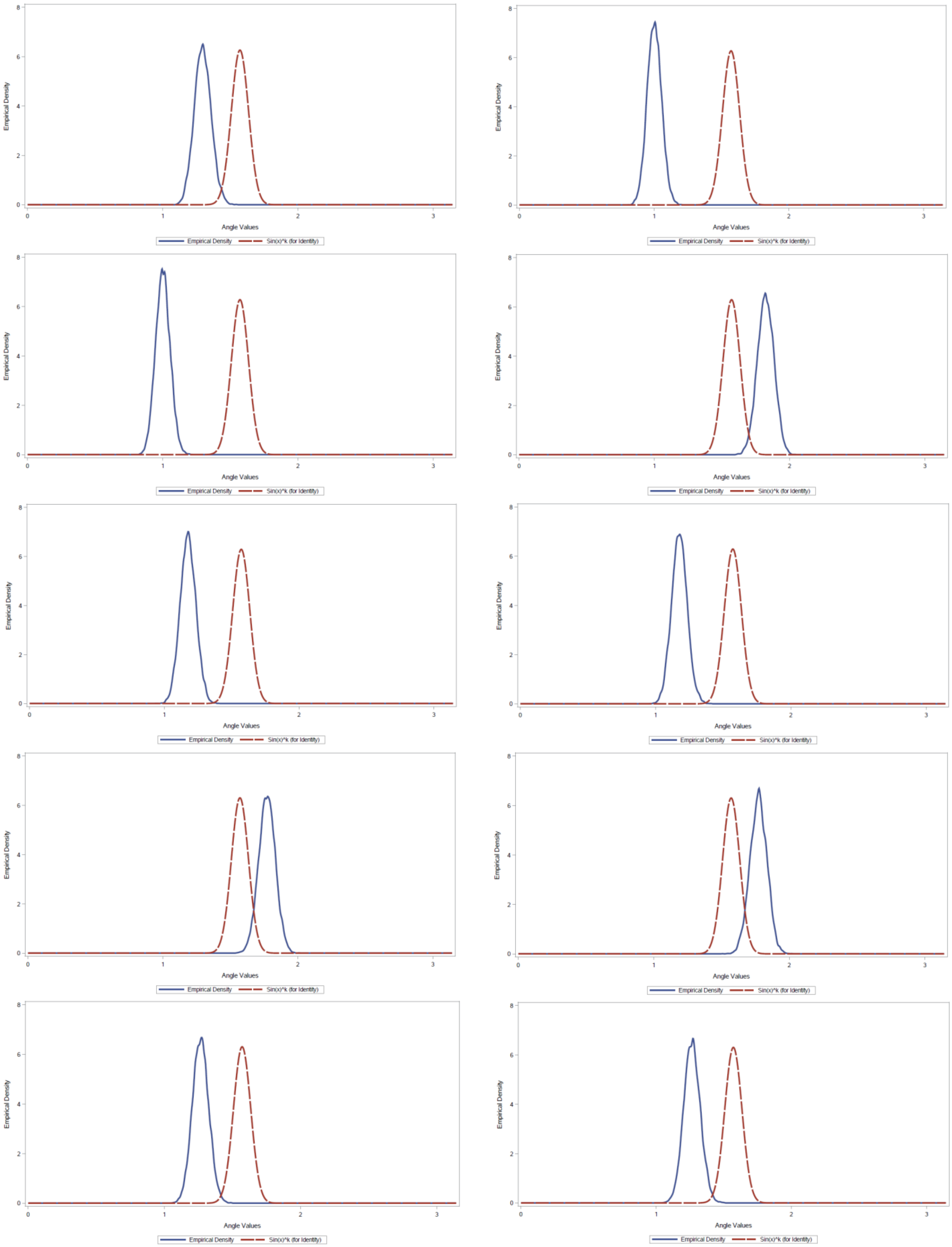




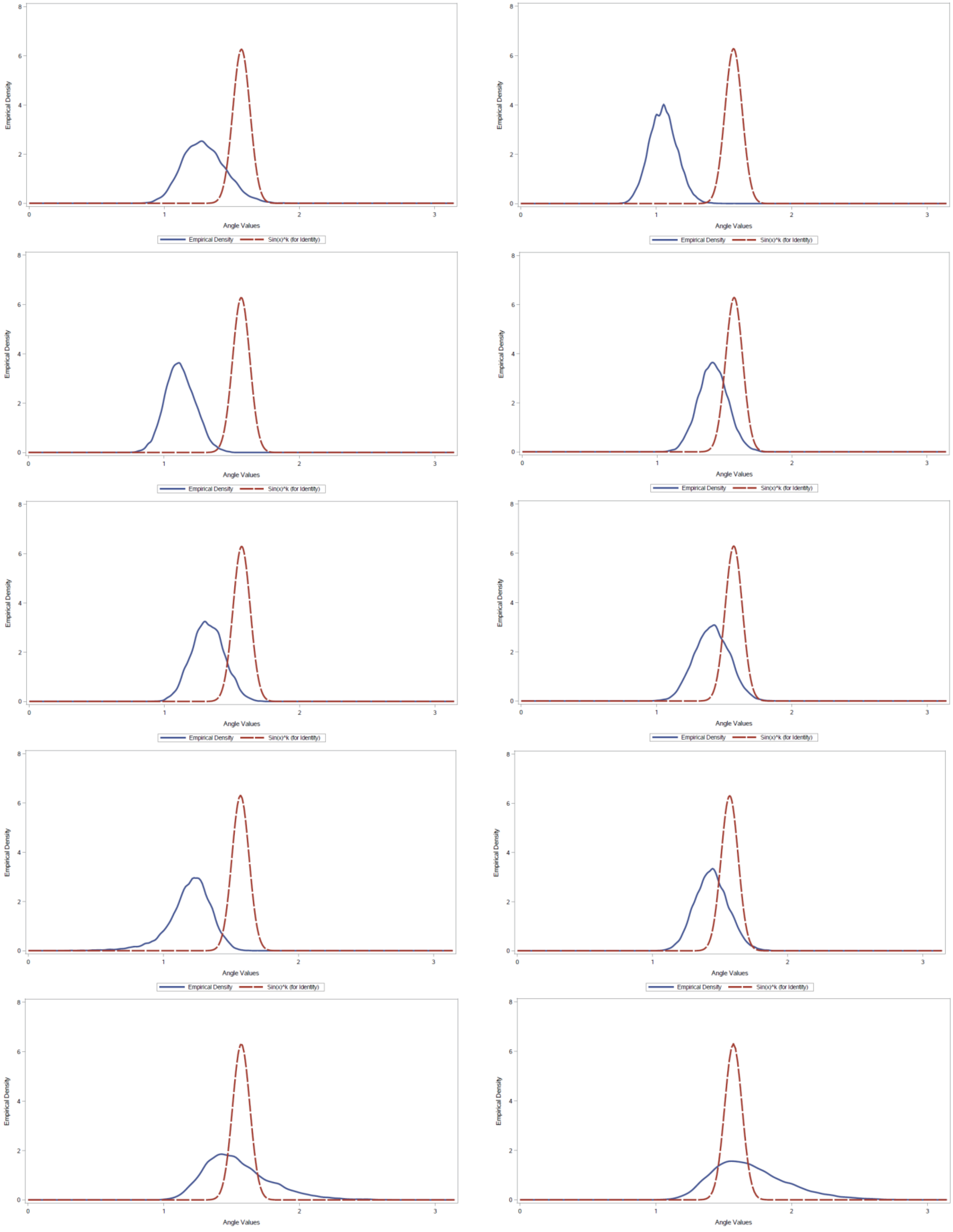




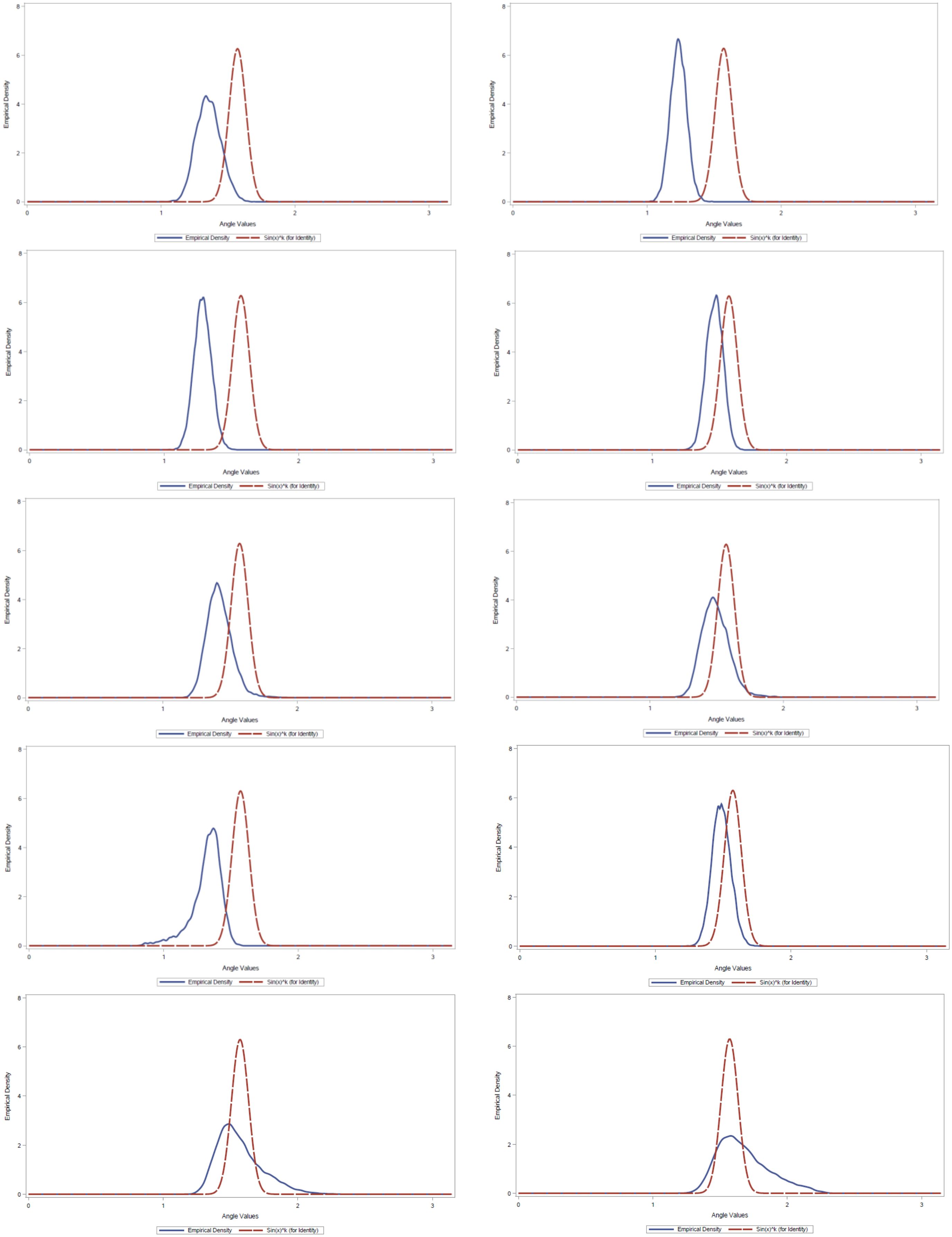
7. APPENDIX A3: Scenario S2 Angles PDFs: n=252, NSims=10k, Cell # Per Appendix A1 (1, 2, ..., 10)



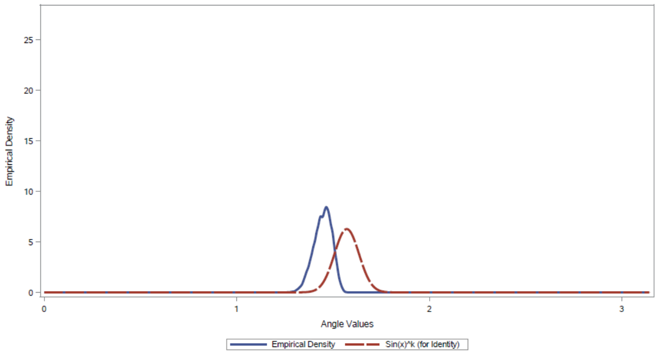
7. APPENDIX A3: Scenario S2A Angles PDFs: n=252, NSims=10k, Cell # Per Appendix A1 (1, 2, ..., 10)

