Psychometrika (2026), **91**, 1092–1126
doi:10.1017/psy.2026.10100

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Does X at Time 1 Cause Y at Time 2? Longitudinal Causal Learning with Hidden Confounders

Dexin Shi[1], Wolfgang Wiedermann[2,3], Amanda J. Fairchild[1] and Bo Zhang[4,5]

[1]Department of Psychology, University of South Carolina, Columbia, SC, USA; [2]Department of Educational, School, and Counseling Psychology, University of Missouri, Columbia, MO, USA; [3]Missouri Prevention Science Institute, University of Missouri, Columbia, MO, USA; [4]School of Labor and Employment Relations, University of Illinois Urbana-Champaign, Champaign, IL, USA; [5]Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL, USA

Corresponding author: Dexin Shi; Email: shid@mailbox.sc.edu.

(Received 7 February 2025; revised 4 February 2026; accepted 10 February 2026)

Abstract

In this study, we develop statistical methods for bivariate causal learning using higher-order moment information from two-wave longitudinal data. Within the framework of linear non-Gaussian models, we derive tests based on joint cumulants and propose a multiple-testing algorithm to detect the existence of a longitudinal causal path in the presence of hidden confounders. By combining temporal information from longitudinal designs with higher-order distributional properties of the observed data, the proposed method allows researchers to draw valid causal conclusions under realistic scenarios commonly encountered in psychological studies. The performance of the proposed algorithm is evaluated through simulation studies. As expected for higher-moment methods, the proposed algorithm may require larger sample sizes than typically needed for second-moment methods to achieve high statistical power. Results demonstrate that the proposed algorithm provides sufficient evidence to establish the existence of a longitudinal causal path, particularly in large-scale data analysis. We then present two real-world data examples to illustrate the application of the causal learning algorithm in psychological research. Finally, we discuss practical implications and potential future research directions.

Keywords: causal learning; joint cumulants; linear non-Gaussian models; longitudinal design

1. Introduction

The primary questions driving research in the social and behavioral sciences aim to uncover cause-and-effect relationships rather than mere associations (Grosz et al., 2020). Traditionally, randomized experiments have been regarded as the gold standard for drawing causal conclusions in many empirical research fields, including psychology (Rubin, 1974). However, successfully randomizing participants or manipulating the independent variable of interest is often unfeasible, unethical, or impractical in many psychological research contexts. Consequently, while many research questions in psychology implicitly involve causality, researchers must often rely on observational data as their only viable option to draw inferential conclusions (Grosz et al., 2020).

In causal reasoning, research objectives can generally be divided into related but distinct categories (Shimizu, 2019). The first involves estimating the magnitude of true causal effects, assuming that one already knows that x causes y. The second is causal structure learning (or causal discovery), where the goal is to uncover the causal structure or determine the directionality of relationships between variables.

© The Author(s), 2026. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

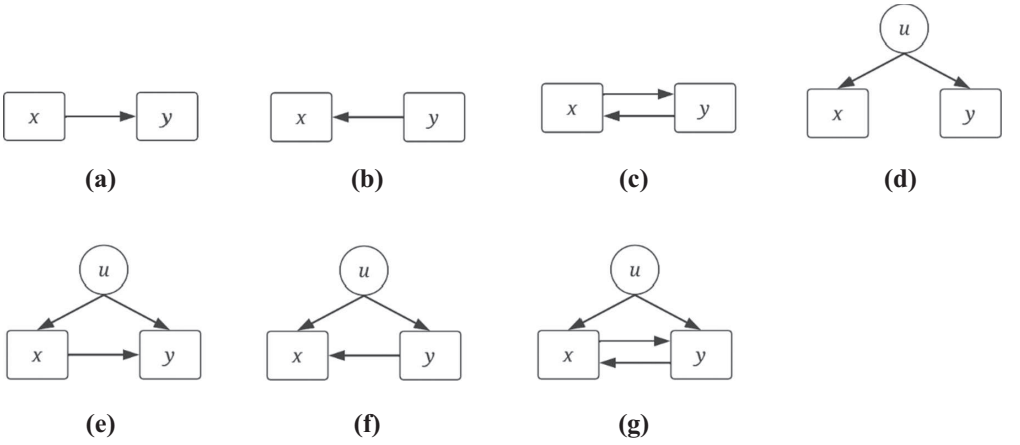


Figure 1. A list of possible data-generating models: *x* is associated with *y*.

The research focus in the present study falls into the latter category, aiming to identify *whether x* causes *y*.

Learning causal structures from observational data is challenging; hence, the well-known adage *correlation does not imply causation*. In other words, a correlation between two events does not prove that one event causes the other. Assuming two events are truly correlated, however, it is philosophically implied that the data-generating process involves some form of underlying causal structure that introduces such correlation. According to *Reichenbach's Common Cause Principle* (Reichenbach, 1956), if two events are correlated, the correlation arises either from a direct causal relationship between the events or from a third factor that generates the correlation, known as a Reichenbachian "common cause" (Hofer-Szabó et al., 2013; Reichenbach, 1956).

In the context of causal structure learning, the Reichenbach Common Cause Principle naturally applies in common research scenarios involving two variables of interest (i.e., *x* and *y*). If *x* and *y* are truly associated (i.e., the association is not due to random or sampling error), at least seven possible causal structures could explain the observed correlation in the data. These seven data-generating models are illustrated as panels (a–g) in Figure 1.[1] Specifically, the association can be the result of a direct causal relationship between the two variables, that is, (a) *x* causes *y* (or in path notation, $x \rightarrow y$); (b) *y* causes *x* ($y \rightarrow x$); or (c) a reciprocal causal structure exists (i.e., $x \rightarrow y$ and $y \rightarrow x$). Alternatively, the observed association may not result from a direct causal relationship but can instead be fully attributed to a (set of) hidden confounder(s), *u*, as depicted in panel d (i.e., $x \leftarrow u \rightarrow y$). In addition, the direct causal paths from models (a–c) and hidden confounders (*u*) can coexist, as shown in models (e–g). For example, in model e, the association between *x* and *y* is driven by both: (1) a direct causal path from *x* to *y* and (2) a (set of) hidden confounders, *u* (i.e., $x \rightarrow y$ and $x \leftarrow u \rightarrow y$).

Distinguishing among competing data-generating models is often a central objective in psychological research, even if not explicitly articulated, with the true causal structure between two variables of interest carrying significant theoretical and practical implications. For example, consider loneliness (*x*) and depression (*y*). If researchers hypothesize that loneliness causes depression (i.e., in path notation, loneliness → depression, as represented by models a, c, e, or g in Figure 1), targeted interventions can be designed to alleviate depressive symptoms by reducing loneliness. Conversely, when the true data-generating model does not include the loneliness → depression path (e.g., models b, d, or f in Figure 1),

[1] We acknowledge that the data-generating models we listed are not exhaustive. For example, a spurious association between two variables may exist due to the presence of a collider bias (e.g., when a collider is controlled for in the data analysis or due to selection bias in the sampling procedure; Pearl, 2009). In this study, we do not address collider variables and assume that no collider is controlled for in the data analysis, as well as no selection bias is present in the sampling process.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

interventions focusing solely on loneliness can be expected to be ineffective in ameliorating depressive symptoms.

Unfortunately, competing models cannot be distinguished using commonly employed linear methods that rely on first- and second-order moments (i.e., means, variances, and covariances), such as correlational analysis, linear regression, and structural equation modeling (SEM). When applying these statistical techniques with only x and y being observed, all seven causally competing models presented in Figure 1 are statistically equivalent (or symmetrical) and provide identical representations of the information in the data. In recent years, various causal discovery algorithms (e.g., Hyvärinen et al., 2001; Mooij et al., 2016; Peters et al., 2014; Shimizu et al., 2006) and statistical methods for directional dependence (e.g., Dodge & Rousson, 2000; Sungur, 2005; von Eye & DeShon, 2012; Wiedermann & von Eye, 2015; Wiedermann et al., 2021) have been developed. In contrast to standard methods of associations, these causal learning techniques make use of higher than second order moments (e.g., skewness, kurtosis, co-skewness, and co-kurtosis). These methodological advancements have allowed researchers to investigate the causal structure between variables using observational data.

While many causal discovery algorithms (e.g., Shimizu et al., 2011; Spirtes et al., 2001) are designed to explore causal structures in multivariate data involving multiple variables, we focus on bivariate causal discovery within the framework of linear non-Gaussian models (Shimizu et al., 2006; Wiedermann et al., 2021). It is worth noting that, although both aim to infer causal direction and may share underlying statistical foundations,[2] bivariate causal discovery is not necessarily a special or simpler case of multivariate causal discovery. Specifically, many multivariate causal discovery techniques rely on information from multiple observed variables, such as conditional independencies used in the Peter-Clark (PC) algorithm (Spirtes et al., 2001), and therefore, they aim to explore the causal structure for part, if not all, of the underlying network of variables. In contrast, in the bivariate case, such as the research scenario considered in the current study, researchers can only leverage information from the two observed variables. As a result, whereas methods for multivariate causal learning are exploratory in nature for the entire causal network, the bivariate approach is often used as a confirmatory framework for testing specific hypotheses and/or comparing competing theories about the causal directionality between two variables of interest. In this study, we focus on the bivariate case, as it frequently addresses fundamental questions regarding psychological theories, particularly during the early stages of research (cf. Wiedermann & von Eye, 2025).

Under linear non-Gaussian models, methodologists aim to differentiate causally competing models by leveraging higher-moment information inherent in nonnormally distributed data. Many properties and tests have been developed to uncover the causal structure of the true data-generating model within this context. The associated asymmetry properties that facilitate this identification have been integrated and unified into the comprehensive framework of Direction Dependence Analysis (DDA; Wiedermann & von Eye, 2025; Wiedermann et al., 2021).

By applying DDA principles, a subgroup of causally competing models (such as those in Figure 1) can be uniquely identified from the data. Specifically, under the assumption of no unmeasured confounding, DDA leverages higher-order information from observed variables and error terms of fitted linear models to differentiate two causally competing models, such as Model 1a ($x \rightarrow y$) and Model 1b ($y \rightarrow x$) (Wiedermann & von Eye, 2025; Wiedermann et al., 2021). These properties have been validated through analytical derivations, and their performance in finite samples has been evaluated in previous simulation studies (e.g., von Eye & DeShon, 2012; Wiedermann & von Eye, 2015). However, the assumption of no unobserved confounding is rather strict and is often violated in practical scenarios.

To account for the presence of unmeasured confounders (u), more general properties and tests based on the independence of predictors and regression errors have been introduced in the DDA

[2]For example, under linear non-Gaussian models, the Darmois-Skitovitch theorem can be used to discover causal directionality in both the bivariate case and multivariate networks, as in the DirectLiNGAM algorithm (Shimizu et al., 2011).

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

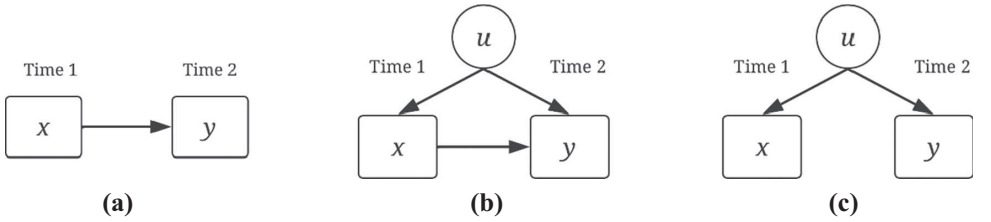


Figure 2. A list of possible data-generating models: x is associated with y, and x is measured before y.

framework (Wiedermann & Sebastian, 2020; Wiedermann et al., 2021). These asymmetry properties enable researchers to distinguish between unconfounded and confounded models. Specifically, it allows one to differentiate among Model 1a ($x \rightarrow y$), Model 1b ($y \rightarrow x$), and a family of models involving latent confounders (i.e., 1d–1f). When unmeasured confounders (u) are present, however, existing approaches provide limited information about either the presence or the direction of causality between x and y. In other words, causally competing models involving u (i.e., 1d–1f) are indistinguishable using extant methods without imposing additional assumptions (e.g., the assumption that the convolution of hidden confounders follows a normal distribution; Chen & Chan, 2013; Schultheiss et al., 2024; Wiedermann, 2022; Wiedermann & von Eye, 2025).

More recently, Shi et al. (2025) proposed an algorithm that shows promise in detecting the directionality of causation in the presence of arbitrarily distributed unmeasured confounders (i.e., distinguishing between Models 1e and 1f). However, the proposed algorithm also relies on additional untenable assumptions: (1) the unobserved confounding is minor and (2) a true causal path exists between x and y. Chen et al. (2024) developed a statistical approach to identify causal links between observed variables and then determine their causal direction in the presence of non-Gaussian latent confounder (i.e., distinguishing among Models 1d, 1e, and 1f). The algorithm by Chen et al. (2024) relies on higher-order cumulants (at least of the fifth order) and, unsurprisingly, requires very large sample sizes (around 100,000) for acceptable performance.

While the framework of linear non-Gaussian models offers valuable insights into bivariate causal discovery, particularly by enabling researchers to identify causal structures between variables without temporal information (i.e., such as in cross-sectional data), only a subset of causally competing models can be distinguished with current methods. Determining the existence and direction of causal paths becomes particularly challenging in the presence of unmeasured or latent confounders. In real-world observational studies, unmeasured confounders are often the norm rather than the exception (D'Onofrio et al., 2020), significantly limiting the applicability of causal discovery techniques in realistic scenarios commonly encountered in psychological research.

Incorporating temporal order information from longitudinal data can complement existing causal discovery algorithms, by offering additional information on the underlying causal structure that generated the observed variable association. One advantage of using a longitudinal design, widely recognized by psychologists, is its capacity to more effectively test causal hypotheses (Maxwell & Cole, 2007; Rutter, 1994). Specifically, a key implicit assumption of longitudinal designs is that an effect always follows its cause in time. Thus, in the context of causal structure learning, when the variables x and y are truly associated, and x is measured earlier in time than y (i.e., when the competing causal path, $y \rightarrow x$, can be ruled out on theoretical grounds), integrating temporal order information reduces the number of possible competing causal models from 7 (Figure 1) to 3, as shown in Figure 2.

Temporal information can therefore bring researchers closer to concluding that x causes y (i.e., Models 2a and 2b), as illustrated in Figure 2, though there remains an alternative causal model where the association between x and y is fully explained by unmeasured confounders u (i.e., Model 2c). As discussed earlier, existing non-Gaussian structure learning algorithms can distinguish Model 2a from

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Models 2b and 2c.[3] However, Model 2a represents a rather unrealistic scenario, as it assumes that there are no unmeasured confounders between x and y. In most applications, the more crucial (and unresolved) question is how to distinguish between Models 2b and 2c.

