

$$f_X(x) = c_k \cdot \sin^k(x), \quad x \in (0, \pi), \quad k = 1, 2, 3, \dots, \text{\#columns} - 1, \quad \text{and } c_k = \frac{\Gamma(k/2 + 1)}{\sqrt{\pi} \Gamma(k/2 + 1/2)}$$

$$F_X(x; k) \sim \frac{1}{2} - \left(\frac{1}{2}\right) \cdot F_{Beta}\left[\cos^2(x); \frac{1}{2}, \frac{1+k}{2}\right] \text{ for } x < \frac{\pi}{2}, \quad \sim \frac{1}{2} + \left(\frac{1}{2}\right) \cdot F_{Beta}\left[\cos^2(x); \frac{1}{2}, \frac{1+k}{2}\right] \text{ for } x \geq \frac{\pi}{2}$$

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

$$F^{-1}(p; k) = \arccos\left(\sqrt{F_{Beta}^{-1}\left(1 - 2p; \frac{1}{2}, \frac{1+k}{2}\right)}\right) \text{ for } p < 0.5;$$

**JD Opdyke**

**Chief Analytics Officer, DataMineit, LLC**

**Invited Speaker, Causal Data Science Meeting 2026,  
Session 4, 5/4/26**

$$= \pi - \arccos\left(\sqrt{F_{Beta}^{-1}\left(1 - 2[1 - p]; \frac{1}{2}, \frac{1+k}{2}\right)}\right) \text{ for } p \geq 0.5$$



# Acknowledgments & Disclaimer

This presentation accompanies a paper of the same title that has been accepted for publication in the book series, Annual Review of Modern Quantitative Finance, vol.2, eds. Itkin, A., World Scientific, 2026-2027.

**Link to Download Paper Preprint, Presentation, and Video:**

[http://www.DataMineit.com/DML\\_publications.htm](http://www.DataMineit.com/DML_publications.htm)

**Disclaimer:** I, John Douglas Opdyke, am the sole author of the work contained herein, and any errors are my own. The views presented herein are solely those of the author and do not necessarily reflect the views of specific institutions.

**Acknowledgments:** I would like to sincerely thank Alejandro Rodriguez Dominguez, PhD, for his multiple, invaluable reads of the paper accompanying this presentation. His insights greatly improved both its clarity and the relevance of its application to capital markets and other empirical research settings.

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 (positive definite) angles space, and then enforce acyclicity 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 a (fast) simulation to define empirical confidence bounds under independence (regardless of the size of the DAG space). We conduct a preliminary empirical study evaluating direct adjacency under nonlinear, asymmetric, heavy-tailed data conditions. The promising results indicate applications to feature selection in quantitative finance and 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.

- 1. Introduction / Summary**
- 2. Background**
  - A. Asymmetric, Directional Dependence**
  - B. Relationship between Measures of Association and Causality**
  - C. The Angles Space**
- 3. Methodology**
  - A. Rationale: Dual Orderings Using the Angles Space of Directional Dependence Measures**
- 4. Preliminary Empirical Study**
- 5. Conclusions and Next Steps**
- 6. References**
- 7. Appendices A, B, C, D**

# 1. Introduction / Summary

## Objective:

- Causal **Discovery** (not Causal Inference, estimating effect sizes, counterfactual outcomes, etc.)

## Framework:

- Directed Acyclic Graph (DAG)
- **Assumptions:** all the usual suspects (acyclicity, causal sufficiency, Markov and faithfulness, independent and identically distributed sampling, reliable measurement, and monotone link from (directional) dependence measure to asymmetry)
- **Limitations:** all the usual suspects (time scale issues, temporality and true acyclicity, etc.)

## Execution: Most causal discovery algos utilizing DAGs execute

- **Sequentially** (e.g. PC algorithms (see Kalisch, M., & Bühlmann, P. (2007)), FCI algorithms (see Spirtes (2001)), and Markov Blanket algorithms (see Aliferis et al. (2010), and Chen et al., (2026)))
- **Iteratively** (e.g. Score-Based methods (NOTEARS, Zheng et al. (2018), DYNOTEARS, Pamfil et al. (2020)), Greedy Equivalence Search (Chickering, D., (2002) and Hauser and Bühlmann (2012)), Kernel-based Score Functions (Huang et al., (2018) and Wang et al., (2024)), and Functional Causal Models (LiNGAM, Shimizu et al. (2006)), VarLiNGAM, Hyvärinen et al. (2010)).

## New Method: Angles-based Directional Dependence, “**ADD**,” executes

- **Simultaneously**, over the entire DAG space

# 1. Introduction / Summary

**New Method**: Angles-based Directional Dependence, “**ADD**”

- **Executes Simultaneously** over the entire DAG space.
- Uses **2 matrix estimations** of any Positive Definite **Directional** Dependence Measure, one in each direction (i.e. 1 original column ordering & 1 reverse column ordering ... covers all possible edges).
- (**Apply conditioning** also if seeking direct adjacency rather than reachability)
- **Angles Space**: Translate the 2 matrices into 2 Angles Matrices (a multivariate bijection).
- **Dual Orderings / One-Direction-Only Rule**: Conduct inference within the Angles space, comparing each cell of the 2 Angles Matrices against distributions under the null of independence obtained via (fast) empirical simulations, or via (finite-sample) analytic results if they exist in the extant literature for that dependence measure; call an edge if statistically significant in one direction, but not in the other direction.
- **DAG requirements**: Apply **pruning** to enforce acyclicity (under constraint of positive definiteness).

**Potential Advantages of ADD**:

- **Increased Coherence**, due to simultaneous edge-calling within a robust, 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.
- **Increased Speed/Scalability**, due to the need for only two matrix estimations, and (possibly) a (fast) simulation to define empirical confidence bounds under independence (regardless of the size of the DAG space).

## 2. Background: Directional Dependence Measures (housekeeping)

Housekeeping: **All-pairwise matrix**, positive definiteness, symmetry enforced (asymmetric measures)

- First, **the dependence structure framework is the all-pairwise matrix**, that is, the matrix of all bivariate relationships between all variables in the portfolio, and **positive definiteness is defined in relation to this framework**, NOT in the sense of multivariate measures of dependence (see Cardin, 2009).

$$R = \begin{bmatrix} 1 & r_{1,2} & r_{1,3} & r_{1,4} \\ r_{2,1} & 1 & r_{2,3} & r_{2,4} \\ r_{3,1} & r_{3,2} & 1 & r_{3,4} \\ r_{4,1} & r_{4,2} & r_{4,3} & 1 \end{bmatrix}$$

i. symmetry:

$$r_{i,j} = r_{j,i}$$

ii. unit diagonals:

$$r_{i=j} = 1$$

iii. bounded non-diagonals, maximum range\*:

$$(-1 \text{ or } 0) \leq r_{i,j} \leq 1$$

iv. positive definiteness, i.e. all eigenvalues

$$\lambda_i > 0$$

- For completeness, and for reference throughout this presentation, I define eigenvalues for  $R \in \mathbb{R}^{p \times p}$  here: if there exists a nonzero vector  $v$  such that  $Rv = \lambda v$  then  $\lambda$  is an eigenvalue of  $R$  and  $v$  is its corresponding eigenvector.  $\lambda$  and  $v$  can be obtained by solving  $\det(\lambda I - R) = 0$ , then  $\det(\lambda I - R)v = 0$ , where  $I$  is the identity matrix and  $\det$  is the determinant.

\* This range varies by measure, by data conditions for a given measure, and for finite samples vs. asymptotic results, as described in more detail below.

## 2. Background: Directional Dependence Measures (housekeeping)

Housekeeping: All-pairwise matrix, **positive definiteness**, symmetry enforced (asymmetric measures)

- Note that even beyond finance, positive definite dependence measures arguably span the majority of those in widespread usage, making this requirement less of a restriction and more of a test for the appropriateness and utility of a given dependence measure under given data conditions.
- For **full transparency**, I classify dependence measures vis-à-vis their positive definiteness:
  1. **Proven Analytically**: Long-established measures (e.g. Pearson's, Spearman's, Kendall's, Tail Dependence Matrix) for which positive definiteness has been proven analytically (see Sabado et al. (2007), and Embrechts et al. (2016)).
  2. **Empirically Consistent Results with 1.**: Newer measures for which positive definiteness has not (yet) been analytically proven, but which exhibit non-positive definiteness (NPD) no more often than what would be expected solely due to numerical issues, based on comparisons to measures in category 1., thus indicating likely structural positive definiteness (which still always must be empirically verified).
  3. **Empirically INconsistent Results with 1.**: Measures which exhibit NPD MORE often than what would be expected solely due to numerical issues.
- All asymmetric dependence **measures utilized with ADD herein fall into category 2.**
- Note that numerical issues can render any **estimated** dependence measure non-positive definite, even if it is a category 1. measure, so empirical validation always must be employed in practice. Consequently, while empirical verification never should be taken on the same level as analytic proof, the need for empirical verification on a sample-by-sample basis makes the use of category 2. measures herein more palatable, or at least a more forgivable offence to the analytical purists among us.

## 2. Background: Directional Dependence Measures (housekeeping)

Housekeeping: All-pairwise matrix, positive definiteness, **symmetry enforced (asymmetric measures)**

- Note that within this all-pairwise matrix framework, for directional measures, **symmetry must be enforced**. The application of a directional measure IN ONE DIRECTION provides only a lower-triangle matrix. This matrix is made symmetric by placing the value in cell  $i,j$  into the zero of cell  $j,i$  and inserting unit diagonals.
- So herein, estimating two (all-pairwise) directional dependence matrices yields two symmetric matrices, one estimating the dependence measure in one direction (e.g. the effect of column  $j$  on row  $i$ ) and the other estimating the dependence measure of the same two variables in the opposite direction. The only difference procedurally is reversing the data column ordering before estimating the second matrix.
- (Note that all formulae below are estimators based on samples of data, as opposed to populations of data.)

## 2. Background: Directional Dependence Measures

**Chatterjee's new correlation coefficient** (Chatterjee (2021), **JASA**) ranges from zero to one asymptotically (it can exceed these bounds slightly under finite samples), and a value of zero does indicate independence. This is an asymmetric dependence measure, that is, the order of X and Y matters:  $\xi_n(X, Y)$  does not necessarily equal  $\xi_n(Y, X)$ , by design. In other words, the dependence of Y on X is not assumed to be identical to the dependence of X on Y: **dependence is directional**. Such measures are not new, in recent times going back over a dozen years (see Zheng et al. (2012), **JASA**), but they have direct origins in work done at the end of the 19th century (see Yule (1897)).

Rank order X and Y pairs according to X values as  $\left( (X_{(1)}, Y_{(1)}), (X_{(2)}, Y_{(2)}), \dots, (X_{(n)}, Y_{(n)}) \right)$ , then

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

Under ties of X values, break ties uniformly at random, then

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

Note that Chatterjee's can be made symmetric by simply taking the maximum of two Chatterjee tests, one in each direction (note that this is distinct from enforced symmetry of the **one-directional** matrix):

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

## 2. Background: Directional Dependence Measures

Chatterjee's breakthrough has spawned many variants, such as the “**improved Chatterjee's correlation**” derived by Xia et al. (2024), the motivation of which is to increase power by using inverse distance weightings of all neighboring data values as opposed to just one.

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

“Improved” Chatterjee's also is directional, with values ranging from zero (independence) to one (perfect dependence), and it also can be made symmetric using the maximum of two tests, one in each direction (note that this is distinct from enforced symmetry of the **one-directional** matrix):

$$5) \text{ ichcorr}_{\text{sym}} = \max \left[ \xi_n^{IM} (X, Y), \xi_n^{IM} (Y, X) \right]$$

## 2. Background: Directional Dependence Measures

Note that 4) typically would be coded with a nested loop, but this can be avoided by calculating the entire equation diagonally, 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.  $p \geq 100$ ) AND these matrices must be calculated many times in many simulations (e.g.  $N_{sim} \geq 10,000$ ).

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

```
*** INPUTS: data4chat is raw input returns data, reverse is a flag to reverse column order;
*** OUTPUT ichtat 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;
  ichtat=J(ncols,ncols,.);
  ichtat[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;
        ichtat[i,j]=1-cumsum/(cumwgts*(nrows+1)/3);
        ichtat[j,i]=ichtat[i,j];
      end;
    end;
  else do;
    print "Data provided to corr_ichat subroutine must be full rank.";
  end;
  return(ichtat);
finish;
```

## 2. Background: Directional Dependence Measures

Another Chatterjee variant is the measure of Zhang (2024a) that **combines Chatterjee's and Spearman's** 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).