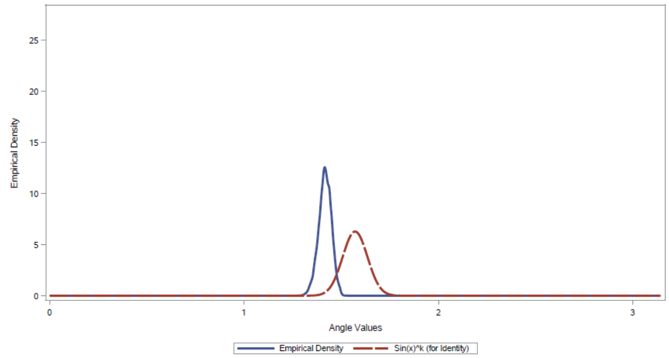


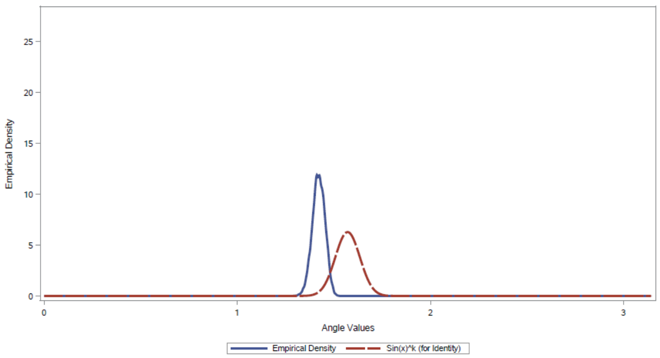
7. APPENDIX A3: Scenario S2B Angles PDFs: n=252, NSims=10k, Cell # Per Appendix A1 (1, 2, ..., 10)

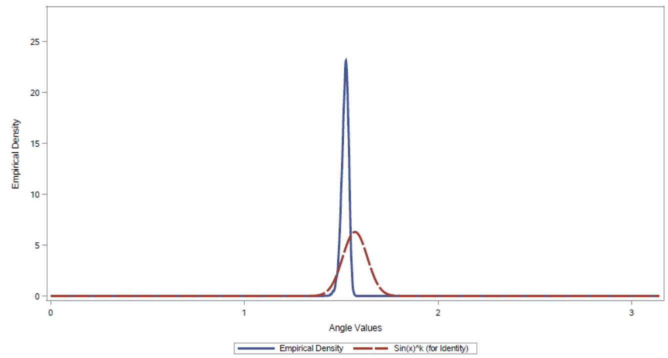


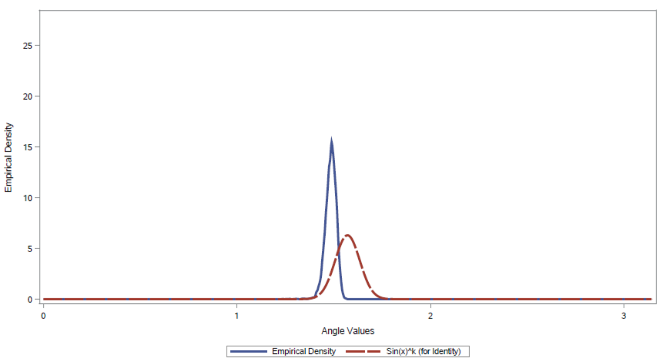
7. APPENDIX A3: Scenario S2C Angles PDFs: n=252, NSims=10k, Cell # Per Appendix A1 (1, 2, ..., 10)