In this study, we aim to integrate insights from both distributional information in the data and temporal information over time to address the issue of learning longitudinal causal structures from observational data. Within the framework of linear non-Gaussian models, we introduce statistical theories and corresponding algorithms to test the existence of longitudinal causal paths (i.e., whether x at time 1 causes y at time 2) when unmeasured confounders are present. In contrast to many existing approaches (e.g., Shi et al., 2025; Wiedermann et al., 2021), which view unmeasured confounders as undesirable external factors hampering sound causal conclusions, the algorithm proposed here explicitly leverages the information carried by latent confounders, as we will demonstrate in detail later.

The remainder of this article is organized as follows. First, we derive the statistical theory and propose an algorithm for bivariate longitudinal causal discovery based on joint higher-order cumulants. Next, we evaluate the performance of the proposed algorithm through Monte Carlo simulations (Study I). In Study II, we further examine the factors influencing the behavior of the proposed test statistics, with the aim of explaining the observed statistical power rates and offering guidelines for researchers applying the proposed methods. We then present two empirical examples to illustrate the real-world application of the proposed algorithms with psychological data: the first real-world example uses a cross-sectional design with temporal information provided from retrospective reports; the second example uses a truly longitudinal design. Finally, we conclude with a discussion of the implications of our findings, best-practice recommendations for empirical researchers and potential directions for future research.

2. Non-Gaussian longitudinal causal learning

2.1. Notations and assumptions

For two variables of interest with temporal information, we consider two competing causal models as shown in Figure 3:

a. Model I: There is no causal relationship between x and y. Any observed association between x and y is fully explained by the unobserved confounder u (i.e., $x \leftarrow u \rightarrow y$).
b. Model II: x causes y, and both x and y are also caused by the unobserved confounder u ($x \rightarrow y$ with $x \leftarrow u \rightarrow y$).

Assuming that the data-generating mechanism can be approximated by the linear model, the population Model I can be expressed as follows (without loss of generality, the intercept terms are omitted):

$$x = \beta_1 u + e_x, \tag{1}$$

$$y = \beta_2 u + e_y, \tag{2}$$

where β_1 and β_2 represent the causal effects from u to x and y, respectively. e_x and e_y denote the corresponding error terms for x and y. For Model II, the true data-generating model can be expressed

[3]Under the framework of linear non-Gaussian models, causal models without unmeasured confounders and those with unmeasured confounders can be distinguished by testing the independence between the predictor and the error term. For example, in the context of differentiating among competing causal models with temporal order information (shown in Figure 2), Model 2a implies that the predictor and the error term are independent. In contrast, for models with latent confounders (i.e., Models 2b and 2c), the predictor and the error term can be expected to be dependent. For further details, see Wiedermann et al. (2021).

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

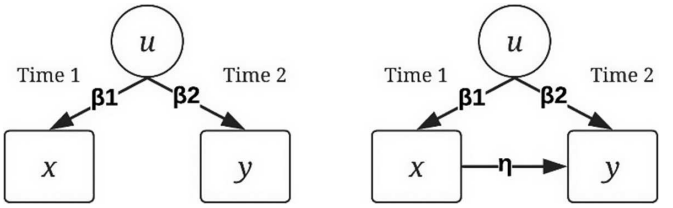


Figure 3. Population causal models considered in the current study.

as follows:

$$x = \beta_1 u + e_x, \tag{3}$$

$$y = \beta_2 u + \eta x + e_y, \tag{4}$$

where η indicates the true causal effect from x to y. Other notations are the same as defined before.

Note that there is only one latent confounder (u) in the above equations. Without loss of generality, however, this global latent confounder (u) can be conceptualized as an additive combination of k independent hidden confounders u_i ($i = 1, \ldots, k$), representing all possible common causes between x and y (see Figure S1 in the Supplementary Material; Chen & Chan, 2013; Wiedermann & von Eye, 2025). Importantly, researchers need not assume that the confounders are independent. The reason for this is that, in linear acyclic causal systems, a set of dependent hidden confounders can always be expressed as the linear sum of hidden independent factors (Hoyer et al., 2008; Shimizu, 2022).

As a key observation, under linearity, Model II can be viewed as a special case of Model I when η = 0. Thus, distinguishing between Model I and Model II is equivalent to testing whether η = 0. The two causally competing models cannot be distinguished when fitting a simple linear regression (or SEM) model that relies solely on the first- and second-order moments of the observed data. Specifically, assuming the data-generating model follows Equations (3) and (4), the estimated ordinary least square regression coefficient of the mis-specified model $y = \eta' x + e_y{}'$ (i.e., the model that erroneously ignores the presence of u) can be expressed as $\widehat{\eta'} \xrightarrow{p} \frac{Cov(x,y)}{Var(x)} = \eta + \frac{\beta_1 \beta_2 var(u)}{\beta_1{}^2 var(u) + var(e_x)}$. As a consequence, the estimated slope $\widehat{\eta}$ will always be nonzero, regardless of its true population value, unless either β_1 or β_2 are zero (i.e., in the absence of u).

Building on the notations introduced above, we make the following additional assumptions to facilitate detecting the presence of longitudinal causal effects from x to y within the linear non-Gaussian framework.

(A1). ***Temporal order.*** Observed variable x precedes y (i.e., the reverse causal path $y \rightarrow x$ is not permitted).

(A2). ***Linearity.*** The relationships between all variables are linear.

(A3). ***Independence.*** e_x and e_y are additive error terms with zero means and constant variances that are independent of u and of each other, that is, $e_x \perp e_y$, $e_x \perp u$, and $e_y \perp u$.

(A4). ***Presence of a hidden confounder.*** There exist hidden factors (u_i) that cause both x and y. For $i > 1$, multiple hidden factors u_i can be represented by a global hidden component u. In the presence of u, $\beta_1 \neq 0$ and $\beta_2 \neq 0$.

(A5). ***Nonnormality.*** Both u and e_x are nonnormally distributed.[4]

Here, Assumption (A5) is in sharp contrast to the distributional assumptions routinely made in standard second-order moment-based methods of association. While OLS regression and SEMs usually rely on the assumption that variables follow a normal distribution, the methods presented here rest

[4]More specifically, as we detail later, the proposed method assumes nonnormality with finite third- or fourth-order moments. For example, u and e_x cannot follow a Cauchy or a Lévy distribution. We thank the anonymous reviewers for pointing this out.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

on the assumption that variables deviate from normality. This requirement is justifiable from both, an empirical and a theoretical perspective. From an empirical perspective, nonnormality of variables constitutes a phenomenon that is ubiquitous across the empirical sciences (see, e.g., Blanca et al. 2013; Micceri, 1989), suggesting that the requirement of nonnormality can be expected to be fulfilled in practice.

From a theoretical perspective, the ubiquitousness of nonnormality can be explained using Cramér's decomposition theorem (Cramér, 1970), stating that when the sum of two independent real-values variates follows a normal distribution, both summand variates must follow a normal distribution as well. This holds accordingly for more than two independent variables, that is, when the sum of many finite independent variables follows a normal distribution, then all independent variables must be normal. As a consequence, the sum of independent variables cannot be exactly normal, except all independent summands are exactly normal (known as the linear closure property of the normal distribution; see, e.g., Glymour et al., 2019). Since this strict requirement is likely to be violated in real-world settings, nonnormality can be expected to be omnipresent in practice.

It should be noted that Cramér's decomposition theorem is not in conflict with the central limit theorem (CLT), which is well known among psychologists.[5] Specifically, under the Lyapunov condition (Lyapunov, 1901), the sum of independent variables converges to a normal distribution as the number of variables approaches infinity. However, with a finite number of variables (e.g., independent confounders[6]), the sum may come close to the normal distribution but will never be exactly normally distributed. Moreover, the rate of convergence to normality can be slow, particularly when the independent variables have large third moments (Zhang et al., 2023). Note that the number of potential confounders never reaches infinity in practice. In most clinical or psychological applications, the number of potential confounders is typically moderate to small. For example, Pocock et al. (2004) reviewed 73 articles in observational epidemiology and found that the median number of potential confounders was 7.

While we believe the assumption of nonnormality is quite realistic, given both empirical evidence and theoretical justifications via Cramér's decomposition theorem, a practical question is whether the variables exhibit enough nonnormality to yield adequate statistical power when implementing the proposed algorithm, as we will demonstrate later in this article. From a practical perspective, justification may draw on prior theory regarding factors such as the number of potential confounders and the degree of their skewness. In addition, we would like to emphasize that, in the theoretical framework, we conceptualize u as a global latent (or hidden) construct of hidden confounding influences not considered in the analysis. This issue of approximate normality can be addressed at the research design stage. For instance, researchers may collect information on major confounders guided by substantive theory and adjust for selected confounders in the data analysis. Consequently, the set of latent (hidden) confounders u decreases with the number of observed confounding covariates which, in turn, can decrease the "Gaussianization effect" of u. This way, data situations can be constructed based on a limited number of confounders, possibly as few as one, that are expected to be strongly nonnormal by design. In addition, incorporating various measured confounders may also serve as a tool for sensitivity analysis of the proposed algorithm, helping to evaluate the robustness of its assumptions and the stability of the causal conclusions.

Under Assumptions (A1)–(A5), we can derive properties and tests to examine the presence of the longitudinal causal path (η) using marginal and joint higher cumulants (Chen & Chan, 2013; Chen et al., 2024; Hyvärinen & Smith, 2013; Wiedermann, 2022). Before we present the main theoretical results, we briefly review basic properties of cumulants.

[5]We thank the anonymous reviewers for pointing this out.

[6]To demonstrate the implications of the relationship between Cramér's decomposition theorem and the CLT, we used the latent confounder as an example. The same logic can be applied to the nonnormality of e_x, which can be conceptualized as an additive combination of independent unique causes of x.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

2.2. Cumulants

Cumulants are a set of descriptive constants of a given (joint) probability distribution that are, from a theoretical perspective, sometimes more useful than moments (Stuart & Ord, 1987).

Definition 1 ***(Joint cumulants)*** (Brillinger, 2001). Let $\mathbf{X} = (X_1, X_2, \ldots X_n)$ be a random vector of length n. The joint cumulants are formally defined as partial derivatives of the joint *cumulant generating function* (*cgf*) at zero and are denoted as follows:

$$\kappa_X(n_1, \ldots n_k) = \left. \frac{\partial^{n_1+n_2+\cdots+n_k} \kappa_X(t_1, t_2 \ldots t_k)}{\partial t_1^{n_1} \partial t_2^{n_2} \ldots \partial t_k^{n_k}} \right|_{t_1=0, t_2=0, \ldots t_k=0}. \tag{5}$$

The joint *cgf* of $\mathbf{X}$ is given by

$$\kappa_X(t) = \log[M_X(t)], \tag{6}$$

where $M_X(t) = \mathbb{E}[\exp(tX)]$ is the *joint moment generating function (mgf).*[7]

By taking the coefficients of the Taylor series expansion of$\kappa_X(t)$, the joint cumulants of $\mathbf{X}$ can be then evaluated using the following function (Brillinger, 2001):

$$Cum(X_{i1}, X_{i2}, \cdots X_{ik}) = \sum_{\pi \in \mathrm{P}} (-1)^{|\pi|-1} (|\pi|-1)! \prod_{B \in \pi} \mathbb{E}\left[\prod_{j \in B} X_{ij} \right], \tag{7}$$

where P is the set of all partitions of the set $\{1, 2, \ldots k\}$, $|\pi|$is the number of blocks in the partition π, and B is the block in the partition π.

For convenience, we use $Cum_i(x)$ to denote the ith order cumulants for a single variable x and $Cum(x, x, \ldots, y, y, \ldots)$ to indicate joint cumulants involving two variables. For example, $Cum_3(x)$ and $Cum_4(x)$ refer to the third- and fourth-order (marginal) cumulants describing the skewness and excess-kurtosis of x, $Cum(x,x,y)$ and $Cum(x,y,y)$ describe the third-order (joint) cumulants, and $Cum(x,x,x,y)$, $Cum(x,y,y,y)$, and $Cum(x,x,y,y)$ describe the fourth-order (joint) cumulants of x and y. For simplicity and without restricting generality, the observed variables x and y are hereinafter assumed to be standardized. All variables in the system are expressed in mean-centered form (mean = 0).

Lemma 1. *Under Model II (defined by Equations (3) and (4) and Assumptions (A2)–(A5), we obtain the following results for the third- and fourth-order joint cumulants of observed variables x and y:*

$$Cum(x,x,y) = \beta_1^2(\beta_2 + \eta\beta_1)\, Cum_3(u) + \eta Cum_3(e_x), \tag{8}$$

$$Cum(x,y,y) = \beta_1(\beta_2 + \eta\beta_1)^2 Cum_3(u) + \eta^2 Cum_3(e_x) \tag{9}$$

$$Cum(x,x,x,y) = \beta_1^3(\beta_2 + \eta\beta_1)\, Cum_4(u) + \eta Cum_4(e_x), \tag{10}$$

$$Cum(x,x,y,y) = \beta_1^2(\beta_2 + \eta\beta_1)^2 Cum_4(u) + \eta^2 Cum_4(e_x) \tag{11}$$

$$Cum(x,y,y,y) = \beta_1(\beta_2 + \eta\beta_1)^3 Cum_4(u) + \eta^3 Cum_4(e_x) \tag{12}$$

A proof for Lemma 1 is given in Appendix A. Note that under Model I, that is, when $\eta = 0$*, joint cumulants simplify to*

$$Cum(x,x,y) = \beta_1^2 \beta_2 Cum_3(u) \tag{13}$$

$$Cum(x,y,y) = \beta_1 \beta_2^2 Cum_3(u) \tag{14}$$

$$Cum(x,x,x,y) = \beta_1^3 \beta_2 Cum_4(u), \tag{15}$$

[7]Even though we defined the joint cumulants using the MGF, as shown later, finite values of $\kappa 3$ and $\kappa 4$ are sufficient for the proposed tests to hold, regardless of whether the MGF is globally defined. We thank the anonymous reviewers for pointing this out.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

$$Cum(x,x,y,y) = \beta_1^2\beta_2^2 Cum_4(u) \tag{16}$$

$$Cum(x,y,y,y) = \beta_1\beta_2^3 Cum_4(u) \tag{17}$$

In the following section, we make use of these higher-order cumulants to develop measures for the presence of longitudinal causal paths under hidden confounding.

2.3. Measures for the presence of a longitudinal causal path (η)

Using the results from Lemma 1, we derive the following theorems.

Theorem 1. *Let* $\Delta_1 = Cum(x,x,x,y) \times Cum(x,y,y,y) - [Cum(x,x,y,y)]^2$. *If no true causal effect from x to y exists (i.e.,* $\eta = 0$*), we have* $\Delta_1 = 0$. *If x truly causes y* ($\eta \neq 0$*), with the faithfulness assumption that* $\eta \neq -\frac{\beta_2}{\beta_1}$, *we have* $\Delta_1 \neq 0$.