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

This measure also ranges from zero (independence) to 1.0 (perfect dependence), and can be made symmetric using the maximum function (see Zhang, 2024b) (note that this is distinct from enforced symmetry of the **one-directional** matrix):

$$7) zcorrsp\_sym = \max \left\{ |sr_{X,Y}|, \sqrt{5/2} \xi_n(X, Y), \sqrt{5/2} \xi_n(Y, X) \right\}$$

## 2. Background: Directional Dependence Measures

For completeness, Spearman's Rho is presented below:

### Spearman's Rho

$$8) sr_{X,Y} = \frac{\sum_{i=1}^n \left( R_{X_i} - \frac{1}{n} \sum_{j=1}^n R_{X_j} \right) \left( R_{Y_i} - \frac{1}{n} \sum_{j=1}^n R_{Y_j} \right) / (n-1)}{\sqrt{\sum_{i=1}^n \left( R_{X_i} - \frac{1}{n} \sum_{j=1}^n R_{X_j} \right)^2 / (n-1)} \sqrt{\sum_{i=1}^n \left( R_{Y_i} - \frac{1}{n} \sum_{j=1}^n R_{Y_j} \right)^2 / (n-1)}}$$

Spearman's (Spearman (1904)) is exactly the same formula as Pearson's (Pearson (1895)) but instead of using the values of the returns of X and Y, their ranks are used instead. It's values also have a maximum range from -1 to 1. 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, the above can be shortened to 9), although 9) also can be adjusted for ties as in 10), 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.

$$9) sr_{X,Y} = 1 - \frac{6 \sum_{i=1}^n (R_{X_i} - R_{Y_i})^2}{n^3 - n}$$
$$10) sr_{X,Y} = \frac{\left( \left[ \frac{n^3 - n}{6} \right] - \sum_{i=1}^n (R_{X_i} - R_{Y_i})^2 - \sum_{i=1}^{mx} (t_{X_i}^3 - t_{X_i}) / 12 - \sum_{i=1}^{my} (t_{Y_i}^3 - t_{Y_i}) / 12 \right)}{\sqrt{\left[ \left( \left[ \frac{n^3 - n}{6} \right] - \sum_{i=1}^{mx} (t_{X_i}^3 - t_{X_i}) / 6 \right) \cdot \left( \left[ \frac{n^3 - n}{6} \right] - \sum_{i=1}^{my} (t_{Y_i}^3 - t_{Y_i}) / 6 \right) \right]}}$$

Computationally, 8) with averaged ranks is preferred to 10) (see Zar, 1999).

## 2. Background: Directional Dependence Measures

Another measure similar to Chatterjee's is the **Differential Distance Correlation** (DDC) of Liu and Shang (2025). DDC's values range from 0 (independence) to 1.0 (perfect dependence), and like 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 11) below:

$$11) DDC_n(X | Y) = 1 - \frac{1}{(n-1)} \sum_{i=1}^{n-1} \|X_{(i)} - X_{(i+1)}\| \bigg/ \left[ \binom{n}{2}^{-1} \sum_{i=1}^n (2i - n - 1) X^{(i)} \right]$$

where  $\left\{ (X_{(i)}, Y_{(i)}) \right\}_{i=1}^n$  are ordered to satisfy  $Y_{(i)} \leq \dots \leq Y_{(n)}$  and

$X^{(i)}$  are ordered to satisfy  $X_{(i)} \leq \dots \leq X_{(n)}$

DDC also can be made symmetric using the maximum function (see Liu, 2025) (note that this is distinct from enforced symmetry of the **one-directional** matrix):

$$12) DDC_{n-sym}(X, Y) = \max \left[ DDC_n(X | Y), DDC_n(Y | X) \right]$$

## 2. Background: Directional Dependence Measures

Another measure similar to Chatterjee's, for which values range from 0 (independence) to 1 (perfect dependence) is the **modified Chatterjee** of Yin (2026), with finite-sample estimator (for continuous variables) shown in (13):

$$13) mch = \xi_{+,n}(X, Y) = n \sum_{i=1}^n \sum_{j=1}^n \left( \frac{R_{i,j}}{n_j} - \frac{R_i}{n} \right)^2 \bigg/ \sum_{i=1}^n R_i (n - R_i) \quad \text{where}$$

$$R_i = \sum_{k=1}^n 1_{\{Y_k < Y_i\}}, \quad n_j = \sum_{k=1}^n K_h(X_k - X_j), \quad R_{i,j} = \sum_{k=1}^n 1_{\{Y_k < Y_i\}} K_h(X_k - X_j), \quad \text{and } K_h(t) = K\left(\frac{t}{h}\right) / h$$

is the Gaussian kernel with bandwidth = h (the simulation study uses the Silverman (1986) bandwidth of  $h = 1.06\hat{\sigma}_X n^{-1/5}$ ).

This carefully derived measure eliminates some of the arguable shortcomings of the other Chatterjee-related measures, such as finite sample distributions whose values extend below zero.

## 2. Background: Directional Dependence Measures

Directional dependence measures also can be applied only to the tails of X and Y, as correlation breakdowns often are associated specifically with (asymmetric) tail dependence: “Extensive evidence has been gathered showcasing the prevalence of ... asymmetric tail interdependence within equity and foreign exchange markets, particularly during times of crisis.” (Pramanik, 2024). And “We provide new evidence that lower tail dependence coefficients increased compared to upper ones for all pairs in the COVID-19 crash...” (Ito and Yoshiba, 2025). Below is the **asymmetric tail dependence** measure of Deidda et al. (2023), which is essentially Kendall’s Tau (see Kendall (1938)) applied conditionally, only when the percentile,  $q$ , of X (or Y) is exceeded:

$$14) \hat{\tau}_{X,Y}(q) = \left(2(k-2)!/k!\right) \sum_{1 \leq i \leq j \leq n} \text{sgn}(X_i - X_j) \text{sgn}(Y_i - Y_j) I(X_i, X_j > X_{(n-k)})$$

where  $q = 1 - k/n$ , and  $k \leq n$  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.

## 2. Background: Directional Dependence Measures

- Other directional / asymmetric dependence measures include the dynamic asymmetric tail dependence measure of Ito and Yoshioka (2025), the QAD measure of Junker et al. (2021), the generalized correlation of Zheng et al. (2012, **JASA**), the extremal directional dependence measure of Garcin and Nicolas (2026), the extreme risk network approach of Klüppelberg and Krali (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).
- We cite these measures for completeness, but do not focus on their implementation in this setting as they are more computationally involved and less readily implemented compared to (1) through (14) presented above.\*
- A takeaway here is the undeniable **richness of the current literature on directional dependence measures**. But it is notable that this history goes all the way back to the 19<sup>th</sup> century! (see Yule (1897), and a related review in Allena and McAleer (2018)).
- Before defining the Angles Space below, and showing how ADD can use it effectively with Directional Dependence Measures for Causal Discovery, it's important to plumb this history and examine its longstanding relationship with Causal Methods.

\* 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.

## 2. Background: Measures of Association & Causality - A Long History

### Relationship between Measures of Association and Causality

- While the development of directional dependence measures is not new, going back to the late nineteenth century (see Yule (1897)), neither is causal modeling, which goes back at least to Wright (1921). So the more recent utilization of the former by the latter should come as no surprise.
- Rodriguez Dominguez (2023, 2025a, 2025c) demonstrate that association-based structures, such as covariance and sensitivity measures, can be embedded within causal frameworks through structured representations of common drivers (much of this work follows the Reichenbach Common Cause Principle (Reichenbach, 1956), a foundational idea in probabilistic causality which emphasizes the identification of common causes through observed **correlations**).
- 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 (2021)), effectively 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, especially those that are asymmetric / directional.**

## 2. Background: Measures of Association & Causality - A Long History

### Relationship between Measures of Association and Causality

- As Moodie and Stephens (2022) find in their review, “...it is false ... to suggest that ordinary, conditional regression models *cannot* provide causal estimates: although there are some settings in which this is true, it is not universally the case.”
- Unfortunately, this is not always appreciated. 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, Lopez 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.
- 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, even in the face of (unverified) claims to the contrary.

## 2. Background: Measures of Association & Causality - A Long History

### Relationship between Measures of Association and Causality

- 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 causal discovery and DAG recovery by combining Directional Dependence Measures and the Angles Space, as described below.
- As Moodie and Stephens (2022) conclude, **“Causal inference should not be seen as conflicting with traditional statistical methods of explanatory analysis, and causal thinking should be considered relevant even in predictive modeling ...”**

## 2. Background: The Angles Space

- Foundational relationship between bivariate Pearson's correlation and angles: **cosine similarity**

$$\cos(\theta) = \frac{\text{inner product}}{\text{product of norms}} = \frac{\langle \mathbf{X}, \mathbf{Y} \rangle}{\|\mathbf{X}\| \|\mathbf{Y}\|} = \frac{\sum_{i=1}^N (X_i - E(X))(Y_i - E(Y))}{\sqrt{\sum_{i=1}^N (X_i - E(X))^2} \sqrt{\sum_{i=1}^N (Y_i - E(Y))^2}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \rho, \text{ with } 0 \leq \theta \leq \pi$$

- The matrix analogue is shown below:

## 2. Background: The Angles Space

### RELATIONSHIP between angles and correlations: Multivariate Bijection

$R = p \times p$  Pearson's matrix  $= BB^T$  where  $B$  is the Cholesky factor of  $R$

$$R = \begin{bmatrix} 1 & r_{1,2} & r_{1,3} & \cdots & r_{1,p} \\ r_{2,1} & 1 & r_{2,3} & \cdots & r_{2,p} \\ r_{3,1} & r_{3,2} & 1 & \cdots & r_{3,p} \\ r_{4,1} & r_{4,2} & r_{4,3} & \cdots & r_{4,p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{p,1} & r_{p,2} & r_{p,3} & \cdots & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \cos(\theta_{2,1}) & \sin(\theta_{2,1}) & 0 & \cdots & 0 \\ \cos(\theta_{3,1}) & \cos(\theta_{3,2})\sin(\theta_{3,1}) & \sin(\theta_{3,2})\sin(\theta_{3,1}) & \cdots & 0 \\ \cos(\theta_{4,1}) & \cos(\theta_{4,2})\sin(\theta_{4,1}) & \cos(\theta_{4,3})\sin(\theta_{4,2})\sin(\theta_{4,1}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \cos(\theta_{p,1}) & \cos(\theta_{p,2})\sin(\theta_{p,1}) & \cos(\theta_{p,3})\sin(\theta_{p,2})\sin(\theta_{p,1}) & \cdots & \prod_{k=1}^{n-1} \sin(\theta_{p,k}) \end{bmatrix}$$

for  $i > j$  angles  $\theta_{i,j} \in (0, \pi)$ .

So from  $R = BB^T$ , for an individual pairwise correlation we have

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

And recursively, in the other direction, for an individual angle  $\theta_{i,j}$  we have:

$$\text{For } j=1 \text{ and } 2 \leq i \leq p: \theta_{i,1} = \arccos(b_{i,1}); \text{ and for } 2 \leq j < i \leq p: \theta_{i,j} = \arccos\left(b_{i,j} / \prod_{k=1}^{j-1} \sin(\theta_{i,k})\right)$$

$$\text{Or equivalently: For } 1 \leq j < i \leq p: \theta_{i,j} = \text{atan2}\left(\sqrt{\sum_{k=j+1}^i b_{i,k}^2}, b_{i,j}\right)$$

## 2. Background: The Angles Space

- **We have a multivariate bijection here: a one-to-one, invertible, non-directional relationship between the matrix of angles and the matrix of correlations.** Transforming back and forth:

1. estimate the correlation/dependence measure matrix
2. obtain the Cholesky factorization of the matrix
3. Use inverse trigonometric functions on 2. to obtain corresponding spherical angles

And in reverse:

3. Start with a matrix of spherical angles
2. apply trigonometric functions to obtain the Cholesky factorization
1. multiply 2. by its transpose to obtain the corresponding correlation/dependence measure matrix

see Rapisarda et al. (2007) and Pourahmadi & Wang (2015) (note a typo in the formula in Pourahmadi & Wang (2015) for the first 3 steps where they mistakenly insert  $\prod_{k=1}^{j-1} \sin(\arccos(b_{i,k}))$  instead of  $\prod_{k=1}^{j-1} \sin(\theta_{i,k})$ . Lucchetti and Pedini (2025) also note this typo.

- **This bijection is true generally, for any positive definite matrix (not just Pearson's).** This transformation preserves probabilistic information, e.g. matrix-level acceptance/rejection regions, but not cell-level pdf's / cdf's, which require the determinant of the Jacobian.
- Note the **inverse relationship between angles and correlations**: correlations decrease monotonically in their corresponding angles, i.e. correlations increase as angles decrease to zero, and decrease as angles increase to  $\pi$ , or to  $\pi/2$  for directional measures (see Zhang et al. (2015) and Lu et al. (2019)). The range from 0 to  $\pi$  rather than 0 to  $2\pi$  is why this is the p-dimensional hyper-hemisphere rather than the hypersphere.