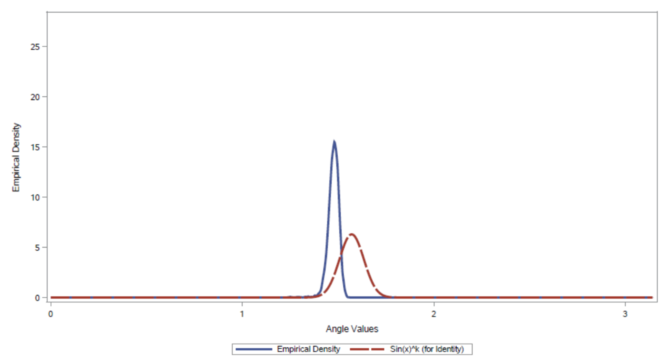


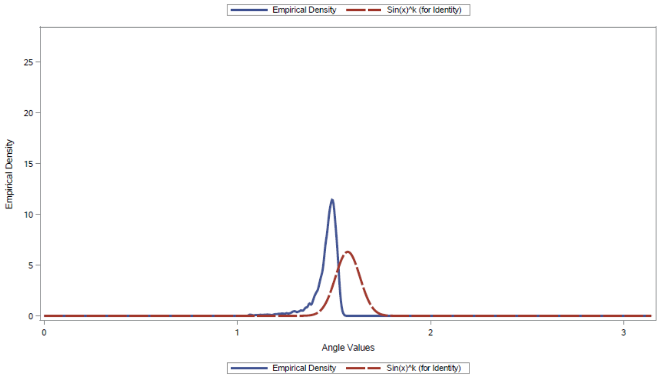


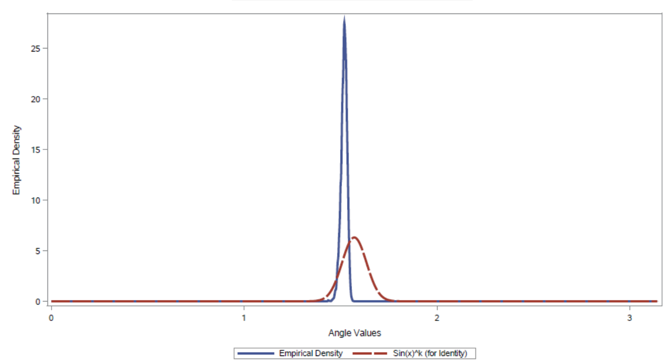


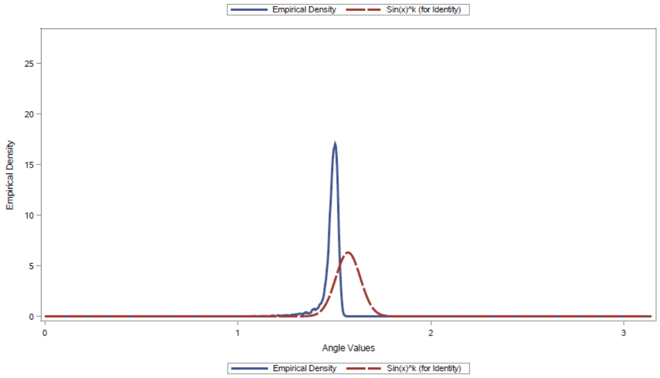


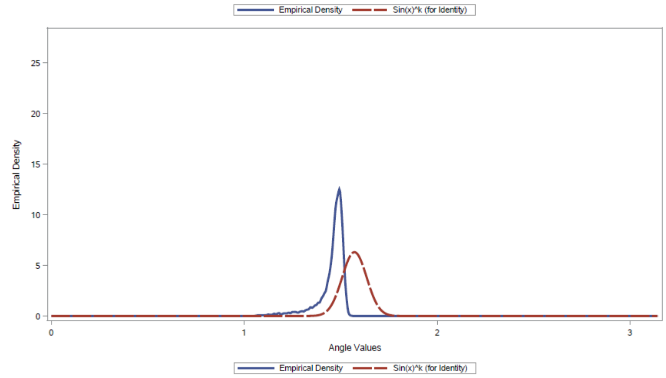












7. Appendix B: Efficient Coding of Improved Chatterjee Correlation of Xia et al. (2025)

SAS/IML code (v9.4): Improved Chatterjee

```
*** INPUTS: data4chat is raw input data, reverse is a flag to reverse column order;
*** OUTPUT ichat is the improved Chatterjee matrix;

  start corr_ichat(data4chat,reverse);
    dim = dimension(data4chat);
    ncols=dim[1,2];
    ncols_m1=ncols-1;
    nrows=dim[1,1];
    nrows_m1=nrows-1;
    ncells=ncols**2;
    ichat=J(ncols,ncols,.);
    ichat[do(1,ncells,ncols+1)]=1;
    if reverse=1 then data4chat=data4chat[,do(ncols,1,-1)];
    if nrows>ncols then do;
      do i=1 to ncols_m1;
        data4chat2=data4chat[,i:ncols];
        call sort(data4chat2);
        do j=i+1 to ncols;
           k=j-i+1;
           rnks = rank(data4chat2[,k]);
           cumsum=0;
           cumwgts=0;
           do q=1 to nrows_m1;
             cumsum=cumsum+sum(abs(dif(rnks,q)/q));
             cumwgts=cumwgts+(nrows-q)/q;
           end;
           ichat[i,j]=1-cumsum/(cumwgts*(nrows+1)/3);
           ichat[j,i]=ichat[i,j];
        end;
      end;
   end;
   else do;
     print "Data provided to corr_ichat subroutine must be rull rank.";
   end;
   return(ichat);
  finish;
```

Note that the "improved Chatterjee correlation" of Xia et al. (2025) typically would be coded with a nested loop, but this can be avoided by calculating the entire equation <u>diagonally</u>, thus increasing speed by an order of magnitude (where an order of magnitude is the dimension of the matrix, p). This can save considerable amounts of real runtime when matrices are large(r) (e.g. p≥100) AND these matrices must be calculated many times in many simulations (e.g. Nsim ≥10,000).