Corresponding Proof. Given Equations (10)–(12), we obtain

$$\begin{aligned}\Delta_1 &= Cum(x,x,x,y) \times Cum(x,y,y,y) - [Cum(x,x,y,y)]^2 \\ &= \left[\beta_1^3(\beta_2 + \eta\beta_1)\, Cum_4(u) + \eta Cum_4(e_x)\right] \times \left[\beta_1(\beta_2 + \eta\beta_1)^3 Cum_4(u) + \eta^3 Cum_4(e_x)\right] \\ &\quad - \left[\beta_1^2(\beta_2 + \eta\beta_1)^2 Cum_4(u) + \eta^2 \mathrm{Cum}_4(e_x)\right]^2. \end{aligned} \tag{18}$$

By setting $\Delta_1 = 0$ and simplifying the result, we obtain the following quadratic equation for η, assuming $Cum_4(u) \neq 0$, $Cum_4(e_x) \neq 0$, $\beta_1 \neq 0$, and $\beta_2 \neq 0$:

$$Cum_4(u)Cum_4(e_x)\beta_1^2\beta_2^2\eta^2 + Cum_4(u)Cum_4(e_x)\beta_1\beta_2^3\eta = 0. \tag{19}$$

Solving Equation (19) for η, the only two solutions are

$$\eta_1 = 0$$

$$\eta_2 = -\frac{\beta_2}{\beta_1}.$$

Therefore, when $\eta = 0$, it follows that $\Delta_1 = 0$ and, assuming faithfulness holds (i.e., $\eta \neq -\frac{\beta_2}{\beta_1}$), when $\eta \neq 0$, then $\Delta_1 \neq 0$.

It should be noted that by utilizing information based on higher-order cumulants, our method is related to prior work in causal discovery (e.g., Chen & Chan, 2013; Chen et al., 2024), although the research goals and algorithms differ in several important ways. For instance, the same analytic result in Theorem 1 was also used by Chen et al. (2024) as the first step (i.e., Case 1) in their causal learning algorithm. In addition, Chen et al. (2024) provided and proved a more general result based on joint cumulants of fourth order or higher (see their Theorem 3). A more detailed comparison between our algorithm and other cumulant-based causal learning methods is presented later in the article.

In a similar manner, using Equations (8)–(12), we can derive two additional theorems by utilizing both third- and fourth-order joint cumulants as follows.

Theorem 2. *Let* $\Delta_2 = Cum(x,x,y) \times Cum(x,x,y,y) - Cum(x,y,y) \times Cum(x,x,x,y)$. *When no true causal effect from x to y exists (i.e.,* $\eta = 0$*), one obtains* $\Delta_2 = 0$. *When x truly causes y* ($\eta \neq 0$*), with the faithfulness assumption that* $\eta \neq -\frac{\beta_2}{\beta_1}$, *one obtains* $\Delta_2 \neq 0$.

The corresponding proof for Theorem 2 is given in Appendix B.

Theorem 3. *Let* $\Delta_3 = Cum(x,x,y) \times Cum(x,y,y,y) - Cum(x,y,y) \times Cum(x,x,y,y)$. *When no true causal effect from x to y exists (i.e.,* $\eta = 0$*), one obtains* $\Delta_3 = 0$. *When x truly causes y* ($\eta \neq 0$*), with the faithfulness assumption that* $\eta \neq -\frac{\beta_2}{\beta_1}$ *and* $\eta \neq -\frac{Cum_4(u)Cum_3(e_x)\beta_2}{Cum_4(u)Cum_3(e_x)\beta_1 - Cum_3(u)Cum_4(e_x)}$, *one obtains* $\Delta_3 \neq 0$.

The corresponding proof for Theorem 3 is given in Appendix C.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

With the assumptions stated above and $n \to \infty$, all three test statistics ($\Delta_1, \Delta_2,$and Δ_3) can be used to detect the presence of a direct causal path from x to y. Specifically, $\eta \neq 0$finds empirical support when $\Delta_1 \neq 0$, $\Delta_2 \neq 0$,or $\Delta_3 \neq 0$. In finite samples where only x and y are observed, the test statistics $\Delta_1, \Delta_2,$and Δ_3 can be computed using the sample joint cumulants estimated from Equation (7). In addition, statistical significance tests can be performed using $100(1-\alpha)$ % nonparametric bootstrapping confidence intervals (Davison & Hinkley, 1997).

A key point to recognize is that Δ_1 relies solely on the fourth-order joint cumulants and presupposes nonzero marginal fourth-order cumulants for both u and e_x. In contrast, both Δ_2 and Δ_3 incorporate third- and fourth-order joint cumulants, requiring the existence of third-order cumulants (i.e., skewness) for u and e_x. In this sense, Δ_1 is regarded as a more general test that relies on fewer assumptions. However, by incorporating third-order information when available, using Δ_2 and Δ_3 offers incremental power to detect the presence of η. Furthermore, even when the third-order cumulants are zero, both Δ_2 and Δ_3 are expected to be zero (as shown in Equations B2 and C2), which does not necessarily lead to incorrect conclusions, particularly within the framework of the multiple testing algorithm introduced in the following section.

2.4. Operationalizing a multiple testing algorithm

We proposed three test statistics ($\Delta_1, \Delta_2,$and Δ_3) for evaluating the presence of a nonzero direct causal effect η. In finite samples, the statistical power to detect a true effect ($\eta \neq 0$) is expected to be influenced by two primary factors: the sample size (n) and the absolute magnitudes of the test statistics ($\Delta_1, \Delta_2,$and Δ_3). As shown in the previous section, the magnitudes of these test statistics can be expected to be influenced by the complex interplay of several parameters, including $Cum_4(u), Cum_4(e_x), Cum_3(u), Cum_3(e_x)$, β_1, and β_2. Therefore, each test statistic is expected to exhibit greater sensitivity to detect η for a given combination of these parameters. In practice, the values of the above parameters are unknown, given that u is unobserved. To achieve higher statistical power, we recommend using all three test statistics in conjunction within a multiple testing algorithm, as described below.

A multiple testing algorithm

1. Set the significance level α and apply the Bonferroni correction (Bonferroni, 1936) for multiple testing using the adjusted $\boldsymbol{\alpha' = \alpha/3}$.
2. Standardize the observed variables x and y in subsequent analyses.
3. Compute $\boldsymbol{\Delta_1}$, and obtain the $100(1-\boldsymbol{\alpha'})$ % percentile bootstrapping confidence interval, $[\boldsymbol{\Delta_1^L}, \boldsymbol{\Delta_1^U}]$.
4. Compute $\boldsymbol{\Delta_2}$, and obtain the $100(1-\boldsymbol{\alpha'})$ % percentile bootstrapping confidence interval, $[\boldsymbol{\Delta_2^L}, \boldsymbol{\Delta_2^U}]$.
5. Compute $\boldsymbol{\Delta_3}$, and obtain the $100(1-\boldsymbol{\alpha'})$ % percentile bootstrapping confidence interval, $[\boldsymbol{\Delta_3^L}, \boldsymbol{\Delta_3^U}]$.
6. If $0 \notin \left[\boldsymbol{\Delta_1^L}, \boldsymbol{\Delta_1^U}\right] \vee 0 \notin \left[\boldsymbol{\Delta_2^L}, \boldsymbol{\Delta_2^U}\right] \vee 0 \notin \left[\boldsymbol{\Delta_3^L}, \boldsymbol{\Delta_3^U}\right]$, then select $x \to y$. Otherwise, fail to detect the presence of a direct causal path.

3. Study I: Assessing the performance of the proposed algorithm

3.1. Data generation

We conducted Monte Carlo simulations to evaluate the performance of the proposed algorithm in the context of longitudinal bivariate causal discovery with unmeasured confounders. We generated data from a linear non-Gaussian model, where the population model is the same as the one presented in Figure 3b. Specifically, the variables of interest include x and y, and we assumed a temporal order such that x occurs before y. Both x and y share a common cause (confounder), u. All exogenous variables (i.e., u) and error terms for the endogenous variables (i.e., x and y) were generated from various Gamma distributions. Since the research goal is to determine the existence of the causal effect from x to y (i.e., $x \to y$), the effect size of the true causal path (η) was manipulated as a key factor of the simulation

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

design, with details discussed below. We also manipulated the sample size, the distribution of the hidden confounder (u), the distribution of the error term (e_x), and the effect size of β_1 and β_2 paths. Linear transformations were applied to the generated endogenous variables and error terms, so that all population path coefficients in the simulation reflect standardized metrics.

Sample size (N). Four levels of sample sizes were considered, including N = 2,000, 5,000, 10,000, and 50,000. It is worth noting that to accurately estimate higher-order statistics beyond the first and second moments, a larger sample size is generally required (Chan et al., 2020). The selection of these sample sizes represents a range from small to large samples for causal discovery under linear non-Gaussian models (Chen et al., 2024; Shi et al., 2023; Shimizu et al., 2006; Tramontano et al., 2024).

Distribution of the hidden confounder (u). The latent confounder u was generated from a Gamma distribution, with the scale parameter fixed at 1 and the shape parameter manipulated to create three levels of population skewness, 0.75, 1.50, and 2.25, mimicking mild, moderate, and highly skewed distributions. These distributional choices align with previous simulation studies in causal learning (Shi et al., 2023; Wiedermann & Li, 2020).

Distribution of the error term (e_x). The error term of x (e_x) was generated from a Gamma distribution with three levels of population: 0.75, 1.50, and 2.25. For simplicity, the error term of y (e_y) was set to follow the same distribution as e_x.

Effect size of η. The $x \rightarrow y$ path represents the true longitudinal causal effect of interest. Four levels of effect size for η were considered: 0, 0.14, 0.39, and 0.59. These population parameters were chosen to represent scenarios with no causal effect between x and y, as well as true causal effects ranging from small to large effect sizes (Cohen, 1988).

Effect sizes of the β_1 and β_2. The strength of the confounder was manipulated by altering the effect sizes of the β_1 and β_2 paths. For each path, three effect sizes were considered: 0.14, 0.39, and 0.59, representing small, medium, and larger confounder effects. These levels were fully crossed for the two paths, resulting in nine unique combinations.

In total, 1,296 (4 × 3 × 3 × 4 × 9) simulation conditions were considered. Note that 36 conditions were not included because they were inadmissible (i.e., conditions with $\beta_1 = \beta_2 = \eta = 0.59$, which yields a negative error variance). Therefore, the final number of simulated conditions was 1,296 – 36 = 1,260. For each simulation condition, 1,000 replications were generated using R 4.1.2 (R Core Team, 2025).

3.2. Data analyses and outcome variables

For each simulated data set, we conducted joint-cumulant-based tests, that is, Δ_1, Δ_2, and Δ_3, to detect the presence of the $x \rightarrow y$ path. Statistical inferences were drawn based on the $100(1 - \alpha)$ % percentile bootstrapping confidence intervals with 2,000 resamples. Conclusions were based on the multiple-testing algorithm described in the previous section. That is, a nonzero causal effect was confirmed, when at least one of the three test statistics significantly differed from zero. To control for Type I error rates due to multiple testing, the Bonferroni correction (Bonferroni, 1936) was applied. Specifically, for $\alpha = 0.05$ with a Bonferroni correction ($\alpha' = 0.05/3$), 98.33% bootstrap confidence intervals were constructed.

Using the proposed multiple-testing algorithm, two possible decisions can be made: (1) x causes y and (2) there is no evidence that x causes y. For each simulation condition, we computed the proportions of these decisions across replications. As the outcome measure, we reported the proportion of cases concluding that x causes y ($x \rightarrow y$) across simulation conditions. Note that the proportions represent Type I error rates when $\eta = 0$, whereas for conditions where $\eta \neq 0$, proportions indicate power rates for detecting the presence of a true causal path $x \rightarrow y$. The Type I error rates across simulated conditions (i.e., when $\eta = 0$) are summarized in Table 1. The power rates to detect a true causal path from x to y (i.e., when $\eta \neq 0$) are reported in Tables 2–4.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 1. Type I error rates in percent for detecting the causal path $x \rightarrow y$ ($\eta = 0$)

Skew. of e_x	β_2	β_1	Skewness of $u = 0.75$				Skewness of $u = 1.50$				Skewness of $u = 2.25$			
			$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$	$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$	$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$
0.75	0.14	0.14	0.4	0.1	0.1	0.0	0.6	0.2	0.3	1.9	0.1	0.0	0.1	0.1
		0.39	0.3	0.0	0.1	0.0	0.0	0.1	0.3	3.2	0.4	0.3	0.7	2.5
		0.59	0.3	0.2	0.3	2.6	1.1	1.4	1.7	1.5	1.8	4.3	2.6	3.0
	0.39	0.14	0.2	0.4	0.1	0.4	0.2	0.2	0.4	3.2	0.4	0.0	0.9	3.4
		0.39	0.3	0.2	1.3	3.4	0.4	2.0	2.4	2.1	1.4	2.9	2.9	2.8
		0.59	0.5	2.9	3.3	3.5	3.1	3.2	4.4	2.5	5.9	4.4	5.0	3.1
	0.59	0.14	0.5	0.8	0.7	1.7	0.5	0.8	3.8	2.7	1.4	3.2	2.8	2.3
		0.39	1.0	2.1	3.2	3.2	3.4	3.4	4.2	3.5	3.7	5.2	3.2	4.1
		0.59	2.0	3.6	3.6	4.8	5.7	4.9	4.6	0.0	6.6	4.9	5.7	3.2
1.50	0.14	0.14	0.0	0.2	0.0	0.0	0.2	0.1	0.0	1.6	0.2	0.0	0.1	0.1
		0.39	0.2	0.3	0.3	0.3	0.3	0.4	0.2	3.3	0.1	0.5	0.7	2.0
		0.59	0.3	0.5	0.6	2.9	0.6	0.7	2.4	0.9	1.1	2.7	2.8	2.4
	0.39	0.14	0.2	0.0	0.1	0.4	0.2	0.1	0.4	3.3	0.0	0.9	1.1	2.4
		0.39	0.3	0.4	0.6	2.8	0.8	2.1	4.2	3.8	1.2	3.5	3.4	3.8
		0.59	0.9	2.5	2.8	4.1	4.6	4.4	3.5	3.7	5.6	6.7	4.1	3.4
	0.59	0.14	0.3	0.2	0.7	1.9	0.8	1.5	1.8	3.7	1.7	3.3	2.3	3.4
		0.39	1.2	1.6	2.6	2.4	3.5	4.8	4.3	4.8	6.2	5.6	4.3	4.2
		0.59	2.6	4.0	4.1	4.0	8.6	6.6	6.5	0.0	6.9	6.3	6.2	4.3

(Continued)