## 2. Background: The Angles Space

- Examples of dependence measure matrices and their corresponding Angles matrices:

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

|         |         |         |         |  |
|---------|---------|---------|---------|--|
|         |         |         |         |  |
| $\pi/2$ |         |         |         |  |
| $\pi/2$ | $\pi/2$ |         |         |  |
| $\pi/2$ | $\pi/2$ | $\pi/2$ |         |  |
| $\pi/2$ | $\pi/2$ | $\pi/2$ | $\pi/2$ |  |

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 |  |

- See APPENDIX C for examples of Angles densities and corresponding Spectral densities.

## 2. Background: The Angles Space

### Correlations to Angles

\* INPUT rand\_R is a valid correlation/dependence measure 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]=acos(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]=acos(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\_corr\_angles is a valid matrix of 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_corr_angles[i,gg]);
        end;
        Bs[i,j]=cos(rand_corr_angles[i,j])*sinprod;
      end;
    else do;
      sinprod=1;
      do gg=1 to (i-1);
        sinprod = sinprod*sin(rand_corr_angles[i,gg]);
      end;
      Bs[i,j]=sinprod;
    end;
  end;
end;
do;
  if i>1 then Bs[i,j]=cos(rand_corr_angles[i,j]);
  else Bs[i,j]=1;
end;
end;
end;
rand_R = Bs*T(Bs);
```

\* OUTPUT rand\_R is the corresponding dependence measure matrix;

## 2. Background: The Angles Space

- For completeness, we include below the definition of the Cholesky factor (see Benoit (1924)) and corresponding formulae:
- Given a real, symmetric positive definite dependence measure matrix  $R$ , 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:

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

- Conceptually, the Cholesky factorization can be thought of as the matrix analog to the square root of a scalar.
- Note, however, that **the Cholesky factor is a function of column ordering**, and this is addressed in APPENDIX B as it relates to ADD.

## 2. Background: The Angles Space

### Why bother with the Angles Space?

- ADD estimates 2 directional dependence measure matrices, one in each direction, translates them to the angles space, and makes statistical inferences within the angles space.\* Why bother?
- The angles space has many useful properties. It i. is **always positive definite**; ii. is generally more **robust** (see Lucchetti and Pedini (2023)), especially under heavy-tailed data (see Zhang and Yang (2025)); iii. allows for **unconstrained optimization** (see Zhang et al. (2015)) that even can be freely characterized via regressions as **functions of covariates**; iv. provides **better estimation** of such structures (see Pinheiro and Bates (1996) and Lucchetti and Pedini (2023)): per 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.”
- Consequently, reliance on spherical angles and hypersphere parameterizations is increasingly common in quantitative finance (see for some examples Li, Q., 2018; Hellton, 2020; Golts & Jones, 2009; Zhang, 2022; Saxena et al., 2023; and Zhang & Yang, 2023), in large part due to its **scale invariance**: it has even been used to define entire financial markets (see Kim and Lee, 2016).
- Importantly, because scale does not matter for measures of dependence, **the angles contain all extant information regarding dependence structure** (see Fernandez-Duren & Gregorio-Dominguez (2023), and Zhang & Yang (2023), as well as Opdyke (2022)).

\* 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 great and clear detail the relationship between correlations and angles.

## 2. Background: The Angles Space

### Why bother with the Angles Space?

- Although we need the determinant of the Jacobian to obtain the pdfs/cdfs of the individual Angles distribution(s), and this has not been derived in useable closed-form under general conditions, the fact that we have a multivariate bijection means that probabilistic information (e.g. matrix-level acceptance/rejection regions) is preserved in the translation. What's more, **inference under the null of independence can be defined nonparametrically** via straightforward and fast simulations (e.g. simply multivariate uniform data, as described below). And if we have the finite-sample analytic distribution of the dependence measure of choice under independence, as is often the case, we do not even need these simulations.
- Additionally, note that in the general case,\* dependence between the measure cells is retained upon translation to the angles matrix, but typically to a lesser degree. Still, this multivariate dependence may well contribute to **increased coherence** in ultimate edge-calling in the DAG.
- So **we use the angles space herein for simultaneous inference of all the pairwise relationships**. And when these inferences are based on directional dependence measures within the framework of dual orderings, **they will allow ADD to establish the direction of any such relationships**.
- Note that an important ancillary benefit of using the angles space here is its robustness and speed for **sample generation**, whereby probability integral transform (PIT) can be used, not only for specific cases\* (see Pourahmadi and Wang (2015) and Opdyke (2024a, 2026)), but also generally, using nonparametric kernel fits to the angles (see Opdyke (2024a, 2026)).

\* For the special case of Pearson's correlation under the identity matrix and Gaussian data, the angles become multivariate independent (see Pourahmadi and Wang (2025), Rapisarda et al. (2007), and Zhang et al. (2015)).

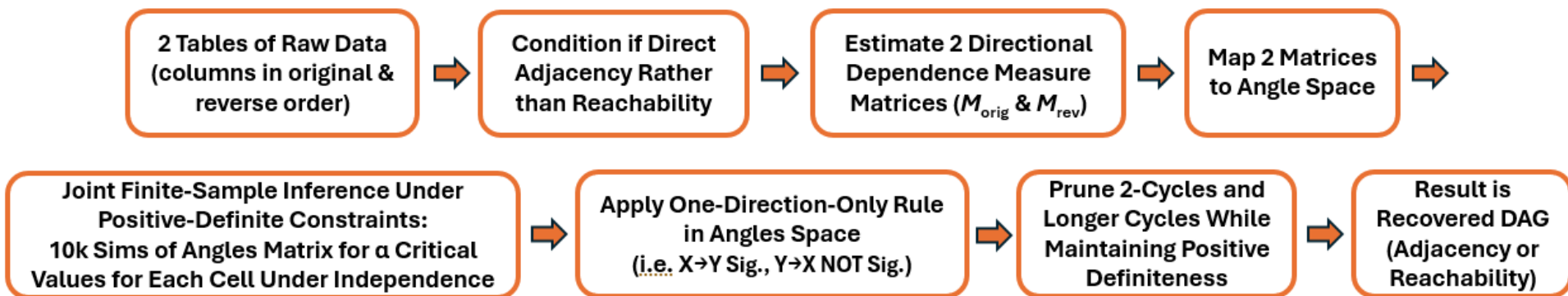
### 3. Method: Dual Orderings on Directional Measures in Angles Space

Angles-based Directional Dependence: “**ADD**” Execution Steps

1. **2 matrix estimations** of any Positive Definite **Directional** Dependence Measure, one in each direction (i.e. 1 original column ordering & 1 reverse column ordering ... covers all possible edges).
2. (**Apply conditioning** also if seeking direct adjacency rather than reachability)
3. **Angles Space**: Translate the 2 matrices into 2 Angles Matrices (a multivariate bijection).
4. **Dual Orderings / One-Direction-Only Rule**: Conduct inference within the Angles space, comparing each cell of the 2 Angles Matrices against distributions under the null of independence obtained via (fast) empirical simulations, or (finite-sample) analytic results if derived in the extant literature for that dependence measure; call an edge if statistically significant in **ONLY** one direction.
5. **DAG requirements**: Apply **pruning** to enforce acyclicity (under constraint of positive definiteness).

# 3. Method: Dual Orderings on Directional Measures in Angles Space

GRAPH 1: Pipeline of ADD's Approach to DAG Recovery



- The angles approach approach to DAG recovery is original: it calls all edges simultaneously, not sequentially, using directional statistical inference in the angles space wherein positive definiteness automatically is enforced.
- This provides potential advantages over competing algorithms which include:
  - a. More coherence in edge calls
  - b. Flexibility to use different (directional) dependence measures under different data conditions to increase power
- Both of these potential advantages will be tested against competing algorithms in a more extensive follow-up empirical study.
- For readable exposition here, I place all the necessary definitions and assumptions as they relate to valid causal interpretation in APPENDIX A (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).

### 3. Method: Dual Orderings on Directional Measures in Angles Space

#### “ADD” Execution Steps: Example

- Take, for example, **Chatterjee’s improved correlation** (“ICH”) coefficient (see Xia et al., 2024), which can be used twice, 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 dependent variable (Y) and relevant covariates (V1, V2, V3), we column-sort the data in one order (e.g. say, 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, with symmetry enforced, will together capture all potential associations, in both directions, of all the variables (factors).
- 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), two findings from this approach – one of a statistically significant, directional effect of X on Y ( $X \rightarrow Y$ ), but NO statistically significant, directional effect of Y on X ( $Y \rightarrow X$ ) – provide evidence of the directional, causal effect of X on Y ( $X \rightarrow Y$ ); the assumptions and definitions provided in APPENDIX A below subsequently allow for a fully causal interpretation, but I present the rationale for this approach here as follows.

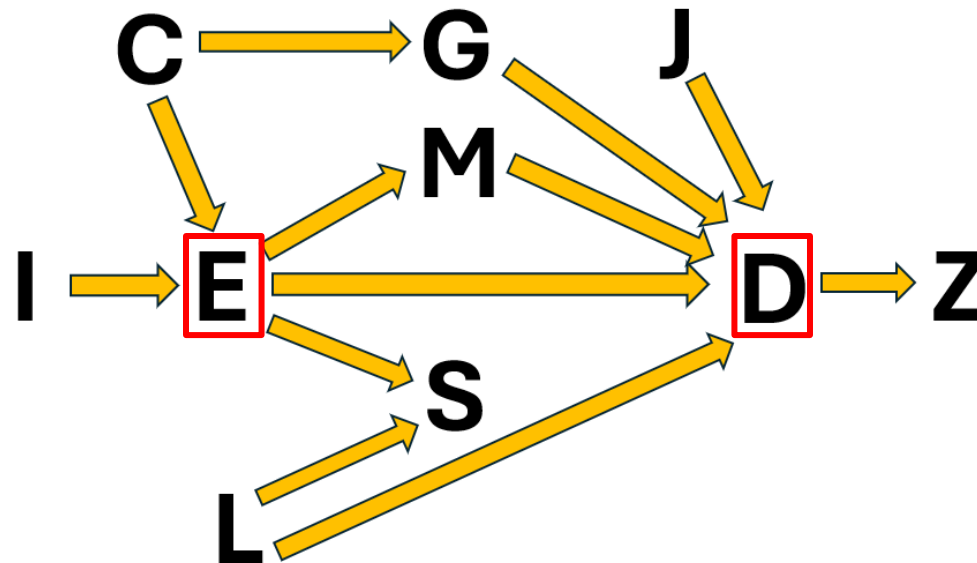
### 3. Method: Dual Orderings on Directional Measures in Angles Space

- RATIONALE: Consider a parent–child pair  $X \rightarrow Y$  with  $Y = f(X, U_Y)$ ,  $U_Y \perp X$ , and NOT  $Y \rightarrow X$ . **Under additive-noise mechanisms**, independence between  $X$  and  $U_Y$  implies that the conditional variability of  $Y$  given  $X$  is dominated by  $U_Y$ , whereas the inverse mapping  $X = g(Y) + \varepsilon$  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 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 measures. So we therefore have that directional functionals  $M$  that reward predictability under independent noise yield  $E[M(X \rightarrow Y)] > E[M(Y \rightarrow X)]$ .
- Now **for linear-Gaussian cases, this asymmetry collapses asymptotically as both directions are equivalent. But for nonlinear and/or non-Gaussian and/or heteroskedastic mechanisms, the asymmetry is restored.** And the angles approach 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 statistically significant and the other direction is not.
- Note that the angles approach is not designed to provide effect sizes, as do regression-based approaches (see for example Pascual-Marqui et al., 2024; Blömbaum et al., 2019; and MacKinnon & Lamp, 2022), nor does it provide counterfactual outcomes. Rather, **the angles approach identifies existing causal structure via DAG recovery.**

## 4. ADD: Preliminary Empirical Study

- For a **preliminary empirical study**, I use the straightforward, but non-toy DAG presented in Digitale et al. (2022) which contains confounders, colliders, and mediators, as well as both causes and consequences of the two primary factors of interest, E and D.

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

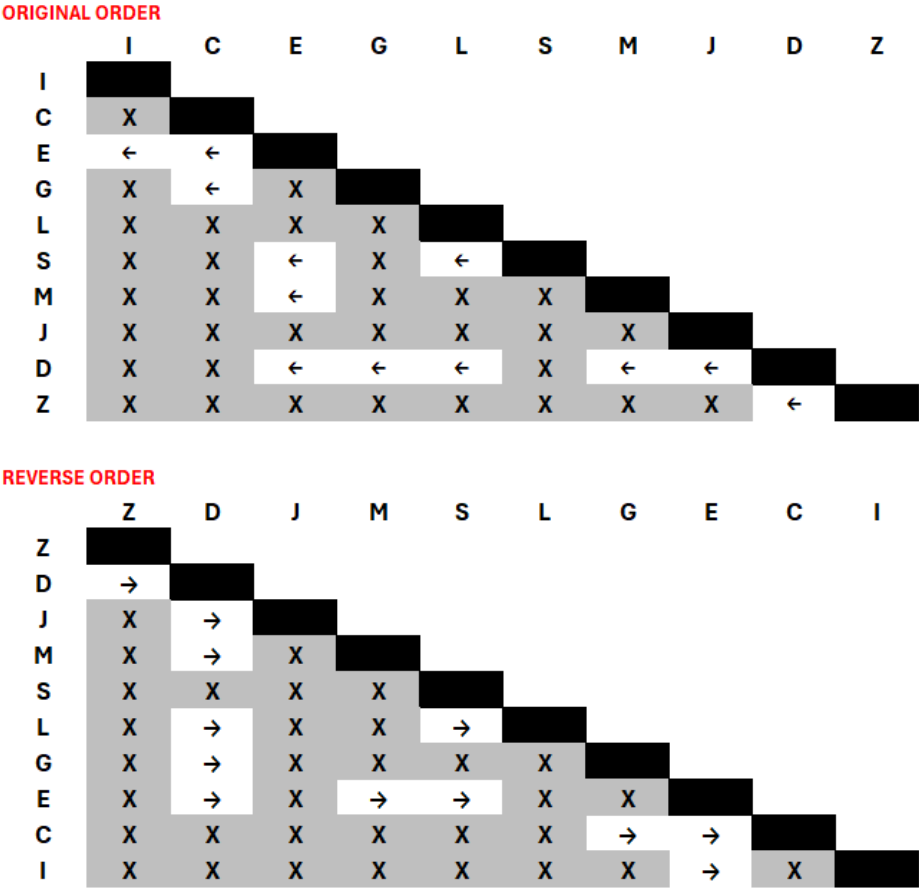


- There are multiple causal paths linking E and D ( $E \rightarrow D$ ,  $E \rightarrow M \rightarrow D$ ) as well as multiple non-causal paths linking E and D ( $E \leftarrow C \rightarrow G \rightarrow D$ ,  $E \rightarrow S \leftarrow L \rightarrow D$ ). Rightly identifying the directional relationships 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).

# 4. ADD: Preliminary Empirical Study

- The dimension of the corresponding all-pairwise matrix, where each cell represents an edge of the graph, is  $p = 10$  factors, yielding  $p(p-1)/2 = 45$  pairwise cells. The 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).

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



## 4. ADD: Preliminary Empirical Study

- I **generate data under the null hypothesis of independence**, to obtain the relevant (one-sided) confidence intervals for the statistical tests on the DAG-generated angles matrices, using multivariate uniform-distributed variables (only independence matters here; distributional changes make no empirical difference in the distributions of the angles). These simulations are very fast. If we have the finite-sample, analytical distribution of the selected dependence measure under the null of independence, as is often the case, this step is even faster. I **generate data based on the alternate hypothesis (the DAG)** via the code in Table C below.
- **For the DAG data, all parametric distributions are heavy-tailed (student's t-distributed), all error terms are asymmetric (lognormal distributed), and all functional relationships are highly nonlinear (most are sinusoidal, with some quadratic terms).** As under the null, every simulation has  $n = 100$  observations (which represents about five months of trading days).
- **Using reasonably small samples is conservative** here, making it more difficult for this approach to rightly identify directional effects, and no effects. Strong results under these conditions indicate good power of the estimator and the angles model approach. Note that other studies treating DAG recovery use sample sizes as large as  $n=1,000$ ,  $n=4,100$ , and  $n=5,000$  (see Olivetti et al. (2026) and <https://hub.crunchdao.com/competitions/causality-discovery>, Xue et al. (2025), and Petkov et al. (2025), respectively).
- Note that the purpose of the extensive use of coefficients in the DAG code in Table C is only to ensure that the additive effects of multiple parent variables (e.g. those of J, E, M, G, and L on D) are not orders of magnitude different in scale, thus swamping the effects of the other parent variables.

## 4. ADD: Preliminary Empirical Study

**TABLE C: Code Generating Data Under Alternate Hypothesis (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;
```