7. Appendix C: Assumptions Required for Valid Causal Interpretation

A. Below I define the assumptions under which the directional associations identified by ADD can be interpreted causally, and the causal meaning of edges inferred by ADD in a DAG.

<u>Preliminaries</u>: Structural Causal Model (SCM) and do-Semantics (see Pearl, 2009)

Let $V=\left\{X_1,\ldots,X_p\right\}$ be observed variables generated by an acyclic Structural Causal Model (SCM):

$X_j = f_j\left(X_{Pa(j)}, U_j\right)$ for $j=1,\ldots,p$

Here Pa(j) are the parents of X_j in a DAG "G," $U_1,\ldots,U_p$ are jointly independent exogenous noises, and each $f_j(\ )$ is a (possibly nonlinear, nonparametric) measurable function. The graph G encodes the causal Markov property and admits do-interventions: for any variable X, the operation do(X:=x) replaces its structural equation with the constant x, severing incoming edges into X.

<u>Causal effect</u>: X has a (possibly context-dependent) causal effect on Y if and only if there exist x ≠ x' such that P(Y | do(X=x)) ≠ P(Y | do(X=x')).

<u>Assumptions for Causal Interpretation</u>:

(A1) <u>Acyclicity</u>: The true causal graph over V is a DAG.

(A2) <u>Causal Sufficiency</u> (no latent confounders among V): The exogenous noises U_j are mutually independent; any common causes of variables in V are included in V. Relaxation: if (A2) is uncertain, interpret edges as ancestral (existence of a directed path) rather than necessarily adjacent (direct).

(A3) <u>Markov + (qualified) faithfulness</u>: The observed joint distribution is Markov with respect to G and violations of direction-revealing asymmetries are non-degenerate (i.e. pathological cancellations are excluded).

(A4) <u>i.i.d. sampling (or weak stationarity for time-series)</u>: Samples are independent draws from the same distribution (future research plans to apply ADD, for time-series extensions, to innovations or appropriate lags).

(A5) <u>Measurement reliability</u>: Variables are measured without systematic differential error that inverts directional asymmetries.

(A6) <u>Conditioning set for direct edges</u>: When inferring adjacency (direct edges), directional dependence is evaluated conditionally on V\{X,Y}. If only marginal directional dependence is used, interpret edge calls as directed reachability (existence of some directed path).

(A7) <u>Monotone link from mechanism to asymmetry</u>: For the chosen directional measure *M* (in the empirical study above, this is ICH – the improved Chatterjee's correlation of Xia et al. (2025)), if X ∈ Pa(Y) and Y ∉ Pa(X), then asymptotically E[M(X→Y)] > E[M(Y→X)], with separation increasing in sample size n under the SCM.

Note that this assumption plays a fundamental role in linking the structural causal model to observable directional asymmetries. It requires that the directional dependence functional M increase in expectation in the true causal direction. This monotone link between mechanism and measurable asymmetry is the critical identifiability condition for ADD-based DAG recovery. In practice, validation of (A7) proceeds through a simulation across a range of nonlinear, non-Gaussian, and/or heteroskedastic settings, ensuring that E[M(X→Y)] > E[M(Y→X)] consistently emerges under the data-generating process.

Among the seven assumptions, only (A7) must be verified empirically.

Edges in ADD:

Let M be a directional dependence functional and $\hat{M}$ its sample estimate. ADD maps the full matrix of $\hat{M}$ values to angles, performs finite-sample joint inference under positive-definiteness, and returns p-values for directed hypotheses.

Adjacency claim (direct edge): Under (A1)–(A7) and using conditional directional dependence M(X→Y | V\{X,Y}): declare a directed edge X→Y if statistically significant(X→Y) and NOT statistically significant(Y→X) at level α (the one-direction-only rule). This is interpreted as X ∈ Pa(Y).

Ancestral/path claim (directed reachability): If only marginal M is used, interpret the same rule as: there exists a directed path X ⇝ Y (not necessarily adjacent); if (A2) is violated, reduce the claim to directional dependence without causal interpretation.