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 1. Continued

Skew. of e_x	β_2	β_1	Skewness of u = 0.75				Skewness of u = 1.50				Skewness of u = 2.25			
			N = 2,000	N = 5,000	N = 10,000	N = 50,000	N = 2,000	N = 5,000	N = 10,000	N = 50,000	N = 2,000	N = 5,000	N = 10,000	N = 50,000
2.25	0.14	0.14	0.0	0.1	0.2	0.0	0.1	0.0	0.0	0.8	0.1	0.0	0.2	0.0
		0.39	0.3	0.0	0.0	0.4	0.3	0.2	0.5	2.2	0.3	0.5	1.3	2.6
		0.59	0.2	0.3	0.7	2.0	0.9	1.1	2.5	1.4	1.8	3.3	3.2	3.7
	0.39	0.14	0.0	0.2	0.0	0.0	0.1	0.4	0.2	4.5	0.3	0.2	0.8	1.5
		0.39	0.3	0.3	0.8	3.4	0.8	2.1	2.5	3.3	0.9	3.4	4.3	4.4
		0.59	1.3	2.0	2.2	3.3	3.2	4.1	4.8	2.7	5.4	6.5	5.7	5.3
	0.59	0.14	0.3	0.1	0.2	1.8	1.0	1.3	2.7	4.5	2.1	2.4	3.2	3.4
		0.39	1.3	2.4	2.0	4.0	3.8	5.1	4.1	5.1	5.2	4.5	5.1	4.1
		0.59	3.1	5.4	6.6	4.3	6.5	7.2	6.0	7.2	8.4	8.3	7.4	5.7

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 2. Power rates in percent for detecting the causal path $x \rightarrow y$ ($\eta = 0.14$)

Skew. of e_x	β_2	β_1	Skewness of $u = 0.75$				Skewness of $u = 1.50$				Skewness of $u = 2.25$			
			$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$	$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$	$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$
0.75	0.14	0.14	2.4	2.8	3.4	3.6	1.4	3.2	2.7	4.1	1.7	2.1	3.7	2.4
		0.39	1.6	2.5	3.1	4.2	2.0	2.7	4.2	3.5	2.4	2.6	3.2	2.6
		0.59	2.2	2.7	2.9	3.7	3.3	4.3	3.5	2.7	4.7	3.0	4.0	4.2
	0.39	0.14	1.2	3.3	4.4	12.0	1.7	3.4	8.3	44.9	0.7	6.4	16.7	**84.5**
		0.39	1.7	4.5	6.8	27.9	2.3	5.8	9.7	52.0	4.8	13.1	32.2	**98.4**
		0.59	2.2	2.8	3.6	7.2	3.6	4.4	7.2	28.4	5.5	6.0	15.0	72.6
	0.59	0.14	1.4	5.3	13.2	76.8	2.6	16.8	49.7	**99.9**	8.1	59.2	**95.5**	**100.0**
		0.39	2.9	7.5	19.3	**90.4**	10.7	38.9	75.0	**100.0**	29.0	73.7	**95.7**	**100.0**
		0.59	2.8	5.7	8.6	43.2	8.2	15.2	32.7	**98.3**	14.3	30.0	56.2	**99.7**
1.50	0.14	0.14	5.2	5.1	3.1	3.5	6.8	5.4	4.2	4.4	5.5	3.9	4.5	5.5
		0.39	5.0	5.0	6.2	4.8	6.6	5.5	4.9	7.4	5.8	4.2	3.0	5.4
		0.59	4.4	5.9	5.0	4.6	4.3	5.6	5.1	3.1	5.5	4.1	2.7	3.0
	0.39	0.14	5.7	7.1	6.3	24.5	6.2	10.1	18.4	76.1	6.3	15.0	32.1	**97.3**
		0.39	5.6	12.2	25.3	**87.6**	7.1	13.6	34.6	**96.5**	6.0	13.8	27.7	**94.4**
		0.59	5.9	12.3	16.4	68.2	4.1	6.1	7.0	28.0	4.3	5.4	9.8	53.6
	0.59	0.14	5.8	14.3	30.9	**97.1**	11.5	54.7	**88.0**	**100.0**	22.9	79.7	**98.1**	**100.0**
		0.39	11.2	51.2	**88.0**	**100.0**	20.6	59.2	**91.8**	**100.0**	40.1	**87.1**	**99.3**	**100.0**
		0.59	7.3	21.3	40.3	**99.5**	6.2	18.3	49.5	**99.9**	18.5	53.6	**81.1**	**99.9**

(Continued)

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 2. Continued

Skew. of e_x	β_2	β_1	Skewness of u = 0.75				Skewness of u = 1.50				Skewness of u = 2.25			
			N = 2,000	N = 5,000	N = 10,000	N = 50,000	N = 2,000	N = 5,000	N = 10,000	N = 50,000	N = 2,000	N = 5,000	N = 10,000	N = 50,000
2.25	0.14	0.14	7.6	5.2	3.8	4.5	7.9	5.7	6.1	4.5	7.3	6.7	4.9	5.9
		0.39	7.6	7.5	6.0	5.8	8.8	5.8	5.9	11.2	8.8	6.2	6.6	13.8
		0.59	7.5	5.9	7.2	6.7	6.1	5.0	5.2	6.7	7.7	6.2	5.6	5.7
	0.39	0.14	10.2	7.7	8.6	25.8	10.5	11.4	20.8	73.3	7.8	16.6	35.1	**95.7**
		0.39	12.3	20.3	32.6	**92.2**	13.5	34.6	67.4	**98.9**	14.7	36.8	66.1	**99.6**
		0.59	10.3	18.3	34.2	**92.7**	8.9	15.3	26.3	**88.2**	4.4	6.4	13.3	49.2
	0.59	0.14	9.6	18.0	34.1	**94.7**	22.7	61.4	**87.8**	**99.9**	44.6	**88.6**	**98.9**	**100.0**
		0.39	29.0	78.7	**97.5**	**100.0**	45.3	**90.0**	**99.3**	**100.0**	49.7	**92.2**	**99.7**	**100.0**
		0.59	20.5	57.9	**87.9**	**99.9**	11.5	35.5	72.4	**100.0**	15.7	48.7	**80.7**	**100.0**

Note: Power rates above 80% are highlighted in bold.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 3. Power rates in percent for detecting the causal path $x \rightarrow y$ ($\eta = 0.39$)

Skew. of e_x	β_2	β_1	Skewness of $u = 0.75$				Skewness of $u = 1.50$				Skewness of $u = 2.25$			
			$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$	$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$	$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$
0.75	0.14	0.14	4.9	5.2	4.6	4.5	6.3	6.0	5.2	5.0	6.3	3.7	4.5	5.3
		0.39	6.8	5.4	4.3	4.6	4.7	4.9	2.8	4.2	3.8	3.3	3.9	7.3
		0.59	5.3	6.6	5.2	4.0	4.4	4.1	3.3	5.5	6.2	6.1	6.4	29.1
	0.39	0.14	5.4	5.6	6.7	18.9	4.8	8.2	12.3	56.3	6.6	11.1	24.9	**94.0**
		0.39	5.2	7.1	14.1	63.9	6.3	14.6	32.1	**98.5**	20.5	64.8	**93.3**	**100.0**
		0.59	3.5	4.7	6.4	27.9	7.9	17.3	44.6	**99.7**	23.5	65.1	**92.2**	**100.0**
	0.59	0.14	6.5	14.2	20.1	**84.3**	13.0	35.2	67.7	**100.0**	26.8	74.4	**98.3**	**100.0**
		0.39	11.7	30.3	59.0	**100.0**	49.0	**92.3**	**99.7**	**100.0**	**93.4**	**99.8**	**99.9**	**100.0**
		0.59	8.1	21.7	47.0	**100.0**	53.6	**88.7**	**99.3**	**100.0**	**81.3**	**97.4**	**99.3**	**100.0**
1.50	0.14	0.14	8.7	6.4	4.2	3.8	7.4	6.6	5.4	5.7	9.1	7.8	7.0	7.0
		0.39	7.6	6.3	7.4	11.5	7.3	6.2	7.7	14.6	6.1	6.9	6.8	13.1
		0.59	7.7	7.7	6.8	12.3	7.9	5.3	5.8	7.3	6.4	4.7	4.6	4.2
	0.39	0.14	8.2	9.7	11.4	29.6	10.4	14.7	22.9	**81.4**	13.7	23.4	41.0	**96.5**
		0.39	14.9	29.5	50.7	**98.2**	18.5	43.3	68.4	**99.9**	16.4	40.2	68.4	**100.0**
		0.59	16.6	29.7	54.2	**98.8**	6.7	15.5	26.9	**97.0**	13.5	40.3	74.4	**99.9**
	0.59	0.14	13.0	22.1	43.6	**95.8**	28.5	63.2	**86.1**	**99.8**	49.2	**85.2**	**98.2**	**100.0**
		0.39	49.8	**89.3**	**97.7**	**100.0**	63.8	**95.7**	**100.0**	**100.0**	**83.7**	**99.3**	**100.0**	**100.0**
		0.59	39.5	**81.5**	**97.6**	**100.0**	46.1	**89.8**	**98.6**	**100.0**	**87.4**	**98.7**	**99.7**	**100.0**

(Continued)

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 3. Continued

Skew. of e_x	β_2	β_1	Skewness of u = 0.75				Skewness of u = 1.50				Skewness of u = 2.25			
			N = 2,000	N = 5,000	N = 10,000	N = 50,000	N = 2,000	N = 5,000	N = 10,000	N = 50,000	N = 2,000	N = 5,000	N = 10,000	N = 50,000
2.25	0.14	0.14	9.7	7.3	7.9	4.5	11.3	7.9	8.5	7.5	10.9	7.2	8.4	8.5
		0.39	9.7	8.5	7.0	12.8	12.0	8.4	10.9	23.0	11.6	10.2	10.3	28.6
		0.59	8.8	7.4	10.3	20.6	8.8	8.1	10.8	25.5	8.5	7.7	6.5	11.4
	0.39	0.14	11.9	12.1	12.6	32.4	14.6	20.3	27.7	79.2	16.3	27.6	46.5	**94.3**
		0.39	23.1	39.6	60.9	**97.7**	35.5	65.7	**89.3**	**99.9**	33.0	69.3	**88.5**	**99.9**
		0.59	28.9	55.6	**80.5**	**99.8**	24.8	46.7	76.0	**99.8**	12.3	26.5	52.5	**99.8**
	0.59	0.14	16.7	27.2	42.4	**92.2**	37.7	67.6	**89.3**	**99.6**	57.7	**88.0**	**97.7**	**100.0**
		0.39	70.9	**94.4**	**99.0**	**100.0**	**89.0**	**98.0**	**99.9**	**100.0**	**89.6**	**99.3**	**99.9**	**100.0**
		0.59	**83.4**	**97.5**	**99.7**	**100.0**	63.7	**94.3**	**99.7**	**100.0**	79.7	**95.9**	**99.2**	**100.0**

Note: Power rates above 80% are highlighted in bold.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 4. Power rates in percent for detecting the causal path $x \rightarrow y$ ($\eta = 0.59$)

Skew. of e_x	β_2	β_1	Skewness of $u = 0.75$				Skewness of $u = 1.50$				Skewness of $u = 2.25$			
			$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$	$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$	$N = 2{,}000$	$N = 5{,}000$	$N = 10{,}000$	$N = 50{,}000$
0.75	0.14	0.14	7.5	5.7	4.8	3.7	6.1	5.7	4.5	4.7	6.8	6.3	5.6	4.9
		0.39	7.4	6.3	4.9	8.5	6.0	5.0	4.1	4.3	6.2	4.0	5.5	16.8
		0.59	5.4	6.6	4.2	4.6	5.6	5.1	4.5	13.6	6.8	10.4	15.1	73.3
	0.39	0.14	7.0	6.2	8.4	23.6	7.6	9.4	14.3	67.2	6.7	15.0	32.8	**98.2**
		0.39	8.2	13.6	25.3	**88.0**	10.8	30.8	64.6	**100.0**	50.1	**92.2**	**99.7**	**100.0**
		0.59	4.9	6.7	14.9	70.4	30.1	75.4	**95.9**	**100.0**	77.1	**97.4**	**100.0**	**100.0**
	0.59	0.14	9.0	16.1	27.3	**90.2**	15.0	36.6	71.5	**100.0**	29.2	76.3	**98.3**	**100.0**
		0.39	18.7	49.0	**81.4**	**100.0**	69.0	**97.1**	**99.9**	**100.0**	**96.4**	**99.7**	**100.0**	**100.0**
		0.59	–	–	–	–	–	–	–	–	–	–	–	–
1.50	0.14	0.14	11.2	7.7	8.1	6.9	10.2	8.9	8.5	8.0	10.0	8.1	8.9	9.7
		0.39	12.5	10.0	9.3	17.7	12.3	11.5	12.6	32.8	9.9	9.2	9.0	21.5
		0.59	11.2	10.7	11.7	31.0	9.5	7.4	6.5	15.9	7.2	5.1	4.6	9.4
	0.39	0.14	13.1	10.3	15.8	41.1	14.6	21.6	34.0	**87.9**	16.5	32.1	52.0	**96.8**
		0.39	25.8	50.3	76.8	**100.0**	35.0	63.7	**88.6**	**100.0**	27.4	63.1	**90.0**	**100.0**
		0.59	32.1	61.0	**88.6**	**100.0**	15.9	37.7	66.5	**100.0**	49.7	**86.6**	**97.7**	**100.0**
	0.59	0.14	15.8	31.8	50.3	**97.9**	37.2	68.5	**88.8**	**99.7**	55.9	**84.6**	**97.6**	**100.0**
		0.39	77.3	**96.7**	**99.6**	**100.0**	77.2	**97.9**	**99.9**	**100.0**	**86.6**	**99.2**	**99.4**	**100.0**
		0.59	–	–	–	–	–	–	–	–	–	–	–	–

(Continued)

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 4. Continued

Skew. of e_x	β_2	β_1	Skewness of u = 0.75				Skewness of u = 1.50				Skewness of u = 2.25			
			N = 2,000	N = 5,000	N = 10,000	N = 50,000	N = 2,000	N = 5,000	N = 10,000	N = 50,000	N = 2,000	N = 5,000	N = 10,000	N = 50,000
2.25	0.14	0.14	13.9	8.9	8.8	7.2	14.4	10.6	12.2	11.0	15.6	12.6	10.8	13.7
		0.39	15.5	13.9	12.7	25.3	15.9	16.6	18.1	44.6	14.3	16.7	23.1	54.2
		0.59	13.5	17.0	18.5	47.0	13.1	16.8	18.7	56.0	12.2	10.1	9.2	29.7
	0.39	0.14	15.5	14.8	17.5	46.1	20.8	27.7	41.8	**88.1**	24.2	42.1	60.7	**96.7**
		0.39	41.1	64.0	**85.1**	**99.7**	62.7	**86.5**	**96.1**	**99.9**	55.8	**83.9**	**96.3**	**100.0**
		0.59	59.8	**89.5**	**98.7**	**100.0**	48.3	79.6	**94.0**	**99.8**	26.0	58.7	**88.3**	**100.0**
	0.59	0.14	22.6	38.0	57.0	**96.1**	50.6	74.7	**91.4**	**99.6**	66.5	**91.1**	**96.2**	**100.0**
		0.39	**90.8**	**99.1**	**99.7**	**100.0**	**93.3**	**98.7**	**99.6**	**100.0**	**89.9**	**98.2**	**99.8**	**100.0**
		0.59	–	–	–	–	–	–	–	–	–	–	–	–

Note: Power rates above 80% are highlighted in bold.