# 4. ADD: Preliminary Empirical Study

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

- This yields, over 10,000 simulations of the DAG, the average Pearson's, Spearman's, and Kendall's matrices shown in Table D. These show essentially no non-directional associations between the factors/variables.\*

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 |

\* 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).

## 4. ADD: Preliminary Empirical Study

- I choose a directional dependence measure that performs reasonably well under these data conditions: the improved Chatterjee correlation coefficient (“ICH”) of Xia et al. (2024). As mentioned above, 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 (the upper confidence bound is  $(1 - \alpha)$  rather than  $(1 - \alpha/2)$  because **this is a one-sided hypothesis test**) 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).\*
- $\alpha$  is set to 0.05, but is conservatively adjusted using a **Bonferroni adjustment ( $\alpha_{\text{adj}} = \alpha/2$ )** 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).\*\*

\* Recall that inference here is based on the angles corresponding to each cell of the matrix, so this is actually the lower  $\alpha$  bound because of the inverse 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  $\pi$  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).

\*\* Under the null hypothesis of independence these two tests will be independent, and under the alternate hypothesis these two tests will be, on average, negatively correlated. The Bonferroni adjustment (see Bonferroni (1936)) is used here not only because it is very conservative but also because it controls Familywise Error Rate (FWER) regardless of the direction of any correlation between these two tests. The effects of multiplicity across cells are ignored for purposes of this exercise.

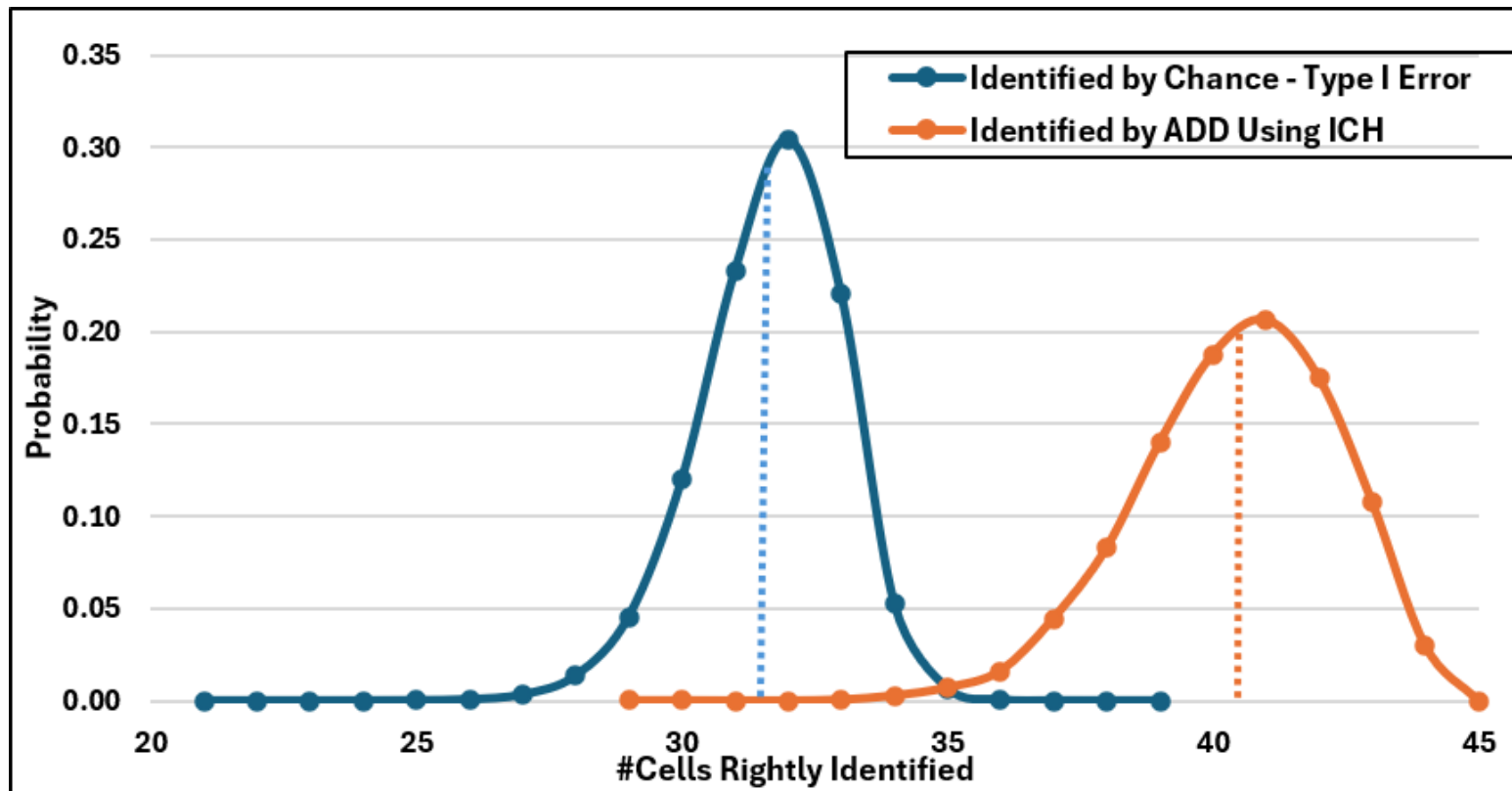
## 4. ADD: Preliminary Empirical Study

- 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.  $\Pr(A.) = \alpha_{\text{adj}} * (1 - \alpha_{\text{adj}}) = 0.025 * 0.975 = 0.024375$
  - B.  $\Pr(B.) = \alpha_{\text{adj}} * (1 - \alpha_{\text{adj}}) = 0.025 * 0.975 = 0.024375$
  - C.  $\Pr(C.) = (1 - \alpha_{\text{adj}}) * (1 - \alpha_{\text{adj}}) = 0.975 * 0.975 = 0.950625$
  - D.  $\Pr(D.) = (\alpha_{\text{adj}}) * (\alpha_{\text{adj}}) = 0.025 * 0.025 = 0.000625$
- The 2-cycles from D. are immediately pruned to preserve the acyclicity 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 A for detail). **All pruning is performed under the constraint of positive definiteness.**
- Once **acyclicity 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 A., and those 33 cells with no directional relationships are rightly chosen as 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.

## 4. ADD: Preliminary Empirical Study

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

**GRAPH 4: Distribution of #Rightly Identified Directional Effects/Non-Effects:  
ADD vs. No Model, All 45 Pairs (Fully Exploratory Analysis, No Priors)**