No-edge claim: If neither direction is statistically significant, refrain from a causal claim.

Cycle handling: If the one-direction-only rule yields a 2-cycle, pruning enforces acyclicality by selecting the edge/direction that has a larger p-value and replacing its dependence measure value with a value just below the α threshold of statistical significance. The same is done when longer cycles are encountered. The specific algorithm for the latter case that was implemented in the above empirical study used Kahn's (1962) topological sorting algorithm to identify cycles, and then i. selected those nodes/variables with the fewest edges; ii. then among those, selected the one with the largest sum of p-values across its related edges (i.e. the weakest cumulative signals/edges); and iii. finally, selected the particular edge from among those that had the largest p-value (weakest signal/edge). This process was repeated until no cycles remained. In the above empirical study, no pruned matrices became non-positive definite, which is not surprising given that the pruning adjustments place the matrix further within the positive definite cone (still, positive definiteness always was empirically verified). But even if positive definiteness had been occasionally violated, non-linear constrained optimization is relatively straightforward to implement here. While this pruning algorithm is not multivariate optimized, it is fast and straightforward, and was effective in the empirical study herein. A follow-up study will explore optimizing this approach, possibly by minimizing an aggregate angle-space loss function under positive definite constraints.

Under (A1)–(A7), the one-direction-only ADD decision rule using conditional *M* is pointwise sound for adjacency: if the rule declares X→Y, then X ∈ Pa(Y) with probability → 1 as n → ∞.

By Markov + faithfulness (A3) and SCM acyclicity (A1), conditioning on V\{X,Y} blocks all non-causal paths between X and Y. If X ∈ Pa(Y), then by (A7) the population asymmetry satisfies *M*(X→Y) > *M*(Y→X); ADD's finite-sample tests are consistent in n (angles-based inference), so the probability of correctly calling X→Y tends to 1. Conversely, if no directed edge X→Y exists, all back-door and collider paths are blocked by the conditioning set, implying symmetric (null) directional dependence; the rule refrains from a causal call with probability → 1.

Note that when *M* is conditional on other variables, an edge X→Y indicates a direct causal relationship, but when *M* is marginal, the same edge indicates only a causal pathway or reachability, not necessarily a direct connection. In the empirical study conducted herein, no conditioning was used, and consequently, the directed edges recovered therein should be interpreted as reachability rather than as direct causal effects.

Finally, note again that ADD, as defined herein, identifies causal structure, but does not estimate effect sizes or counterfactual outcomes.

B. Below I place ADD within the Causal Inference Literature

<u>Proposition</u>:

Let $V = \left\{X_1,\ldots,X_p\right\}$ be generated by acyclic structural equations $X_j = f_j\left(X_{Pa(j)},U_j\right)$ with mutually independent exogenous noises U_j. Under the causal Markov property and (qualified) faithfulness, the observed distribution is Markov to the causal DAG "G." I adopt these semantics and make the following explicit: (i) acyclicity; (ii) causal sufficiency (or an ancestral interpretation if violated); (iii) i.i.d./weak-stationarity; (iv) measurement reliability. Directional inference in ADD uses a directional dependence functional *M*; edges are interpreted as direct parents when *M* is evaluated conditionally on V\{X,Y}, and as directed reachability when evaluated marginally.

This makes ADD comparable with established approaches in the causal literature (see Zanga et al., 2025). Some of these include constraint-based methods, such as PC algorithms (see Kalisch, M., & Bühlmann, P. (2007)), FCI algorithms (see Sprites (2001)), and Markov Blanket algorithms (see Aliferis et al. (2010), and Chen et al., (2026)); they also include score-based methods, such as NOTEARS (see Zheng et al. (2018)) and its variants (e.g. DYNOTEARS, see Pamfil et al. (2020)), Greedy Equivalence Search (see Chickering, D., (2002) and Hauser and Bühlmann (2012)) and Kernel-based Score Functions (see Huang et al., (2018) and Wang et al., (2024)); finally, they include functional causal models, such as LiNGAM (see Shimizu et al. (2006)), and its variants such as VarLiNGAM (see Hyvärinen et al. (2010)). Unlike those examples, however, ADD takes a completely original approach by performing finite-sample inference with positive-definite constraints at the matrix level, which may very well improve coherence of edge calls across the graph, especially compared to those algorithms that identify causal edges sequentially. This will be tested in a more extensive, follow-up empirical study pitting ADD against many of the above-mentioned competitors.