3.3. Results: Type I error rates

As shown in Table 1, the Type I error rates of the proposed multiple-testing algorithm were well-controlled under the simulated conditions. Across the 324 conditions with $\eta = 0$, the Type I error rates ranged from 0% to 8.6%. Only three simulation conditions with relatively small sample sizes (i.e., $N <$ 5,000) produced Type I error rates marginally above the 7.5% criterion suggested by Bradley (1978). Thus, we conclude that researchers using the proposed algorithm face a low risk of falsely concluding that x causes y when no true causal path exists, particularly in large-sample contexts.

3.4. Results: Power rates

As shown in Tables 2–4, under conditions where x is truly a cause of y (i.e., $\eta \neq 0$), the proposed multiple-testing algorithm can successfully detect the presence of the causal path of $x \rightarrow y$. The power rates for detecting a true causal path varied across simulation conditions. Below, we summarize the main factors that drive power rates.

First, not surprisingly, the power to detect the true causal path (η) increased with the sample size. For example, as indicated in Table 2, when the population skewness of the latent confounder (u) was 1.50 and the population skewness of e_x was 0.75, with N = 2,000, $\beta_1 = 0.14$, and $\beta_2 = 0.59$, the power to detect a causal $x \rightarrow y$ path of $\eta = 0.14$ was only 10.7 %. Keeping all other factors the same, the power increased to 38.9%, 75.0%, and 100% as the sample size increased to 5,000, 10,000, and 50,000, respectively.

Second, the power rates to detect the $x \rightarrow y$ path generally increased with the true effect size (η). For example, when the population skewness of $u = 0.75$, the population skewness of $e_x = 0.75$, $\beta_1 = \beta_2 =$ 0.39, and N = 50,000, the power rates were 27.9%, 63.9%, and 88.0% for $\eta = 0.14$, $\eta = 0.39$, and $\eta = 0.59$, respectively.

Third, the power rates were related to the strength of the latent confounder, particularly the magnitude of β_2 (i.e., the $u \rightarrow y$ path). For instance, as demonstrated in Table 3, when the population skewness of the latent confounder (u) was 1.50, the population skewness of the e_x was 2.25, with a relatively small sample size of N = 2,000, $\beta_1 = 0.39$, and $\beta_2 = 0.14$, the proposed algorithm was underpowered (i.e., 12.0%) to detect the presence of the causal path $x \rightarrow y$ with an effect size of $\eta =$ 0.39. However, holding all other factors constant, the power increased to 35.5% when $\beta_2 = 0.39$ and to 89.0% when $\beta_2 = 0.59$.

Finally, higher power rates were also observed as the level of nonnormality of u and/or e_x increased. Taking $\eta = 0.59$ (Table 4) as an example, with N = 10,000 and $\beta_2 = 0.39$, when both u and e_x are mildly nonnormal (i.e., population skewness = 0.75), the power rates ranged from 8.4% to 25.3%, indicating inadequate power to detect a causal path with a large effect size. Nevertheless, under conditions where both u and e_x are severely nonnormal (i.e., population skewness = 2.25), the power rates greatly improved, ranging from 60.7% to 96.3%.

In summary, the proposed algorithm generally yielded higher power to detect the true causal path under desirable conditions, including larger sample sizes, larger effect sizes of the true causal path, stronger confounders (particularly with a stronger $u \rightarrow y$ path), and greater levels of nonnormality of u and/ore_x. Roughly speaking, when the sample size was relatively small (i.e., N = 2,000), adequate power (i.e., $\geq$ 80%) could be achieved only for detecting a causal path with a medium or large effect size, and under conditions in which the effect size of the $u \rightarrow y$ path was large (i.e., $\beta_2 = 0.59$), and either u or e_x showed considerable deviations from normality. To achieve power rates $\geq$ 80% for detecting a causal path with a small effect size (i.e., $\eta = 0.14$), a sample size of 10,000 with a large $u \rightarrow y$ path (i.e., $\beta_2 = 0.59$) and moderate to serve level of nonnormality in u ande_x is required. When the sample size reached 50,000, most conditions yielded power rates approaching 100%, regardless of the effect size of the true causal path, provided that the effect size for the $u \rightarrow y$ path was medium or large, except under undesirable conditions where the level of nonnormality in both u and e_x was mild.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

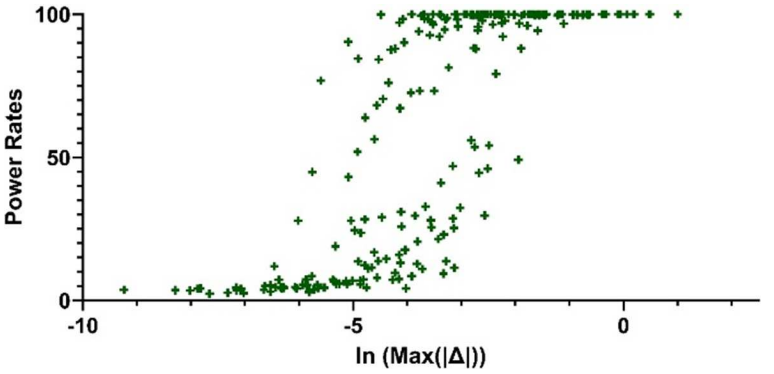


Figure 4. Relationship between ln (max|Δ|) and power rates (N = 50,000).

4. Study II: Evaluating the relationship of power rates and test statistics ($\Delta_1, \Delta_2,$ and Δ_3)

In Study I, a notable finding was that the power rates of the proposed algorithm were not determined solely by sample size and the magnitude of the true causal path (η). For example, our simulation results indicated that low statistical power could still occur under certain conditions when the true causal path had a large effect size (η = 0.59), even with N = 50,000.

This observation can be explained by the fact that the power rates can be expected to be related to the magnitudes (in absolute values) of the test statistics (i.e., Δ_1, Δ_2 and Δ_3), which are not a strictly monotonic function of η. As shown in the derivations for Equations (18), (B1), and (C1), the magnitudes of the test statistics ($\Delta_1, \Delta_2,$ and Δ_3) are influenced not only by the true causal effect size (η) but also by a combination of several other parameters. These parameters include higher-order cumulants of u and e_x [e.g., $Cum_3(u)$, $Cum_3(e_x)$, $Cum_4(u)$, $Cum_4(e_x)$], as well as the path coefficients β_1 and β_2. In Study II, we further examined the factors influencing the behavior of the proposed test statistics, with the aim of explaining the observed statistical power rates and offering guidelines for researchers applying the proposed methods.

To better illustrate the influence of the magnitudes of the test statistics on the statistical power, we first calculated the average sample estimates of $\Delta_1, \Delta_2,$ and Δ_3 for each simulation condition with N = 50,000. We then obtained the maximum absolute value among the three average test statistics (i.e., $max|\Delta|$). The relationship between $max|\Delta|$ and the observed power rates is presented in Figure 4. Note that we applied a logarithmic transformation to the x-axis to reduce compression of smaller values and to enhance the visibility of the pattern. As shown in the figure, the power rates increased with the magnitude of $max|\Delta|$.

Given the importance of the magnitudes of the test statistics, we further examined the relationship between the magnitudes of $\Delta_1, \Delta_2,$ and Δ_3 and the simulated factors that were manipulated. Specifically, we used the same simulated conditions as in Study I and computed the population values of all relevant higher-order joint cumulants and the three test statistics.[8] The results of the population values across 315 simulated conditions are provided in the Supplementary Material. We also plotted the relationship between the maximum population values of $\Delta_1, \Delta_2,$ and Δ_3 and the various simulated conditions, as presented in Figures S2–S4 in the Supplementary Material. The results indicated that the main factors driving the magnitude of the test statistics matched the factors influencing power, as observed in Study I. In general, greater magnitudes of max|Δ| are associated with larger effect sizes of the true causal path, stronger confounders (particularly with a stronger $u \rightarrow y$ path), and greater levels of nonnormality in u and/or e_x.

[8]The population values were approximated using simulations based on a sample size of 10 million.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

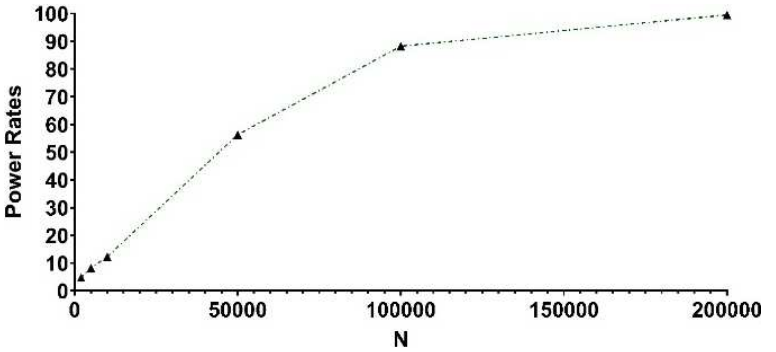


Figure 5. Power rate as a function of sample size (N): Skewness of u = 1.50, Skewness of e_x = 0.75, β_1 = 0.14, β_2 = 0.39, and η = 0.39.

On the other hand, the results also indicated that certain combinations of undesirable conditions (e.g., small effect size of η, a weak latent confounder, and a very mild level of nonnormality in u and/or e_x) can yield test statistics that are approximately, though not exactly, zero – even when $\eta \neq 0$ and all assumptions are satisfied. In practice, approximately zero test statistics may also occur under other conditions not examined in this study.[9] We thereby drew attention to the fact that in applied research, true causal effects can be masked when the magnitudes of the test statistics (Δ_1, Δ_2, and Δ_3) approach zero. Nonetheless, it should be acknowledged that observing undesirable conditions in which the effect sizes of the test statistics (Δ_1,Δ_2 and Δ_3) are close to zero do not necessarily imply that the proposed tests are technically flawed. Specifically, Theorems 1–3 demonstrate that, in the presence of a direct causal path (i.e., $\eta \neq 0$), the test statistics asymptotically do not equal zero, provided all assumptions are satisfied. As a result, under undesirable conditions with a small max|Δ|, the proposed algorithm is expected to achieve high statistical power as the sample size continues to increase.

To provide evidence for this argument, we selected one such undesirable condition with a population skewness of u = 1.50, population skewness of e_x = 0.75, β_1 = 0.14 and β_2 = 0.39, and η = 0.39 (leading to population values of Δ_1 = 0.010, Δ_2 = 0.008, and Δ_3 = 0.006). With the above combination of population parameters, we conducted additional simulations by further manipulating the sample size, ranging from 2,000 to 200,000. The power curve against sample size for the proposed algorithm is presented in Figure 5. As indicated in the figure, as the sample size continues increasing, the power rates increase from 4.8% (N = 2,000) to 99.5% (N = 200,000).

It is also important to re-emphasize that the proposed algorithm is not intended to demonstrate the absence of a causal path. Although the algorithm may have limited statistical power to detect a true causal effect—particularly under conditions where the effect (η) is masked because the test statistics (Δ_1, Δ_2, and Δ_3) approach zero—a reassuring implication is that, from a practitioner's perspective, a Type II error does not imply an incorrect causal conclusion. Instead, researchers will make an inconclusive decision (i.e., fail to detect the presence of a direct causal path).

5. Empirical Example I: Do adverse childhood experiences cause depression?

In the first empirical data example, we demonstrate the application of the proposed algorithm to detect the structure linking *adverse childhood experiences* to *depressive symptoms* in adulthood. Data were obtained from the Behavioral Risk Factor Surveillance System (BRFSS), a national, large-scale telephone survey collecting state-level data on U.S. residents' health-related behaviors, chronic conditions, and

[9]For example, as one reviewer noted, when both u and e_x follow symmetric nonnormal distributions, the opposite signs of their kurtoses can yield $\Delta_1 \approx 0$.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

preventive service use. For this analysis, we used data from the 2010 cohort to examine measures of interest.

The sample size from the BRFSS 2010 cohort was 451,075, including adults from all 50 states, the District of Columbia, Puerto Rico, Guam, and the Virgin Islands. In the current application, we included only respondents who provided complete responses to all items in the relevant measures, resulting in a final sample size of $N = 18{,}472$.

Adverse childhood experiences (ACEs) were measured with the 11-item BRFSS ACE questionnaire (Merrick et al., 2018). Participants were asked to reflect on their lives before the age of 18 and indicate whether specific events occurred during their childhood. Five statements (e.g., "Did you live with anyone who was a problem drinker or alcoholic?") were rated dichotomously as "Yes" or "No." Participants rated the frequency of the remaining six items (e.g., "How often did a parent or adult in your home ever swear at you, insult you, or put you down?") on a 3-point Likert scale ranging from 1 (never) to 3 (more than once). The Cronbach's alpha for the ACE scores in the current sample was $\alpha = 0.764$.

Depressive symptoms were measured using the 8-item Patient Health Questionnaire-8 (PHQ-8; Kroenke et al., 2009). Participants indicated the number of days over the past 2 weeks (i.e., 0–14 days) they experienced each depressive symptom (e.g., "felt down, depressed, or hopeless"). The depression scores demonstrated high reliability, with Cronbach's alpha of $\alpha = 0.842$ in the current sample.

We calculated sum scores for each measure used in the analysis. Descriptive statistics suggested that the scores were nonnormally distributed for both adverse childhood experiences ($M = 13.23$, $SD = 3.03$; Skewness = 1.83; Excess-Kurtosis = 3.47; Kolmogorov–Smirnov $D = 0.231$, $p < 0.001$) and depression ($M = 12.04$, $SD = 17.82$; Skewness = 2.58; Excess-Kurtosis = 7.73; Kolmogorov–Smirnov $D = 0.250$, $p < 0.001$). The Pearson correlation between the two variables is 0.335 ($p < 0.001$). Under the framework of linear non-Gaussian models, we tested the assumption of hidden confounders using independence tests based on the Hilbert-Schmidt Independence Criterion (HSIC; Gretton et al., 2008; Wiedermann & Li, 2018, 2020). Results showed that when fitting the model ACEs → Depression, the predictor and error terms are significantly dependent. The test results suggested that hidden confounders exist between ACEs and depression.[10]

The data indicated that adverse childhood experiences were associated with depressive symptoms in adulthood. However, this association does not necessarily imply a true causal link, as such an association can be entirely generated by latent confounders (e.g., parental mental health status; Van Santvoort et al., 2015; Walsh et al., 2002). We therefore applied our proposed algorithm to examine whether a causal relationship exists between adverse childhood experiences and depressive symptoms in adulthood.

Both variables were standardized prior to the analysis. We then computed the three test statistics, Δ_1, Δ_2, and Δ_3, and their corresponding confidence intervals separately and made collective decisions based on all three tests. To control the overall Type I error rate at $\alpha = 0.05$, a Bonferroni correction (Bonferroni, 1936) was applied, resulting in an adjusted $\alpha' = 0.05/3 = 0.017$. Accordingly, 98.33% confidence intervals (CIs) were constructed for each statistic using 5,000 nonparametric bootstrap resamples based on the percentile method. Results are shown in Table 5.