## 4. ADD: Preliminary Empirical Study

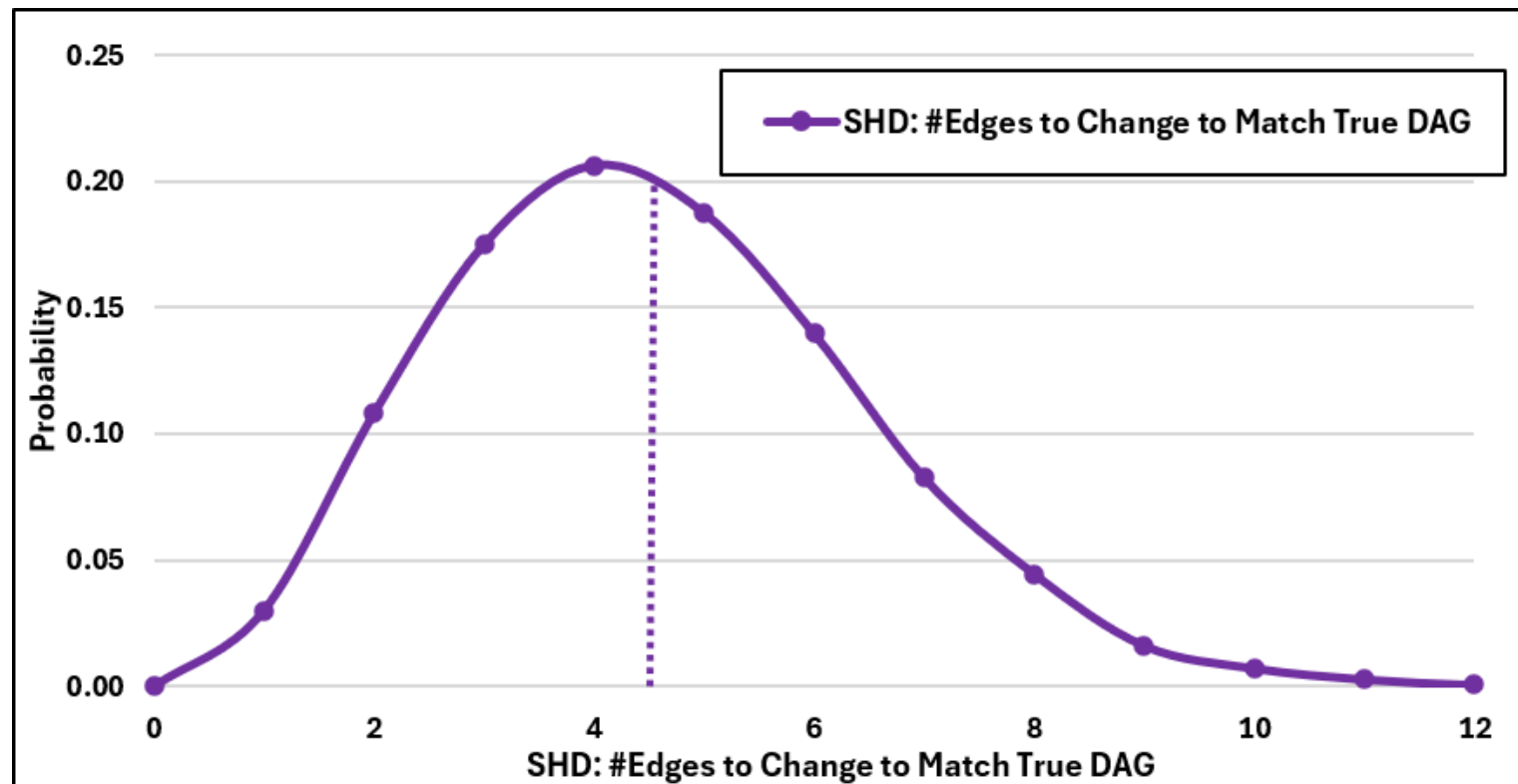
- ICH is applied to the data with the factors in original and reverse orders, per Graph 3, and the resulting two matrices of ICH values are saved. The rows and columns of the reverse order matrix are reversed so that the two matrices can be compared easily, cell for cell.
- If, for a given cell, the result of the first test is statistically significant, but the result of the second test is 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, acyclicity 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 40.44, compared to that under the null which is 31.66.
- **The difference between the two distributions is material and highly statistically significant, with  $p\text{-value} < 0.00001$ .**\*

\* This is a one-sided p-value based on a permutation test of 100,000 samples with sample sizes of  $n_1=n_2=45$  (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.

## 4. ADD: Preliminary Empirical Study

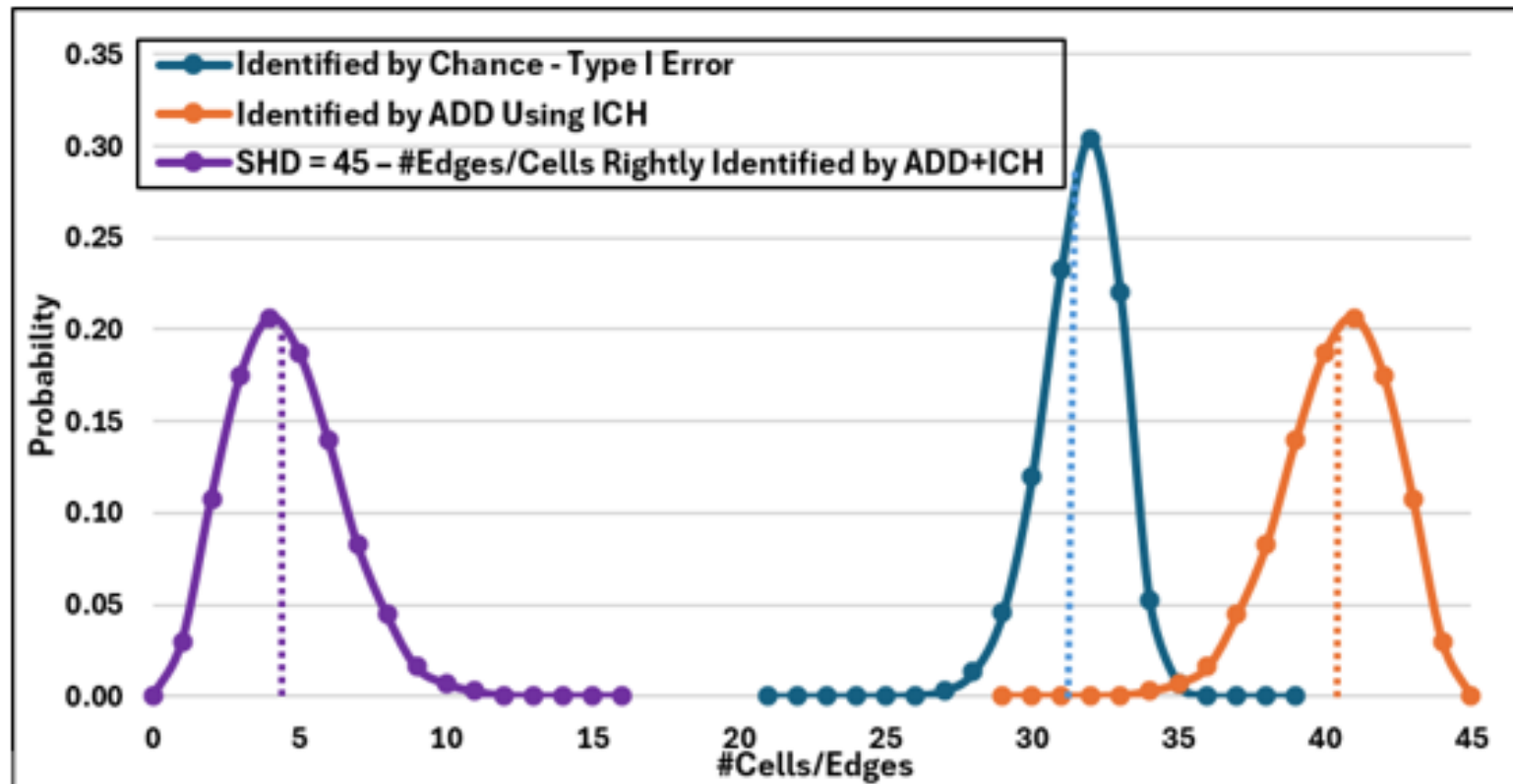
- Note that this **#right metric simply is the inverse of the Structural Hamming Distance (SHD)** (see Acid and deCampos (2003)), that is, the number of edge modifications that need to be changed for an estimated DAG to match the true DAG. ADD's corresponding SHDs across the 10,000 simulations, with a mean of 4.56, are shown in Graph 4a. I combine Graphs 4 and 4a in Graph 4b below to put them on the same scale.

**GRAPH 4a: ADD's Distribution of SHD: #Edges Changed to Match True DAG**



## 4. ADD: Preliminary Empirical Study

GRAPH 4b: ADD's Distribution of #Rightly Identified Directional Cells/Edges vs. SHD and No Model



- In addition to SHD, I present ADD's confusion / error matrix below in Table E, along with various common model performance metrics based on it. These values all indicate **material predictive power of ADD under these conditions**.

# 4. ADD: Preliminary Empirical Study

TABLE E: ADD’s Error Matrix, and Associated Performance Metrics

| Error Matrix - Numbers |            |            |         |
|------------------------|------------|------------|---------|
|                        | Predicted  |            | TOTAL   |
|                        | Right      | Not Right  |         |
|                        | (Edge+Dir) | (Edge+Dir) |         |
| True Edge              | 80,949     | 39,051     | 120,000 |
| Non-Edge               | 6,552      | 323,448    | 330,000 |

| Error Matrix - Rates |            |            |       |
|----------------------|------------|------------|-------|
|                      | Predicted  |            | TOTAL |
|                      | Right      | Not Right  |       |
|                      | (Edge+Dir) | (Edge+Dir) |       |
| True Edge            | 67.5%      | 32.5%      | 100%  |
| Non-Edge             | 2.0%       | 98.0%      | 100%  |

|             |       |          |       |                       |       |
|-------------|-------|----------|-------|-----------------------|-------|
| Precision = | 0.925 | MCC =    | 0.732 | Jaccard =             | 0.640 |
| Recall =    | 0.675 | F1 =     | 0.780 | Prevalence Threshold= | 0.146 |
| Accuracy =  | 0.899 | Youden = | 0.655 | Balanced Accuracy =   | 0.827 |

# 4. ADD: Preliminary Empirical Study

- Table F presents ADD's %right classifications, across all simulations, for each specific cell/edge. Red font indicates lower relative scores on the %right metric. For the effect of M on D the red 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**.

TABLE F: ADD's %Right Relationships Identified in Each Cell / Edge

|   | I     | C     | E     | G     | L     | S | M     | J     | D     | Z |
|---|-------|-------|-------|-------|-------|---|-------|-------|-------|---|
| I |       |       |       |       |       |   |       |       |       |   |
| C |       |       |       |       |       |   |       |       |       |   |
| E | 85.4% | 85.2% |       |       |       |   |       |       |       |   |
| G |       | 98.1% |       |       |       |   |       |       |       |   |
| L |       |       |       |       |       |   |       |       |       |   |
| S |       |       | 71.7% |       | 72.4% |   |       |       |       |   |
| M |       |       | 72.1% |       |       |   |       |       |       |   |
| J |       |       |       |       |       |   |       |       |       |   |
| D |       |       | 12.5% | 60.0% | 55.4% |   | 30.6% | 70.3% |       |   |
| Z |       |       |       |       |       |   |       |       | 95.9% |   |

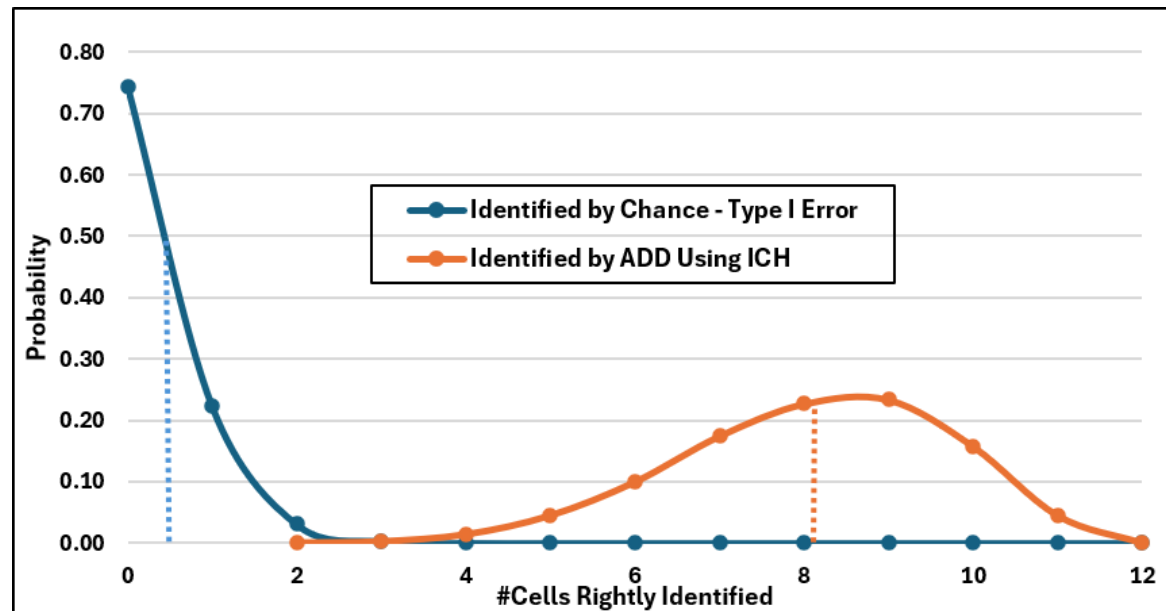
  

|   | I     | C     | E     | G     | L     | S     | M     | J     | D | Z |
|---|-------|-------|-------|-------|-------|-------|-------|-------|---|---|
| I |       |       |       |       |       |       |       |       |   |   |
| C | 98.3% |       |       |       |       |       |       |       |   |   |
| E |       |       |       |       |       |       |       |       |   |   |
| G | 95.9% |       | 98.0% |       |       |       |       |       |   |   |
| L | 96.2% | 96.3% | 97.2% | 98.5% |       |       |       |       |   |   |
| S | 99.1% | 99.0% |       | 97.0% |       |       |       |       |   |   |
| M | 99.8% | 99.8% |       | 97.6% | 98.2% | 99.0% |       |       |   |   |
| J | 95.8% | 96.2% | 98.0% | 96.9% | 98.7% | 98.0% | 98.9% |       |   |   |
| D | 98.1% | 99.3% |       |       |       | 99.3% |       |       |   |   |
| Z | 96.7% | 97.9% | 98.3% | 98.4% | 98.3% | 97.7% | 99.3% | 99.1% |   |   |

## 4. ADD: Preliminary Empirical Study

- If we look only at the 12 cells with the directional relationships, ADD's results remain notable. Under the null 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, 8.09. **This material difference also is highly statistically significant**, with p-value < 0.00001.\* The angle model's %right value is 67.5% here (note that this is just recall, the percent of "positives" rightly identified), while that under the null is approximately 2.0%, which is  $\approx \Pr(A.)$ .

Graph 5: ADD's Distribution of #Rightly Identified Directional Effects for 12 Effect Pairs v No Model

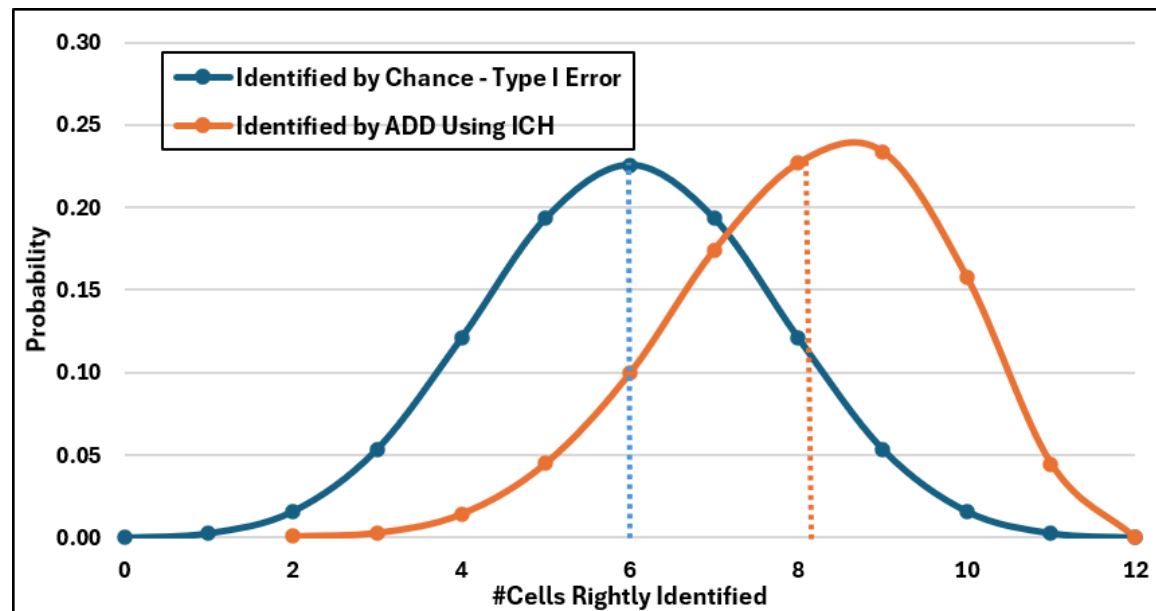


\* For this permutation test, sample sizes  $n1 = n2 = 12$ , and 100,000 simulations were run.

## 4. ADD: Preliminary Empirical Study

- Even if we gave a researcher an unfair advantage over the angles-based model and he or she somehow KNEW that the 12 cells were associated with an effect (i.e. he or she had 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 approximately 0.01.\* All of these results are summarized below in Table G.

**Graph 6: ADD's Distribution of #Rightly Identified Directional Effects for 12 Effect Pairs vs. No Model with Perfect Priors on which 12 Pairs Have SOME Effect (but not the Direction)**



\* For this permutation test, sample sizes  $n_1 = n_2 = 12$ , and 100,000 simulations were run.

## 4. ADD: Preliminary Empirical Study

**TABLE G: DAG Recovery, ADD-ICH vs. No Model**

All 45 Pairs, Fully Exploratory Analysis, No Priors

| no model*    |              | ADD-ICH      |              | p-value*           |
|--------------|--------------|--------------|--------------|--------------------|
| # right      | % right      | # right      | % right      |                    |
| <b>31.66</b> | <b>70.4%</b> | <b>40.44</b> | <b>89.9%</b> | <b>&lt;0.00001</b> |

\*Binomial vector:  $p_{12} = \alpha_{\text{adj}}(1 - \alpha_{\text{adj}})$ ,  $p_{33} = p = [1 - \alpha_{\text{adj}}]^2$

\*p-value: permutation test, nsims=100k,  $n_1 = n_2 = 45$

12 Pairs with Effects, Fully Exploratory Analysis, No Priors

| no model**  |             | ADD-ICH     |              | p-value**          |
|-------------|-------------|-------------|--------------|--------------------|
| # right     | % right     | # right     | % right      |                    |
| <b>0.29</b> | <b>2.4%</b> | <b>8.09</b> | <b>67.5%</b> | <b>&lt;0.00001</b> |

\*\*Binomial distribution,  $p_{12} = \alpha_{\text{adj}}(1 - \alpha_{\text{adj}})$

\*\*p-value: permutation test, nsims=100k,  $n_1 = n_2 = 12$

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

| no model~   |              | ADD-ICH     |              | p-value~       |
|-------------|--------------|-------------|--------------|----------------|
| # right     | % right      | # right     | % right      |                |
| <b>6.00</b> | <b>50.0%</b> | <b>8.09</b> | <b>67.5%</b> | <b>0.01174</b> |

~Binomial distribution,  $p_{12} = 0.5$

~p-value: permutation test, nsims=100k,  $n_1 = n_2 = 12$

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

| no model~~   |              | ADD-ICH      |              |
|--------------|--------------|--------------|--------------|
| # right      | % right      | # right      | % right      |
| <b>31.37</b> | <b>95.1%</b> | <b>32.34</b> | <b>98.0%</b> |

~~Binomial distribution,  $p_{33} = [1 - \alpha_{\text{adj}}]^2$

## 4. ADD: Preliminary Empirical Study

- This is a relatively narrow, preliminary study, but **the empirical results are promising**.
- **The study conditions are reasonably challenging**: all bivariate relationships are highly nonlinear, with heavy-tailed distributions and asymmetric error terms, **yet ADD exhibits predictive power** even when sample sizes are relatively small ( $n=100$ ).
- **ADD's simultaneous, as opposed to sequential or iterative, estimation of edge calls, within the positive definite angles space, may very well contribute to greater coherence in DAG recovery**. 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.
- ADD's results will vary depending on the (directional) dependence measure used, but this is a strength rather than a weakness of this approach: ADD is adaptable to different and varying data conditions, under which some measures will perform better than others. So **having the flexibility to test and verify data conditions ex ante could lead to notable power gains**.
- Finally, the requirement of only two matrix estimations, and a (fast) simulation (in the absence of measure-specific analytic pdf derivations) to establish confidence bounds under independence, appears to offer ADD relative **speed and scalability**.
- This all will be explicitly addressed in a more extensive follow-up empirical study that will include a wide range of sample sizes, many different data types with varying signal-to-noise ratios, and results from competitor algorithms designed for identifying causal structure.

## 5. Conclusions and Next Steps

- We end on a note of caution by recognizing that this dive into (directional) causality and 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)).
- Even beyond quantitative finance, across a broad range of settings in fact, many researchers have expressed strong caution regarding the limitations of DAGs (see de Lara, (2023), Gong et al., (2024), Padh et al. (2025), and Kaiser and Sipos (2022)).\*
- But **alongside the new potential benefits of ADD, progress continues to be made here on frameworks that augment DAGs**, or at least are not inconsistent with DAGs.

\* 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.

## 5. Conclusions and Next Steps

- Recent innovative work on **geometric causal models** is providing both broader and more flexible structural frameworks for causal discovery (see Weinstein and Blei (2026) and Usta and Bollt (2024), respectively).
- Regarding the critical issue of **temporality**, 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 enables 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) embeds 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.

## Contact Information:

**JD Opdyke**

**Chief Analytics Officer, DataMineit, LLC**

**[JDOpdyke@DataMineit.com](mailto:JDOpdyke@DataMineit.com)**

**[JDOpdyke@gmail.com](mailto:JDOpdyke@gmail.com)**



# 6. References

- Acid, S., and deCampos, L.M., (2003), "Searching for Bayesian network structures in the space of restricted acyclic partially directed graphs," *Journal of Artificial Intelligence Research*, 18, 445–490.
- 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," *Journal of Machine Learning Research*, 11(7):171–234.
- Allena, D., and McAleer, M., (2018), "Generalized Measures of Correlation for Asymmetry, Nonlinearity, and Beyond: Comment," ICAE – Instituto Complutense de Análisis Económico, Working Paper 1823.
- Benoit, (1924), Note Sur Une Méthode de Résolution des équations Normales Provenant de L'Application de la Méthode des Moindres Carrés à un Système d'Équations Linéaires en Nombre Inférieur à Celui des Inconnues. — Application de la Méthode à la Résolution d'un Système Défini D'équations Linéaires. *Bull. Géodésique* 2, 67–77 (1924). <https://doi.org/10.1007/BF03031308>
- Blömbaum, P., Janzing, D., Washio, T., Shimizu, S., and Schölkopf, B., (2019), "Analysis of Cause-Effect Inference by Comparing Regression Errors," *PeerJ Computer Science*, 5:e169.
- 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.
- Cardin, M., (2009), Multivariate Measures of Positive Definiteness," *International Journal of Contemporary Mathematical Sciences*, 4(4), 191–200.
- 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.
- Chiriac, R., and Voelv, V., (2010), "Modelling and Forecasting Multivariate Realized Volatility," *Journal of Applied Econometrics*, 26(6), 922–947.
- Church, Christ (2012). "The asymmetric t-copula with individual degrees of freedom", Oxford, UK: University of Oxford Master Thesis, 2012.
- Cüppers, J., Xu, S., Musa, A., and Vreeken, J., (2024), "Causal Discovery from Event Sequences by Local Cause-Effect Attribution," *38<sup>th</sup> Conference on Neural Information Processing Systems (NeurIPS'24)*.
- De Lara, L., (2024), "On the (in)compatibility between potential outcomes and structural causal models and its signification in counterfactual inference," arXiv:2309.05997v3 [math.ST].
- Digitale, J., Martin, J., and Glymour, M., (2022), "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].
- Fernandez-Duran, J.J., and Gregorio-Dominguez, M.M., (2023), "Testing the Regular Variation Model for Multivariate Extremes with Flexible Circular and Spherical Distributions," arXiv:2309.04948v2.
- Garcin, M., and Nicolas, M., (2026), "Directional Dependence of Extreme Events," working paper, arXiv:2604.03215v1 [stat.ME].
- Gong, H., Lu, C., and Zang, Y., (2024), "Distribution-consistency Structural Causal Models" arXiv:2401.15911v2 [cs.AI]
- Golts, M. and Jones, G., (2009), "A Sharper Angle on Optimization," <https://ssrn.com/abstract=1483412>.
- Hauser, A., and Bühlmann, P., (2012), "Characterization and Greedy Learning of Interventional Markov Equivalence Classes of Directed Acyclic Graphs," *J. 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.
- Itaya, Y., Tamura, J., Hayashi, K., and Yamamoto, K., (2024), "Asymptotic Properties of Matthews Correlation Coefficient," *Statistics in Medicine*, 44(1–2), e10303.
- Ito, K., and Yoshida, 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.
- 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.
- Kahn, A.B., (1962), "Topological Sorting of Large Networks," *Communications of the ACM*, 5(11), 558–562.
- 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.
- Kendall, M. (1938), "A New Measure of Rank Correlation," *Biometrika*, 30 (1–2), 81–89.
- Kim, W., and Lee, Y., (2016), "A Uniformly Distributed Random Portfolio," *Quantitative Finance*, Vol. 16, No. 2, pp.297–307.
- KlÜppelberg, C., and Krali, M., (2026), "Causal Analysis of Extreme Risk in a Network of Industry Portfolios," *Canadian Journal of Statistics*, 54(1).
- Li, X., and Joe, H., (2024), "Multivariate Directional Tail-weighted Dependence Measures," *Journal of Multivariate Analysis*, Vol 203.
- 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.
- Liu, Y., and Shang, P., (2025), "Differential Distance Correlation and Its Applications," *Journal of Multivariate Analysis*, Vol. 214, Issue C.
- Liu, Y., (2025), email correspondence with author Yixiao Liu, July 9, 2025.
- Lopez de Prado, M., (2023), *Causal Factor Investing*, Cambridge University Press.
- Lu, F., Xue, L., and Wang, Z., (2019), "Triangular Angles Parameterization for the Correlation Matrix of Bivariate Longitudinal Data," *J. of the Korean Statistical Society*, 49:364–388.
- Lucchetti, R., and Pedini, L. (2023), "The Spherical Parametrisation for Correlation Matrices and its Computational Advantages," *Computational Economics*, 64, 1023–1046.
- Lucchetti, R., and Pedini, L. (2025), "Correction to: The Spherical Parametrisation for Correlation Matrices and its Computational Advantages," *Computational Economics*, 65, 2449–2450.
- 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.
- Moodie, E., and Stephens, D., (2022), "Causal Inference: Critical Developments, Past and Future," *The Canadian Journal of Statistics*, 50(4), 1200–1320.
- Olivetti, E., Zoonekynd, V., Yam, P., Lopez de Prado, M., Imbens, G., and Hernan, M., (2026), "Can AI Learn Causal Structure? Evidence from ADIA Lab's Causal Discovery Challenge," <https://ssrn.com/abstract=6125566>.