Results indicated that the 98.33% CI for Δ_2 included zero, failing to provide evidence for a direct causal path. In contrast, the test statistics for both Δ_1 and Δ_3 are significantly different from zero, as reflected in their CIs. Using the multiple-testing algorithm and controlling for α inflation, the findings support the presence of a causal link from adverse childhood experiences to depressive symptoms in

[10]Specifically, when fitting the linear non-Gaussian model ACEs → Depression, the HSIC between the predictor and the estimated error terms was significant, HSIC (ACEs → Depression) = 175.5, $p < 0.001$. The independence tests were conducted using nonparametric bootstrapping with 1,000 replications. It should be noted that without temporal information, the presence of latent confounders requires testing independence by fitting both competing causal models: HSIC (ACEs → Depression) and HSIC (Depression →ACEs). In this context, because the model Depression →ACEs can be ruled out based on temporal ordering, testing only HSIC (ACEs → Depression) is sufficient for detecting latent confounders. For further details on the use of independence tests to identify hidden confounders, see Wiedermann and Li (2018, 2020) and Maeda and Shimizu (2020).

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 5. Results from the empirical Example I

Test statistics	Test	Estimate	98.33% confidence interval
Δ_1	$cum(x,x,y,y) \times cum(x,x,y,y) - cum(x,x,x,y) \times cum(x,y,y,y)$	**−0.514**	**[−0.733, −0.313]**
Δ_2	$cum(x,x,x,y) \times cum(x,y,y) - cum(x,x,y,y) \times cum(x,x,y)$	0.015	[−0.036, 0.066]
Δ_3	$cum(x,x,y,y) \times cum(x,y,y) - cum(x,y,y,y) \times cum(x,x,y)$	**−0.240**	**[−0.316, −0.172]**

Note: Test statistics that are significant are highlighted in bold.

adulthood. The conclusion aligns with prior theoretical and empirical work suggesting that early life experiences play a causal role in shaping mental health outcomes (Hughes et al., 2017; Tan & Mao, 2023).

It should be acknowledged that, in this example, the temporal order information is derived from retrospective reports. Specifically, ACEs capture experiences recalled from childhood, whereas depressive symptoms reflect reports of the two weeks preceding data collection. This temporal structure allows us to examine potential longitudinal causal processes. However, strictly speaking, the data in this example are cross-sectional with respect to the timing of data collection. Reliance on retrospective reports introduces the risk of several types of bias, such as "recollection biases" (remembering past states as better or worse than it actually was), "implicit theory of change biases" (reconstructing former states from the current state and assumptions on how the state changed over time), and "present state effect biases" (using information about the current state to generate a judgment on former states; for an overview, see, e.g., Blome & Augustin, 2015). Therefore, although the proposed algorithm can be applied, researchers should exercise additional caution when interpreting findings from cross-sectional studies that rely on retrospective reporting.

6. Empirical Example II: Does loneliness cause anxiety in older adults?

In the second data example, we apply the proposed algorithm to a truly longitudinal design, aiming to detect the causal path from *loneliness* (at Time 1) to *anxiety symptoms* (at Time 2) among older adults. The analyses used data from the Health and Retirement Study (HRS), a large-scale, nationally representative longitudinal study that surveys approximately 20,000 Americans over the age of 50 every 2 years. For the current analysis, we used data from subsample A (i.e., a rotating random 50% subsample of the longitudinal panel), collected in the 2018 and 2022 waves. After removing incomplete observations, the final sample size was $N = 3{,}005$.

Loneliness was assessed using an 11-item subscale from the 20-item Revised UCLA Loneliness Scale (Russell et al., 1980). Participants were asked to rate how often they experienced different aspects of loneliness (e.g., "lack companionship") on a 3-point Likert scale ranging from 1 (Hardly ever or never) to 3 (Often). The Cronbach's alpha for loneliness scores in the 2018 wave was $\alpha = 0.89$.

Anxiety symptoms were measured using a five-item scale selected from the Beck Anxiety Inventory (BAI; Beck et al., 1988). Participants indicated how often they experienced each anxiety symptom (e.g., "I had fear of the worst happening") using a 4-point Likert scale ranging from 1 (Never) to 4 (Most of the time). The Cronbach's alpha for anxiety scores collected in the 2022 wave was $\alpha = 0.82$.

The sum scores indicated nonnormality for both measures of loneliness ($M = 16.68$, $SD = 4.92$; Skewness = 0.83; Excess-Kurtosis = 0.07; Kolmogorov–Smirnov $D = 0.133$, $p < 0.001$) and anxiety ($M = 7.78$, $SD = 2.92$; Skewness = 1.12; Excess-Kurtosis = 0.88; Kolmogorov–Smirnov $D = 0.171$, $p < 0.001$). Loneliness at Time 1 and anxiety symptoms at Time 2 were significantly associated, with a Pearson correlation of 0.321 ($p < .001$). Under a linear non-Gaussian model, the independence test suggested the presence of hidden confounder(s) between the two variables of interest.[11] We then aim to investigate whether loneliness causes later anxiety symptoms in older adults or whether the longitudinal association is merely due to unobserved confounders (e.g., personality traits; Buecker et al., 2020; Hakulinen et al.,

[11] HSIC (Loneliness → Anxiety) = 13.00, $p < 0.001$.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 6. Results from the empirical Example II

Test statistics	Test	Estimate	95% CI
Δ_1	$cum(x,x,y,y) \times cum(x,x,y,y) - cum(x,x,x,x) \times cum(x,y,y,y)$	**0.040**	**[0.005, 0.104]**
Δ_2	$cum(x,x,x,y) \times cum(x,y,y) - cum(x,x,y,y) \times cum(x,x,y)$	**−0.050**	**[−0.099, −0.017]**
Δ_3	$cum(x,x,y,y) \times cum(x,y,y) - cum(x,y,y,y) \times cum(x,x,y)$	−0.006	[−0.500, 0.029]

Note: Test statistics that are significant are highlighted in bold.

2015). The same test procedure described in Example 1 was followed, and the results are reported in Table 6.

As shown in Table 6, although the 98.33% CI for Δ_3 includes zero, both Δ_1 and Δ_2 were significantly different from zero, as reflected in their CIs. The results support the presence of a causal link between loneliness and anxiety symptoms among older adults. The findings provide causal evidence that loneliness leads to adverse mental health outcomes and align with prior theories and empirical research (Mann et al., 2022; Park et al., 2020).

7. Discussion and conclusion

7.1. Summary and discussion of major findings

In this study, we developed statistical methods for bivariate causal discovery using longitudinal data and higher-order information from joint cumulants. Within the framework of linear non-Gaussian models, we derived three test statistics and proposed a multiple-testing algorithm to detect the existence of a longitudinal causal path in the presence of hidden confounders. By integrating temporal information from longitudinal designs and higher-order distributional information from the data, the proposed method allows researchers to draw valid causal conclusions under realistic scenarios commonly encountered in psychological studies. The performance of the proposed multiple-testing algorithm was evaluated using Monte Carlo simulations.

Simulation results showed that the proposed algorithm yielded Type I error rates close to the 5% nominal level. This observation suggests that by applying the proposed algorithm, researchers are not likely to incorrectly detect a causal path that does not exist in the population data-generating model. With regard to power, the proposed multiple-testing algorithm can successfully detect the existence of a true causal path, and the power rates were found to be dependent on various characteristics of the data-generating model. Based on the conditions considered in the current simulation study, higher statistical power rates were associated with larger sample sizes, large effect sizes of the true causal path (η), and a stronger latent confounder (i.e., largerβ_1 and $\beta_2$paths, particularly β_2).

Since the proposed method relies on estimating higher-order joint cumulants, not surprisingly, a larger sample size is required to achieve adequate statistical power. Therefore, the proposed algorithm is particularly suited for studies using data from national-level large-scale longitudinal surveys, such as the Health and Retirement Study (HRS, https://hrs.isr.umich.edu/), the National Longitudinal Survey of Youth 1979 (NLSY79; https://www.nlsinfo.org/), and the Adolescent Brain Cognitive Development Study (ABCD; https://abcdstudy.org/). We also provided two empirical examples to illustrate the real-world application of the proposed algorithm. As demonstrated in Example I, the algorithm can also be applied to cross-sectional designs in which temporal information is obtained through retrospective reporting, a context that typically allows for larger sample sizes. However, in such cases, researchers should exercise caution and account for potential biases associated with retrospective recall when interpreting the findings.

7.2. Recommendations for applied researchers

Based on the findings from the current study, we offer recommendations for applied researchers to consider when implementing the proposed algorithm. To begin with, it is important to note that the

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

proposed algorithm follows the basic logic of Null Hypothesis Significance Testing (NHST). Specifically, if the multiple-testing algorithm yields significant results, researchers gain empirical evidence to support the conclusion that x at Time 1 causes y at Time 2. Our simulations indicated that Type I error rates are well-controlled at their nominal level when a causal conclusion is made. However, following the standard NHST guideline, failing to reject the hypothesis that the test statistics (Δ_1,Δ_2,and Δ_3) significantly differ from zero does not provide evidence that they are equal to zero in the population. Therefore, if the results of the multiple-testing algorithm are not significant, researchers simply fail to detect a causal path based on the available evidence from the data. In this case, researchers should not blatantly conclude that x does not cause y, because nonsignificant results may arise from conditions with low statistical power, as demonstrated in our simulations.

In practice, the proposed algorithm is recommended as a methodological approach for providing sufficient evidence to support the existence of a longitudinal causal path, rather than as proof for the absence of a causal effect. In this sense, even though the proposed algorithm may require a large sample size to achieve higher statistical power, a promising takeaway is that, from a practitioner's perspective, a Type II error does not necessarily imply that the researcher is making an incorrect causal conclusion (an inconclusive decision is made instead).

In addition, it is essential to acknowledge that the validity of the proposed algorithm depends on several key assumptions. Table 7 summarizes the key assumptions and implications of potential violations. Below, we clarify and elaborate on several critical points related to some of these assumptions.

1) ***Assumption A1 (temporal order)***: In this study, we focus on longitudinal causal discovery and assume that the temporal order of the observed variables is known (i.e., x precedes y). This assumption allows researchers to rule out the causally competing path $y \rightarrow x$. Without this temporal order assumption, the proposed algorithm cannot distinguish between causally competing models 1e, 1f, and 1g, as the properties from the developed theorems also hold in the presence of the causal path $y \rightarrow x$.
2) ***Assumptions A2 (linearity) and A3 (independence):*** The assumptions of linearity and independent errors are routinely made in conventional linear models. Although the linearity assumption can be tested when the variables are observed, it is technically untestable in the current research setting, where the confounding variables are unobserved.
3) ***Assumption A4 (presence of latent confounder)***: Under linear non-Gaussian models, the presence of hidden confounders (A4) can be tested based on the independence properties between the predictor and error terms (Wiedermann et al., 2021), using omnibus independence tests such as the Hilbert-Schmidt Independence Criterion (see, e.g., Wiedermann & Li, 2020)
4) ***Assumption A5 (nonnormality):*** The proposed algorithm assumes that both u and $e_x$are nonnormally distributed. As discussed earlier, nonnormality can be considered the norm with a theoretical justification provided by Cramér's Decomposition Theorem (Cramér, 1970). That is, for variables u and e_x, if both have multiple independent causes (the sum of finite independent variables), they are expected to be nonnormally distributed as long as any of the independent causes does not follow a normal distribution (Glymour et al., 2019). In addition, using Cramér's Decomposition, assumption A5 is ***partially*** statistically testable by conducting normality tests on the observed variable x. Specifically, when x (where u and e_x are independent summands) is nonnormally distributed, either u or $e_x$must also be nonnormally distributed.
5) ***Assumption A6 (faithfulness)***: In deriving Theorems 1–3, we made additional faithful assumptions regarding specific magnitudes of η. These assumptions are common in the field of causal discovery and are not testable.

It should be emphasized that, under violations of Assumptions A4–A6, the test statistics (Δ_1,Δ_2,and Δ_3) are theoretically expected to converge to zero irrespective of the parameter η. This result has two practical implications. *First*, from a statistical perspective, the violation of any of the

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Table 7. Assumptions of the proposed approach.

Assumption	Definition	Consequences of assumption violation
A1. Temporal order	Observed variable x precedes y (i.e., no causal path is permitted from y to x)	Models with $y \rightarrow x$ cannot be ruled out, and it is not possible to distinguish between causally competing models 1e, 1f, and 1g
A2. Linearity	The relationships between all variables are approximately linear	Test statistics no longer hold under nonlinear relationships, and the tests are not trustworthy.
A3. Independence	$e_x \perp e_y$, $e_x \perp u$, and $e_y \perp u$.	Lemma 1 no longer holds, and the tests are not trustworthy.
A4. Presence of a latent confounder	There exist hidden factors (u_i) that cause both x and y. For $i > 1$, multiple hidden factors u_i are assumed to form a global hidden component u.	The test statistics fail to detect a true causal path.
A5. Nonnormality	Both u and e_x are nonnormally distributed. Δ_1: $Cum_4(e_x) \neq 0$, and $Cum_4(u) \neq 0$ Δ_2: $Cum_3(e_x) \neq 0$, $Cum_3(u) \neq 0$, $Cum_4(e_x) \neq 0$, and $Cum_4(u) \neq 0$ Δ_3: $Cum_3(e_x) \neq 0$, $Cum_3(u) \neq 0$, $Cum_4(e_x) \neq 0$, and $Cum_4(u) \neq 0$	The test statistics fail to detect a true causal path.
A6. Faithful assumptions	$\Delta 1$: $\eta \neq -\frac{\beta_2}{\beta_1}$ $\Delta 2$: $\eta \neq -\frac{\beta_2}{\beta_1}$ $\Delta 3$: $\eta \neq -\frac{\beta_2}{\beta_1}$, and $\eta \neq -\frac{Cum_4(u)Cum_3(e_x)\beta_2}{Cum_4(u)Cum_3(e_x)\beta_1 - Cum_3(u)Cum_4(e_x)}$	The test statistics fail to detect a true causal path.

Assumptions A4–A6 leads only to a Type II error (i.e., failing to detect a true causal effect). As we discussed and recommended in the revised manuscript, if the algorithm's results are not significant, researchers should interpret this outcome as a failure to detect a causal path based on the available evidence, rather than as proof that x does not cause y. Therefore, from a practitioner's standpoint, a Type II error does not necessarily imply an incorrect causal conclusion; instead, it results in an inconclusive decision. *Second*, and related to the above point, because violation of assumptions leads only to Type II errors, in practice, if the proposed algorithm identifies a causal path from x to y, and all other assumptions are satisfied with Type I error adequately controlled (as evidenced by the simulation results), then all the Assumptions A4–A6 can be regarded as indirectly verified when the algorithm yields a significant result, suggesting the presence of a causal path.