# 6. References

- 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](http://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](http://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](http://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].
- Pearson, K., (1895), "VII. Note on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, 58: 240-242.
- Petkov, H., MacLellan, C., and Dong, F., (2025), "DAGAF: A Directed Acyclic Generative Adversarial Framework for Joint Structure Learning and Tabular Data Synthesis," *Applied Intelligence*, 55:602.
- 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., 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 Assets 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, Alejandro, (2025c), "Causal PDE-Control Models for Dynamic Portfolio Optimization with Latent Drivers," <https://ssrn.com/abstract=5747083>.
- 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.
- Silverman, B.W., (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.
- Spearman, C., (1904), "General Intelligence, Objectively Determined and Measured," *The American Journal of Psychology*, 15(2), 201-292.
- 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.
- Stollenwerk, M., (2026), "Probability Distributions for Realized Covariance Measures," *Journal of Econometrics*, Vol. 254(A).
- Szekely, G., Rizzo, M., and Bakirov, N., (2007), "Measuring and Testing Dependence by Correlation of Distances," *The Annals of Statistics*, 35(6), pp2769-2794.
- Usta, O., and Bollt, E., (2024), "Fractal Conditional Correlation Dimension Infers Complex Causal Networks," arXiv:2411.19284 [cs.IT].
- 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.
- Weinstein, E., and Blei, D., (2026), "Geometric Causal Models," arXiv:2607.05153v1 [stat.ML].
- 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., Sankararaman, S., and Pimentel, H., (2025), "dotears: Scalable and Consistent Directed Acyclic Graph Estimation Using Observational and Interventional Data," *iScience*, 28, 111673.
- Yin, C., (2026), "A Modified Dependence Measure Related to Chatterjee's Rank Correlation: Theoretical Properties and Asymptotic Analysis," arXiv:2608.07844v1 [math.ST].
- 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., Ozkirimli, E., and Stella, F., (2025), "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, 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, Y., (2022), "Angle Based Dependence Measures in Metric Space," arXiv:2206.01459v1 [stat.ME].
- 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].
- Zhang, Y., and Yang, S., (2025), "Kernel Angle Dependence Measures in Metric Spaces," *Journal of Computational Statistics*, 34(1), 280-289.
- 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 A: ADD's Assumptions for Valid Causal Interpretation

**A.** Below we 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.

**Preliminaries:** Structural Causal Model (SCM) and do-Semantics (see Pearl, 2009)

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

$$X_j = f_j(X_{\text{Pa}(j)}, U_j) \text{ for } j = 1, \dots, p$$

Here  $\text{Pa}(j)$  are the parents of  $X_j$  in a DAG “G,”  $U_1, \dots, U_p$  are jointly independent exogenous noises, and each  $f_j(\cdot)$  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  $\text{do}(X=x)$  replaces its structural equation with the constant x, severing incoming edges into X.

**Causal effect:** X has a (possibly context-dependent) causal effect on Y if and only if there exist  $x \neq x'$  such that  $P(Y \mid \text{do}(X=x)) \neq P(Y \mid \text{do}(X=x'))$ .

## **Assumptions for Causal Interpretation:**

(A1) **Acyclicity:** The true causal graph over V is a DAG.

(A2) **Causal Sufficiency** (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) **Markov + (qualified) faithfulness:** 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) **i.i.d. sampling (or weak stationarity for time-series):** Samples are independent draws from the same distribution (future research plans to apply ADD, for time-series extensions, to innovations or appropriate lags).

(A5) **Measurement reliability:** Variables are measured without systematic differential error that inverts directional asymmetries.

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

(A7) **Monotone link from mechanism to asymmetry:** 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  $X \in \text{Pa}(Y)$

## 7. Appendix A: ADD's Assumptions for Valid Causal Interpretation

and  $Y \notin \text{Pa}(X)$ , then asymptotically  $E[ M(X \rightarrow Y) ] > E[ M(Y \rightarrow X) ]$ , 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  $M$  increases 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  $E[ M(X \rightarrow Y) ] > E[ M(Y \rightarrow X) ]$  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  $M(X \rightarrow Y | V \setminus \{X, Y\})$ : declare a directed edge  $X \rightarrow Y$  if statistically significant( $X \rightarrow Y$ ) and NOT statistically significant( $Y \rightarrow X$ ) at level  $\alpha$  (the one-direction-only rule). This is interpreted as  $X \in \text{Pa}(Y)$ .

Ancestral/path claim (directed reachability): If only marginal  $M$  is used, interpret the same rule as: there exists a directed path  $X \rightsquigarrow Y$  (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 both directional tests are significant, yielding a 2-cycle, pruning enforces acyclicity by selecting the edge/direction that has a larger p-value and replacing its angle value with a value just below what would indicate statistical significance at the  $\alpha$  threshold. 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 convex 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

# 7. Appendix A: ADD's Assumptions for Valid Causal Interpretation

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  $X \rightarrow Y$ , then  $X \in \text{Pa}(Y)$  with probability  $\rightarrow 1$  as  $n \rightarrow \infty$ .

By Markov + faithfulness (A3) and SCM acyclicity (A1), conditioning on  $V \setminus \{X, Y\}$  blocks all non-causal paths between  $X$  and  $Y$ . If  $X \in \text{Pa}(Y)$ , then by (A7) the population asymmetry satisfies  $M(X \rightarrow Y) > M(Y \rightarrow X)$ ; ADD's finite-sample tests are consistent in  $n$  (angles-based inference), so the probability of correctly calling  $X \rightarrow Y$  tends to 1. Conversely, if no directed edge  $X \rightarrow Y$  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  $\rightarrow 1$ .

Note that when  $M$  is conditional on other variables, an edge  $X \rightarrow Y$  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, conditioning was used, allowing for the former interpretation.

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

**B.** Below we place ADD within the Causal Inference Literature

## **Proposition:**

Let  $\mathcal{V} = \{X_1, \dots, X_p\}$  be generated by acyclic structural equations  $X_j = f_j(X_{\text{Pa}(j)}, U_j)$  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.” We 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  $V \setminus \{X, Y\}$ , 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 Spirtes (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

## 7. Appendix A: ADD's Assumptions for Valid Causal Interpretation

sequentially. This will be tested in a more extensive, follow-up empirical study pitting ADD against many of the above-mentioned competitors.

## 7. Appendix B: Column Ordering and the Cholesky Factor

- As noted above, as a mathematical issue, **the values of the Cholesky factor do depend on the order of the columns of the underlying data. However, in this setting** (i.e. all-pairwise dependence measure matrices that are unit diagonal, positive definite, and not extremely sparse or on scales that differ by orders of magnitude) **this appears to be more of a computational/calculation issue than a statistical one.**
- The conditions under which the Cholesky factor's values will change materially based on changing column order include: the pathological case of the dependence measure matrix being very nearly singular (e.g. nearly perfectly collinear variables), the matrix being extremely sparse (i.e. the vast majority of cells are zeros), and/or the matrix contains values that are highly non-uniform (e.g. with large clusters of values at or very near zero, and large clusters of values at or very near one). These data conditions are relatively extreme and readily observed, so if the degree to which they exist is a concern, the potential materiality of **the effects of column ordering can be checked empirically.**

## 7. Appendix B: Column Ordering and the Cholesky Factor

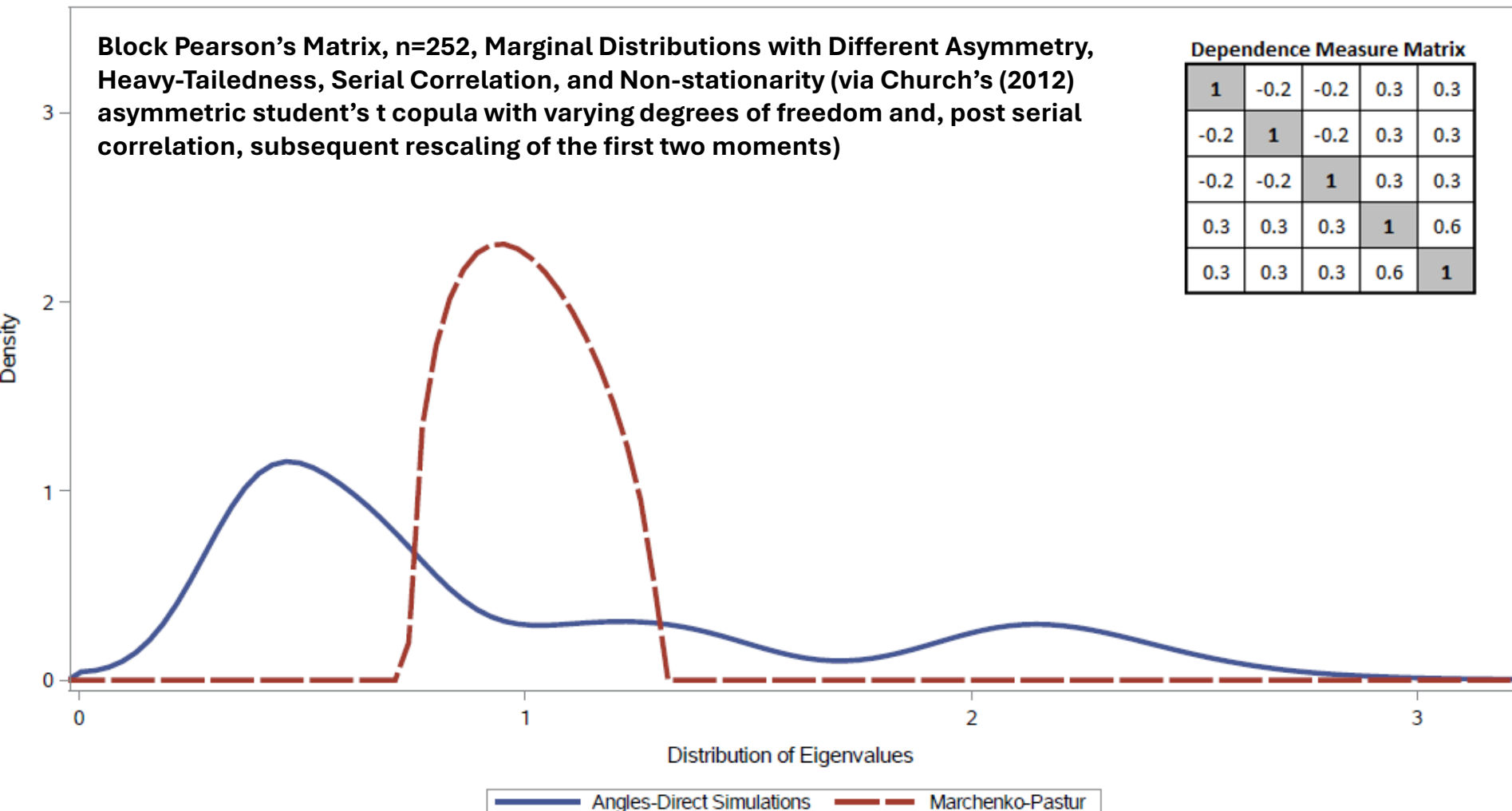
- **We perform such a check based on the empirical study performed in the paper, randomly re-ordering the columns of the input data variables ten different ways (and changing nothing else):** we find that, compared to the original findings from Table D, the average values of true positives, true negatives, and MCC are virtually identical, while their standard deviations across all 10 runs all are smaller than one percent (see table below).

|                         | True Positives | True Negatives | MCC    |
|-------------------------|----------------|----------------|--------|
| Original                | 67.46%         | 98.01%         | 73.15% |
| Mean: 10 Column Orders  | 67.72%         | 98.05%         | 73.43% |
| Stdev: 10 Column Orders | 0.88%          | 0.17%          | 0.98%  |

- In addition to its small absolute size, this result for MCC (0.0098) compares reasonably well against the approximate standard error of the MCC estimator (0.0011) based on its estimated 95% confidence interval (see Itaya et al. (2024), eq. (7)).
- For settings where the conditions listed above (e.g. extreme sparsity, nearly perfect collinearity, etc.) are not uncommon, this would not be an unreasonable check to perform when implementing ADD. But empirically here, even under this study's reasonably challenging data conditions, **the effects of column-ordering are not material**, and it is worth noting that **two other studies confirm this empirical result of non-materiality vis- à-vis column ordering** (see Stollenwerk (2026) and Chiriac and Voev (2010)).