7.3. Comparison of the proposed algorithm and other cumulant-based causal learning algorithms

Our proposed algorithm is rooted in the framework of linear non-Gaussian models (Shimizu, 2022; Shimizu et al., 2006) and follows a line of causal discovery methods that utilize higher-order cumulant information (e.g., Chen & Chan, 2013; Chen et al., 2024; Wiedermann, 2022). Nevertheless, the algorithm introduced in this study differs from previous work in several important ways, and we illustrate this by comparing it with the algorithm of Chen et al. (2024) below.

As discussed earlier, both algorithms build on the concept of higher-order cumulants, and the same analytic result from our Theorem 1 had previously been used as Case 1 in the algorithm proposed by

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Chen et al. (2024). However, the two algorithms focus on fundamentally different substantive research problems. Specifically, Chen et al.'s (2024) work addresses causal structure learning (distinguishing between $x \to y$ and $y \to x$) in the presence of hidden confounder(s) with cross-sectional data. For Case 1, their algorithm uses a test statistic based on fourth-order joint cumulants (the same cumulants that form the basis of our Δ_1) as an initial step to detect the presence of a causal link. If a causal link is detected, the *direction of causality* is inferred by solving a quadratic equation using fifth-order joint cumulants (i.e., Case 2). In contrast, our study focuses on causal learning in longitudinal designs and aims to provide an innovative solution to a substantively important and challenging research question common in psychological studies: *Does x at Time 1 cause y at Time 2, or is the observed association solely due to hidden confounder(s)?*

In addition, whereas Chen et al.'s (2024) algorithm detects causal links using only fourth-order joint cumulants (Δ_1), our study derives additional test statistics (Δ_2 and Δ_3) that incorporate information from third-order joint cumulants. We further propose a multiple-testing procedure to enhance statistical power in detecting true longitudinal causal paths. Finally, regarding statistical inference, in Case 1, Chen et al.'s (2024) algorithm performs a single-sample sign test (Dixon & Mood, 1946) based on sample estimates of Δ_1 calculated from 100 resamples. The proposed test exhibits limited statistical foundation, as evidenced by the highly inflated Type I error rates shown in Figure 2 of their study. By contrast, our algorithm relies on 95% bootstrap confidence intervals, which appropriately control Type I error while preserving statistical power.

While it is important to recognize that the two algorithms focus on different research scenarios, making their performance not directly comparable from a substantive standpoint, differences in their behavior can still be observed from a statistical perspective. Based on the simulation results, we identify several differences between the performance of our algorithm and *Case 1* of Chen et al.'s (2024) approach. First, by using formal bootstrap-based statistical tests, our proposed algorithm does not exhibit the same issue of severely inflated Type I error rates observed in Chen et al.'s (2024) algorithm. Moreover, as higher-moment methods, both algorithms are expected to require larger sample sizes to achieve favorable performance. However, by incorporating information from third-order joint cumulants and conducting tests based on multiple test statistics, the sample size required for implementing our algorithm is substantially smaller than that recommended for Chen et al.'s (2024) algorithm, thus making our algorithm more practical for psychological studies.

7.4. Relationship between the proposed algorithm and other causal inference techniques

During the review process, an interesting question was raised about how the proposed cumulant-based algorithm compares with approaches more familiar to psychologists. More specifically, to address potential confounders (whether observed or unobserved) in observational studies, causal inference techniques such as propensity score methods (Thoemmes & Kim, 2011), instrumental variable estimation (Maydeu-Olivares et al., 2020), and sensitivity analysis (Bi et al., 2025) are well established and commonly used techniques. This leads to the question of why a new method is needed, particularly given that the proposed algorithm relies on the specific assumption of nonnormality and requires a large sample size. In this section, we explain the relationship between the proposed algorithm and other well-established causal inference techniques, with the aim of emphasizing the theoretical motivation for the proposed approach and providing practical guidelines for applied researchers in selecting appropriate methods.

First, the proposed algorithm and many classic techniques mentioned above aim to address different causal questions. For example, propensity score methods and instrumental variable analyses focus on estimating the magnitude of causal effects, assuming that the causal structure is already known. In contrast, the proposed algorithm, developed within the field of causal discovery or causal structure learning (Shimizu, 2019), seeks to identify the underlying causal structure, that is, the presence of the causal path.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

It is noted that the two distinct causal questions can be related. Specifically, by correctly estimating the true causal effect between two variables of interest (e.g., β_{yx}), researchers can also determine the presence of a causal path (i.e., by testing $H_0 : \beta_{yx} = 0$). In addition, sensitivity analyses provide a useful tool for assessing how strong the unmeasured confounder would need to be to negate the observed effect. However, correctly estimating the true causal effect is not a necessary condition for detecting the presence of a causal path. The proposed algorithm, along with other causal learning methods, offers a different perspective for testing whether a causal path or causal structure exists between variables without requiring consistent estimation of the target causal path.

Second, in studying causality, confounders – whether observed or unmeasured – are a central concern. Most existing techniques view confounders as undesirable factors to be mitigated. For example, propensity score methods focus on controlling for observed confounders. Unobserved confounders can also be addressed by incorporating valid instrumental variables, or the risk posed by them can be quantified through sensitivity analyses. In contrast, the proposed algorithm, as discussed earlier, explicitly leverages higher-order information carried by latent confounders, turning what is typically seen as a limitation into a valuable source of causal evidence.

In connection with the above points, by incorporating higher-order information from latent confounders, one assumption made by the proposed algorithm is that the latent confounder is nonnormal, which is technically an untestable and relatively uncommon assumption to make. Nevertheless, all other causal inference techniques also rely on distinct sets of assumptions. For instance,[12] propensity score methods assume that all relevant confounders are observed and appropriately specified in the propensity score model. Implementation of instrumental variable estimation, on the other hand, requires the availability of at least one valid instrumental variable. To make the results from sensitivity analysis interpretable, researchers typically need to assume and specify a plausible range for the strength of the unmeasured confounders. Technically speaking, the above assumptions are not testable and are not necessarily weaker – and may sometimes arguably be stronger – in psychological studies.

Moreover, in the context of causal structural learning, violations of key assumptions across different methods may lead to different types of errors. For example, using an invalid instrumental variable results in biased estimation of the true causal effect. Thus, when no causal path exists (i.e., the true causal effect is zero), instrumental variable analysis may still produce a biased, nonzero estimate. In other words, researchers may mistakenly conclude that a causal path exists when it does not. For the proposed algorithm, as discussed earlier, if the latent confounder is normally distributed, all three test statistics ($\Delta 1$, $\Delta 2$, and $\Delta 3$) are expected to be zero, indicating that researchers would fail to detect the causal path, regardless of whether it exists or not. In this sense, when key assumptions are at risk of being violated, the proposed algorithm can be considered more conservative in drawing conclusions about the presence of a causal link compared with other established causal inference techniques.

It should be acknowledged that the methods to which we compared our proposed approach were selected because they are commonly known and applied in psychological studies. We note that the alternative approaches discussed above are not meant to be exhaustive. Other alternative approaches exist that allow researchers either to estimate causal effects or to discover the presence of a causal path. For example, without distributional or linearity assumptions, the Fast Causal Inference (FCI) algorithm (Spirtes et al., 2001) enables researchers to identify certain local causal paths. Nevertheless, to detect the existence of a target causal path (i.e., $x \rightarrow y$), FCI requires additional conditions, such as particular conditional independence patterns among x, y, and other variables (e.g., so-called Y-structures), along with other standard assumptions underlying constraint-based causal discovery (e.g., faithfulness; Glymour et al., 2019).

The main point from the above comparison is that, objectively, all causal inference/learning methods have their own assumptions and appropriate application scenarios. We do not intend to suggest that the proposed algorithm is universally superior for identifying causal effects in longitudinal designs.

[12]Note that the assumptions listed are not exhaustive.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

In practice, we recommend that applied researchers select methods based on their research questions, study design, and the plausibility of underlying assumptions. Finally, the output from the proposed algorithm (as with other techniques) should not be interpreted as definitive evidence of longitudinal causal paths in the absence of substantive theories. Rather, the results are intended to help develop, complement, expand, and refine existing theories. As with any data-driven statistical technique, the integration of domain-specific knowledge and substantive theory into the decision-making process remains essential.

7.5. Limitations and future directions

We conclude this article by discussing limitations and outlining several future directions for our work. First, even though we conducted a large-scale simulation study with 1,260 unique conditions, there are many characteristics of the data-generating model that we did not consider in the simulation design. For example, we only generated nonnormal data following Gamma distributions. Future studies should examine the performance of the proposed algorithm under broader modeling conditions and data scenarios. Second, in this initial study, we estimated the sample high-joint cumulants based on the function of expected values provided in Equation (7). Methodologists have proposed alternative estimation methods for obtaining the higher-order cumulants, which could yield less biased results compared to the corresponding sample mean estimators, especially in small samples (e.g., Chan et al., 2020). Future studies should explore and compare different estimation methods for computing joint cumulants and test statistics, as well as their impact on the performance of the proposed algorithm.

Third, the proposed algorithm is developed within the framework of linear non-Gaussian models. Although the assumption of linearity (or approximate linearity) is consistent with widely accepted methodological practice and many empirical application scenarios (see Nickel et al., 2019), we acknowledge that in certain psychological applications, relationships between variables may be nonlinear or take more complex functional forms (Jebb et al., 2018). In our initial manuscript, we assumed that all relationships among u, x, and y are (approximately) linear. However, the proposed algorithm may also be applied even if the potential causal relationship between x and y is nonlinear. Specifically, when there is no causal link between x and y, the relationships given in Equations (13)–(17) still hold, implying that the test statistics are asymptotically zero, provided that the relationships between x and u and between y and u are approximately linear. When x causes y, the higher-order joint cumulants become more complicated with additional terms involving x, which can lead to nonzero test statistics (under different sets of faithfulness assumptions depending on the form of the causal relationship).

To verify this conjecture, we conducted a small additional simulation study in which the relationship between x and y was nonlinear; details are provided in the Supplementary Material. The results supported the applicability of the proposed algorithm in detecting the presence of a causal link between x and y (including nonlinear forms) while maintaining well-controlled Type I error rates. Future studies are expected to advance this line of work by extending cumulant-based approaches to more effectively accommodate complex nonlinear relationships (Hernandez-Lobato et al., 2016).

Fourth, in this study, we assumed that the association between the two variables of interest was not introduced by collider bias. Future research should extend the proposed algorithm to address situations involving potential collider bias, such as selection bias arising from the sampling process. Finally, this study focused on a simple design with two variables of interest, conducting longitudinal causal learning in a bivariate setting. The proposed algorithm can be naturally extended to applications with more complex longitudinal designs or modeling scenarios, such as longitudinal mediation models (Maxwell & Cole, 2007) and cross-lagged panel models (Zyphur et al., 2020). It is anticipated that future research will build upon the proposed algorithm and assess its potential in these more complex models.

Supplementary material. To view supplementary material for this article, please visit http://doi.org/10.1017/psy.2026.10100.

Data availability statement. Publicly available datasets were analyzed in this study. These data can be accessed at https://www.cdc.gov/brfss/annual_data/annual_data.htm (Empirical Example 1) and https://hrs.isr.umich.edu/ (for Empirical Example 2).

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Funding statement. This work is supported by an ASPIRE grant from the Office of the Vice President for Research at the University of South Carolina. D.S. would like to gratefully acknowledge the sabbatical leave granted to him by the University of South Carolina, which provided the time to help make this research possible.

Competing interests. The authors declare that they have no competing interests.

Disclosure of artificial intelligence-generated content (AIGC) tools. The authors produced and take full responsibility for all written content. Note that an AIGC tool (ChatGPT-4o) was used for several grammatical checks and language polishing.

References

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*(6), 893–897.

Bi, K., Merrin, G. J., Li, T., Sun, X., Chai, Y., Lu, Z., & Chen, M. S. (2025). Navigating unmeasured confounding in nonexperimental psychological research: A practical guide to computing and interpreting e-value. *Advances in Methods and Practices in Psychological Science*, *8*(2), 25152459251326571.

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology*, *9*(2), 78–84.

Blome, C., & Augustin, M. (2015). Measuring change in quality of life: Bias in prospective and retrospective evaluation. *Value in Health*, *18*(1), 110–115.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, *8*, 3–62.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152.

Brillinger, D. R. (2001). *Time series: Data analysis and theory*. Society for Industrial and Applied Mathematics.

Buecker, S., Maes, M., Denissen, J. J., & Luhmann, M. (2020). Loneliness and the big five personality traits: A meta–analysis. *European Journal of Personality*, *34*(1), 8–28.

Chan, L. H., Chen, K., Li, C., Wong, C. W., & Yau, C. Y. (2020). On higher-order moment and cumulant estimation. *Journal of Statistical Computation and Simulation*, *90*(4), 747–771.

Chen, W., Huang, Z., Cai, R., Hao, Z., & Zhang, K. (2024). Identification of causal structure with latent variables based on higher order cumulants. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(18), 20353–20361.

Chen, Z., & Chan, L. (2013). Causality in linear non-Gaussian acyclic models in the presence of latent Gaussian confounders. *Neural Computation*, *25*(6), 1605–1641.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cramér, H. (1970). *Random variables and probability distributions* (3rd ed.). Cambridge University Press.

D'Onofrio, B. M., Sjölander, A., Lahey, B. B., Lichtenstein, P., & Öberg, A. S. (2020). Accounting for confounding in observational studies. *Annual Review of Clinical Psychology*, *16*(1), 25–48.

Davison, A., & Hinkley, D. (1997). *Bootstrap methods and their application*. Cambridge University Press.

Dixon, W. J., & Mood, A. M. (1946). The statistical sign test. *Journal of the American Statistical Association*, *41*(236), 557–566.

Dodge, Y., & Rousson, V. (2000). Direction dependence in a regression line. *Communications in Statistics-Theory and Methods*, *29*(9–10), 1957–1972.

Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, *10*(524).

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Scholkopf, B., & Smola, A. J. (2008). A kernel statistical test of independence. *Advances in Neural Information Processing Systems*, *20*, 585–592.

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, *15*(5), 1243–1255.

Hakulinen, C., Elovainio, M., Pulkki-Råback, L., Virtanen, M., Kivimäki, M., & Jokela, M. (2015). Personality and depressive symptoms: Individual participant meta-analysis of 10 cohort studies. *Depression and Anxiety*, *32*(7), 461–470.

Hernandez-Lobato, D., Morales-Mombiela, P., Lopez-Paz, D., & Suarez, A. (2016). Non-linear causal inference using gaussianity measures. *Journal of Machine Learning Research*, *17*(28), 1–39.

Hofer-Szabó, G., Rédei, M., & Szabó, L. E. (2013). *The principle of the common cause*. Cambridge University Press.

Hoyer, P. O., Shimizu, S., Kerminen, A. J., & Palviainen, M. (2008). Estimation of causal effects using linear non-gaussian causal models with hidden variables. *International Journal of Approximate Reasoning*, *49*(2), 362–378.

Hughes, K., Bellis, M. A., Hardcastle, K. A., Sethi, D., Butchart, A., Mikton, C., … Dunne, M. P. (2017). The effect of multiple adverse childhood experiences on health: A systematic review and meta-analysis. *The Lancet Public Health*, *2*(8), e356–e366.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent components analysis*. Wiley.