# 7. Appendix C: Example of Spectral and Angles Distributions

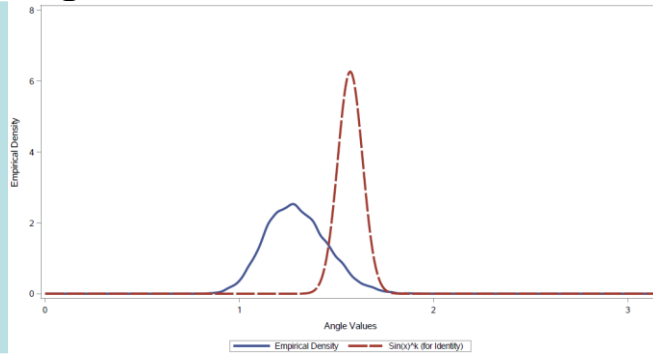
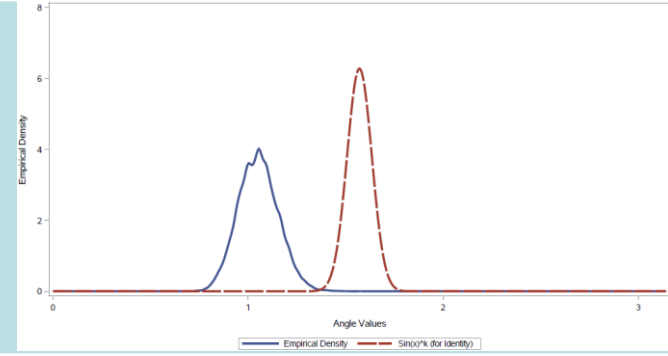
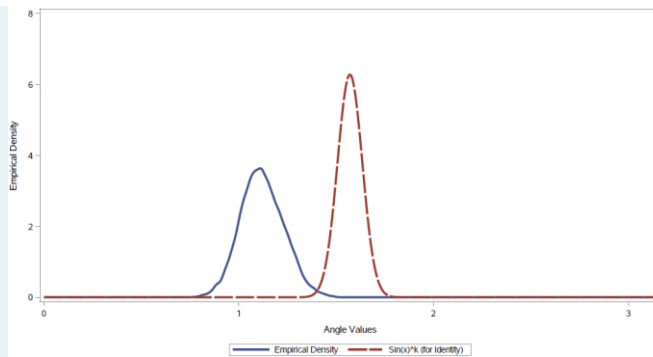
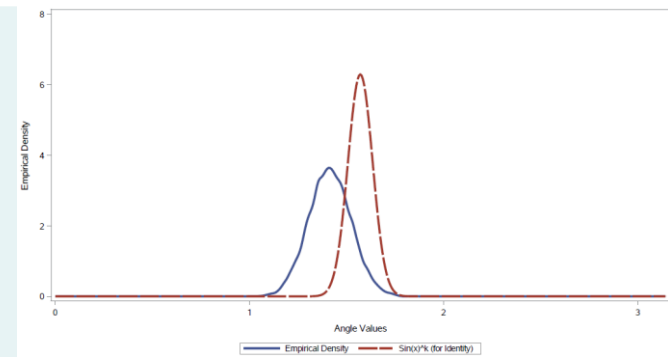
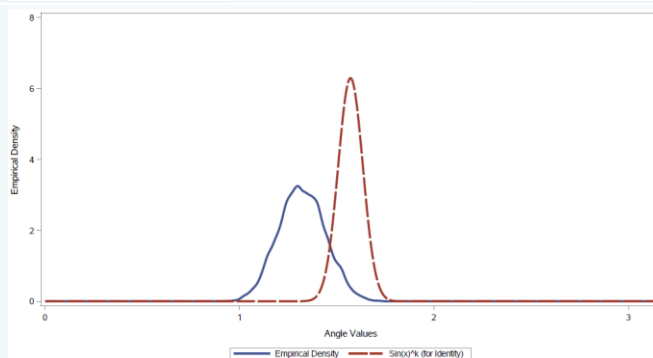
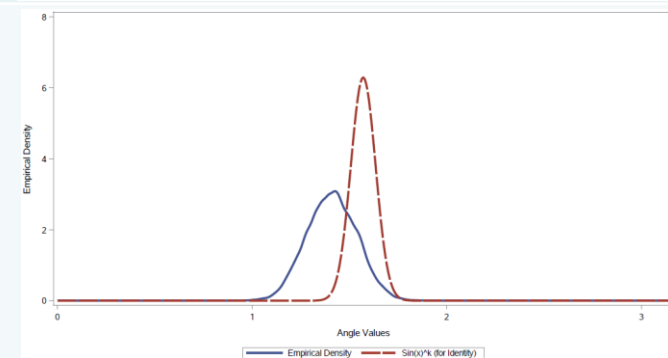
**Graph 7C1: Marchenko-Pastur vs. Spectral Distribution of Pearson's Based on Scenario 1**



# 7. Appendix C: Example of Spectral and Angles Distributions

## Graphs 7C2-7C11:

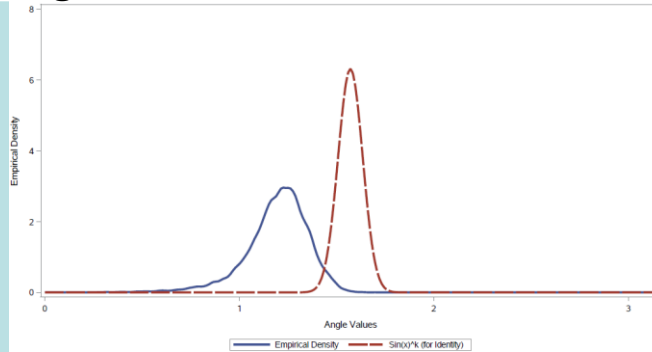
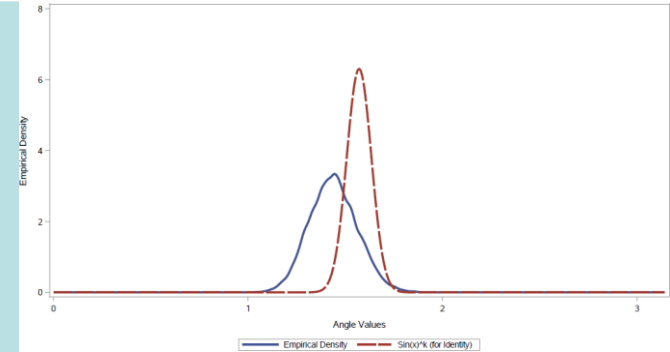
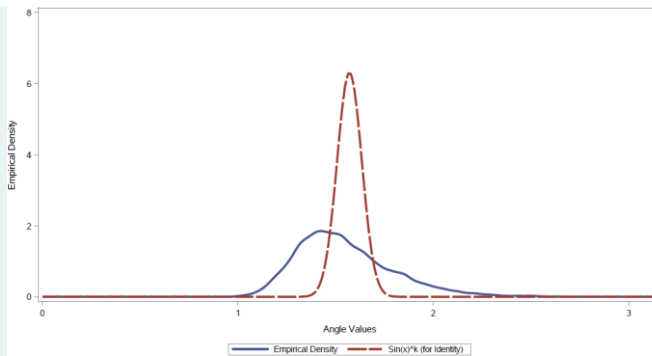
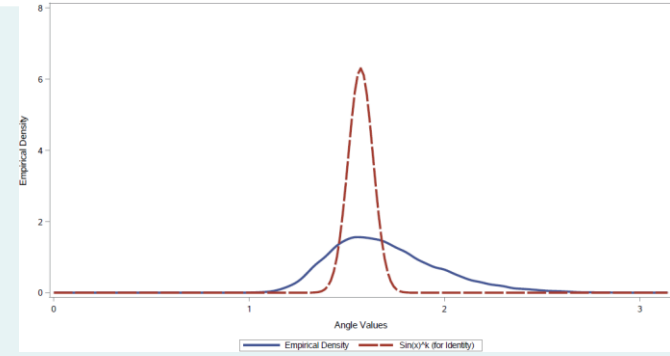
### Angles Distributions of Pearson's Identity Matrix vs. Pearson's Based on Scenario 1



# 7. Appendix C: Example of Spectral and Angles Distributions

## Graphs 7C2-7C11:

### Angles Distributions of Pearson's Identity Matrix vs. Pearson's Based on Scenario 1



Cell # Matrix for PDFs

|    |   |   |   |  |
|----|---|---|---|--|
|    |   |   |   |  |
| 7  |   |   |   |  |
| 8  | 4 |   |   |  |
| 9  | 5 | 2 |   |  |
| 10 | 6 | 3 | 1 |  |

Cell Numbers for Graphs 7C2-7C11:

# 7. Appendix D: Conditioning Algorithm

## 7. Appendix D: Conditioning Algorithm

We condition on the factors whose edges are identified by ADD to make inferences regarding direct adjacency rather than reachability.

The conditioning algorithm is described below, step by step.

The conditioning approach is a residual-based approximation and conservative in that it conditions separately on both the parent and the child factors of an edge/cell, regardless of the direction of the dependence of the other factors on either. In other words, regressions are run separately on the parent and the child, using only those factors significantly associated with BOTH, in either direction, and the residuals of each regression are used to (re)assess the (potential) dependence between the parent and the child (note that the significance threshold used to identify regression covariates is, conservatively, the same Bonferroni-adjusted threshold as that used to identify significant edges). To avoid overfitting and instability, the model uses a simple linear OLS based on the two most significant factors of the remaining eight (out of ten) factors in the Digitale et al. (2022) DAG (if there are more than two significantly associated with BOTH the parent and the child). Higher order polynomial regressions undoubtedly would identify more cases that invite conditioning, but increasing power with regressions based on more, and more complex covariates would come at the price of increasing type 1 error, i.e. the approach would be more vulnerable to overfitting and instability. Conditioning on both directions, and separately on both the parent and child factors, while using straightforward models to do so provides a reasonable balance for this tradeoff, and a principled approach to conditioning in this setting.

Notably, just as with pruning, this conditioning induces non-positive definiteness in zero cases amongst all the 10,000 simulations run in the study. This result is not surprising vis-à-vis pruning, as our method of pruning places the all-pairwise dependence matrix deeper into the convex cone and thus, farther from singularity. Conditioning appears to be having a similar effect here, as it generally will dampen observed associations and cause ADD to miss some true edges even as it is correctly rejecting more nonexistent edges. Nevertheless, positive definiteness always is verified empirically, and even though not encountered under these data conditions, any singular edge cases can be made positive definite using “nearest PD matrix” algorithms like Higham (2002). This approach, however, would not be advisable in situations where cases of non-positive definiteness occur systematically, say, more than one percent of the time, as this would distort the sample space of the dependence measure matrix, invalidate associated inference, and call into question the wisdom of using that particular dependence measure at all, at least under the given data conditions. But that is not an issue herein.

## 7. Appendix D: Conditioning Algorithm

Conditioning Algorithm:

Nsims = 10000;

For each of Nsims simulations, DO:

1. Calculate two all-pairwise dependence measure matrices (for example, using improved Chatterjee's correlation), one in each direction
2. Identify cells (factor pairs) from 1. with a statistically significant dependence measure value (where  $\alpha=0.025$ , as described above in the text), in EITHER direction
3. For each cell/pair in 2., identify those remaining factors that are statistically significantly related to BOTH variables (parent and child) of the pair, regardless of the direction of association (note that the same Bonferroni-adjusted significance threshold ( $\alpha=0.025$ ) is used here to conservatively identify relevant conditioning factors)
4. If there are more than two factors identified in 3. for a given pair/cell, keep only the two largest, where "largest" is defined as follows: each factor has two directional effects on each parent and each child – take the maximum of the two for the parent, and the maximum of the two for the child; then add these two (maximum) effects for each factor, rank these sums across the remaining factors and keep only the 2 largest.
5. Run two OLS regressions using the factors identified in 3. and 4.: one regression on the parent and one regression on the child, and keep the residuals of both regressions
6. (Re)calculate the dependence measure between the parent and child, in both directions, based on the residuals obtained from 5. and insert this value into both the all-pairwise dependence measure matrices, one in each direction (and insert the value in both cell  $i,j$  and cell  $j,i$  as the matrix must remain symmetric, even though the dependence measure is asymmetric). Verify that the matrix remains positive definite (apply Higham (2002) if rare edge cases are singular).
7. Apply pruning as described in the text.

END;