Hyvärinen, A., & Smith, S. M. (2013). Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *Journal of Machine Learning Research*, *14*, 111–152.

Jebb, A. T., Tay, L., Diener, E., & Oishi, S. (2018). Happiness, income satiation and turning points around the world. *Nature Human Behaviour*, *2*(1), 33–38.

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, *114*(1–3), 163–173.

Lyapunov, A. M. (1901). Nouvelle forme du théorème sur la limite de probabilité. Mémoires de l'Académie Impériale des Sciences de St.-Pétersbourg VIIIe Série. *Classe Physico-Mathématique*, *12*, 1–24.

Maeda, T. N., & Shimizu, S (2020). RCD: Repetitive causal discovery of linear non-Gaussian acyclic models with latent confounders. In *International Conference on artificial intelligence and statistics* (pp. 735–745). PMLR

Mann, F., Wang, J., Pearce, E., Ma, R., Schlief, M., Lloyd-Evans, B., … Johnson, S. (2022). Loneliness and the onset of new mental health problems in the general population. *Social Psychiatry and Psychiatric Epidemiology*, *57*(11), 2161–2178.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23–44.

Maydeu-Olivares, A., Shi, D., & Fairchild, A. J. (2020). Estimating causal effects in linear regression models with observational data: The instrumental variables regression model. *Psychological Methods*, *25*(2), 243–258.

Merrick, M. T., Ford, D. C., Ports, K. A., & Guinn, A. S. (2018). Prevalence of adverse childhood experiences from the 2011–2014 behavioral risk factor surveillance system in 23 states. *JAMA Pediatrics*, *172*(11), 1038–1044.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*(1), 156–166.

Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *The Journal of Machine Learning Research*, *17*(1), 1103–1204.

Nickel, L. B., Roberts, B. W., & Chernyshenko, O. S. (2019). No evidence of a curvilinear relation between conscientiousness and relationship, work, and health outcomes. *Journal of Personality and Social Psychology*, *116*(2), 296–312.

Park, C., Majeed, A., Gill, H., Tamura, J., Ho, R. C., Mansur, R. B., … McIntyre, R. S. (2020). The effect of loneliness on distinct health outcomes: A comprehensive review and meta- analysis. *Psychiatry Research*, *294*, 113514.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Peters, J., Mooij, J. M., Janzing, D., & Schölkopf, B. (2014). Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, *15*, 2009–2053.

Pocock, S. J., Collier, T. J., Dandreo, K. J., de Stavola, B. L., Goldman, M. B., Kalish, L. A., … McCormack, V. A. (2004). Issues in the reporting of epidemiological studies: A survey of recent practice. *BMJ*, *329*(7471), 883.

R Core Team. (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org/

Reichenbach, H. (1956). *The direction of time*. University of California Press.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688.

Russell, D., Peplau, L. A., & Cutrona, C. E. (1980). The revised UCLA loneliness scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology*, *39*(3), 472.

Rutter, M. (1994). Beyond longitudinal data: Causes, consequences, changes, and continuity. *Journal of Consulting and Clinical Psychology*, *62*(5), 928–940.

Schultheiss, C., Bühlmann, P., & Yuan, M. (2024). Higher-order least squares: Assessing partial goodness of fit of linear causal models. *Journal of the American Statistical Association*, *119*(546), 1019–1031. https://doi.org/10.1080/01621459.2022.2157728

Shi, D., Fairchild, A. J., & Wiedermann, W. (2023). One step at a time: A statistical approach for distinguishing mediators, confounders, and colliders using direction dependence analysis (DDA). *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000619.

Shi, D., Zhang, B., Wiedermann, W., & Fairchild, A. J. (2025). Distinguishing cause from effect in psychological research: An independence-based approach under linear non-Gaussian models. *British Journal of Mathematical and Statistical Psychology*, *78*(3), 965–995.

Shimizu, S. (2019). Non-Gaussian methods for causal structure learning. *Prevention Science*, *20*, 431–441.

Shimizu, S. (2022). *Statistical causal discovery: LiNGAM approach*. Springer.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., Kerminen, A., & Jordan, M. (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, *7*(10), 2003–2030.

Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Washio, T., … Bollen, K. (2011). DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, *12*, 1225–1248.

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (2nd ed.). MIT Press.

Stuart, A., & Ord, J. K. (1987). *Kendall's advanced theory of statistics*. Oxford University Press, Inc.

Sungur, E. A. (2005). A note on directional dependence in regression setting. *Communications in Statistics-Theory and Methods*, *34*(9–10), 1957–1965.

Tan, M., & Mao, P. (2023). Type and dose-response effect of adverse childhood experiences in, predicting depression: A systematic review and meta-analysis. *Child Abuse & Neglect*, *139*, 106091.

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, *46*(1), 90–118.

Tramontano, D., Kivva, Y., Salehkaleybar, S., Drton, M., & Kiyavash, N. (2024). Causal effect identification in LiNGAM models with latent confounders. In *Proceedings of the 41st International Conference on Machine Learning* (Vol. 235, Article 1982). JMLR.org. Vienna, Austria.

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Van Santvoort, F., Hosman, C. M., Janssens, J. M., Van Doesum, K. T., Reupert, A., & Van Loon, L. M. (2015). The impact of various parental mental disorders on children's diagnoses: A systematic review. *Clinical Child and Family Psychology Review*, *18*(4), 281–299.

von Eye, A., & DeShon, R. P. (2012). Directional dependence in developmental research. *International Journal of Behavioral Development*, *36*(4), 303–312.

Walsh, C., MacMillan, H., & Jamieson, E. (2002). The relationship between parental psychiatric disorder and child physical and sexual abuse: Findings from the Ontario Health Supplement. *Child Abuse & Neglect*, *26*(1), 11–22.

Wiedermann, W. (2022). Third moment-based causal inference. *Behaviormetrika*, *49*(2), 303–328.

Wiedermann, W., & Li, X. (2018). Direction dependence analysis: A framework to test the direction of effects in linear models with an implementation in SPSS. *Behavior Research Methods*, *50*(4), 1581–1601.

Wiedermann, W., & Li, X. (2020). Confounder detection in linear mediation models: Performance of kernel-based tests of independence. *Behavior Research Methods*, *52*, 342–359.

Wiedermann, W., Li, X., & von Eye (2021). Direction dependence analysis: Statistical foundations and applications. In W. Wiedermann, D. Kim, E. A. Sungur, & A. von Eye (Eds.), *Direction dependence in statistical modeling: Methods of analysis* (pp. 9–46). Wiley & Sons.

Wiedermann, W., & Sebastian, J. (2020). Direction dependence analysis in the presence of confounders: Applications to linear mediation models using observational data. *Multivariate Behavioral Research*, *55*(4), 495–515.

Wiedermann, W., & von Eye, A. (2015). Direction-dependence analysis: A confirmatory approach for testing directional theories. *International Journal of Behavioral Development*, *39*(6), 570–580.

Wiedermann, W., & von Eye, A. (2025). *Direction dependence analysis: Foundations and statistical methods*. Cambridge University Press.

Zhang, X., Astivia, O. L. O., Kroc, E., & Zumbo, B. D. (2023). How to think clearly about the central limit theorem. *Psychological Methods*, *28*(6), 1427–1445.

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., … Diener, E. (2020). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*, *23*(4), 651–687.

Appendix A: Proofs for Lemma 1

Lemma 1. *Under Model II (defined by Equations (3) and (4) and assumptions (A2)–(A5), we obtain the following results for the third- and fourth-order joint cumulants of observed variables x and y:*

$$Cum(x,x,y) = \beta_1^2(\beta_2 + \eta\beta_1)\,Cum_3(u) + \eta Cum_3(e_x) \tag{L1}$$

$$Cum(x,y,y) = \beta_1(\beta_2 + \eta\beta_1)^2 Cum_3(u) + \eta^2 Cum_3(e_x) \tag{L2}$$

$$Cum(x,x,x,y) = \beta_1^3(\beta_2 + \eta\beta_1)\,Cum_4(u) + \eta Cum_4(e_x) \tag{L3}$$

$$Cum(x,x,y,y) = \beta_1^2(\beta_2 + \eta\beta_1)^2 Cum_4(u) + \eta^2 Cum_4(e_x) \tag{L4}$$

$$Cum(x,y,y,y) = \beta_1(\beta_2 + \eta\beta_1)^3 Cum_4(u) + \eta^3 Cum_4(e_x) \tag{L5}$$

A Proof for (L1)

From Equations (3) and (4), we obtain

$$\begin{aligned} Cum(x,x,y) &= \mathbb{E}\left[x^2 y\right] = \mathbb{E}\left[(\beta_1 u + e_x)^2(\beta_2 u + \eta x + e_y)\right] \\ &= \mathbb{E}\left[\eta(e_x)^3 + (e_x)^2 e_y + \beta_2(e_x)^2 u + 3\beta_1\eta(e_x)^2 u + 2\beta_1 e_x e_y u + 2\beta_1\beta_2 e_x u^2 + 3\beta_1^2\eta e_x u^2 + \beta_1^2 e_y u^2 + \beta_1^2\beta_2 u^3 + u^3\right] \\ &= \eta\mathbb{E}\left[(e_x)^3\right] + \mathbb{E}\left[(e_x)^2 e_y\right] + (\beta_2 + 3\beta_1\eta)\,\mathbb{E}\left[(e_x)^2 u\right] + 2\beta_1\mathbb{E}\left[e_x e_y u\right] \\ &\quad + (2\beta_1\beta_2 + 3\beta_1^2\eta)\,\mathbb{E}\left[e_x u^2\right] + \beta_1^2\mathbb{E}\left[e_y u^2\right] + (\beta_1^2\beta_2 + \beta_1^3\eta)\,\mathbb{E}\left[u^3\right]. \end{aligned}$$

Given the independence assumption (A3), $Cum(x,x,y)$ can be further simplified as

$$\begin{aligned} Cum(x,x,y) &= \beta_1^2(\beta_2 + \beta_1\eta)\,\mathbb{E}\left[u^3\right] + \eta\mathbb{E}\left[(e_x)^3\right] \\ &= \beta_1^2(\beta_2 + \eta\beta_1)\,Cum_3(u) + \eta Cum_3(e_x). \end{aligned}$$

A Proof for (L2)

Repeating the procedure and assumptions used in the proof for (L1), we can obtain

$$\begin{aligned} Cum(x,y,y) &= \mathbb{E}\left[xy^2\right] = \mathbb{E}\left[(\beta_1 u + e_x)(\beta_2 u + \eta x + e_y)^2\right] \\ &= \beta_1(\beta_2 + \eta\beta_1)^2\mathbb{E}\left[u^3\right] + \eta^2\mathbb{E}\left[(e_x)^3\right] \\ &= \beta_1(\beta_2 + \eta\beta_1)^2 Cum_3(u) + \eta^2 Cum_3(e_x). \end{aligned}$$

A Proof for (L3)

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

From Equations (3) and (4), we obtain

$$\begin{aligned}
Cum(x,x,x,y) &= \mathbb{E}\left[x^3y\right] - 3\mathbb{E}\left[xy\right]\mathbb{E}\left[x^2\right] = \mathbb{E}\left[\left(\beta_1 u + e_x\right)^3\left(\beta_2 u + \eta x + e_y\right)\right] - 3\mathbb{E}\left[\left(\beta_1 u + e_x\right)\left(\beta_2 u + \eta x + e_y\right)\right]\mathbb{E}\left[\left(\beta_1 u + e_x\right)^2\right] \\
&= \mathbb{E}\left[\eta(e_x)^4 + (e_x)^3 e_y + \beta_2 (e_x)^3 u + 4\beta_1\eta(e_x)^3 u + 3\beta_1 (e_x)^2 e_y u + 3\beta_1\beta_2 (e_x)^2 u^2\right. \\
&\quad + 6\beta_1^2\eta(e_x)^2 u^2 + 3\beta_1^2 e_x e_y u^2 + 3\beta_1^2\beta_2 e_x u^3 + 4\beta_1^3\eta e_x u^3 + \beta_1^3 e_y u^3 + \beta_1^3\beta_2 u^4 + \beta_1^4\eta u^4\Big] \\
&\quad - 3\mathbb{E}\left[\eta(e_x)^2 + e_x e_y + \beta_2 e_x u + 2\beta_1\eta e_x u + \beta_1 e_y u + \beta_1\beta_2 u^2 + \beta_1^2\eta u^2\right]\mathbb{E}\left[(e_x)^2 + 2\beta_1 e_x u + \beta_1^2 u^2\right] \\
&= \eta\mathbb{E}\left[(e_x)^4\right] + \mathbb{E}\left[(e_x)^3 e_y\right] + (\beta_2 + 4\beta_1\eta)\,\mathbb{E}\left[(e_x)^3 u\right] + 3\beta_1\mathbb{E}\left[(e_x)^2 e_y u\right] \\
&\quad + \left(3\beta_1\beta_2 + 6\beta_1^2\eta\right)\mathbb{E}\left[(e_x)^2 u^2\right] + 3\beta_1^2\mathbb{E}\left[e_x e_y u^2\right] + \left(3\beta_1^2\beta_2 + 4\beta_1^3\eta\right)\mathbb{E}\left[e_x u^3\right] + \beta_1^3 E\left[e_y u^3\right] \\
&\quad + \left(\beta_1^3\beta_2 + \beta_1^4\eta\right)E\left[u^4\right] - 3\eta\left\{\mathbb{E}\left[(e_x)^2\right]\right\}^2 - \left(3\beta_1\beta_2 + 6\beta_1^2\eta\right)\mathbb{E}\left[(e_x)^2 u^2\right] - 3\left(\beta_1^3\beta_2 + \beta_1^4\eta\right)\left\{E\left[u^2\right]\right\}^2
\end{aligned}$$

Given the independence assumption (A3), $Cum(x,x,x,y)$ can be further simplified as

$$\begin{aligned}
Cum(x,x,x,y) &= \beta_1^3(\beta_2 + \eta\beta_1)\,\mathbb{E}\left[u^4\right] - 3\beta_1^3(\beta_2 + \eta\beta_1)\left\{\mathbb{E}\left[u^2\right]\right\}^2 \\
&\quad + \eta\mathbb{E}\left[(e_x)^4\right] - 3\eta\left\{\mathbb{E}\left[(e_x)^2\right]\right\}^2 \\
&= \beta_1^3(\beta_2 + \eta\beta_1)\,Cum_4(u) + \eta Cum_4(e_x).
\end{aligned}$$

A Proof for (L4)

Repeating the procedure and assumptions used in the proof for (L3), we can obtain

$$\begin{aligned}
Cum(x,x,y,y) &= \mathbb{E}\left[x^2y^2\right] - 2\left\{\mathbb{E}\left[xy\right]\right\}^2 - \mathbb{E}\left[x^2\right]\mathbb{E}\left[y^2\right] = \mathbb{E}\left[\left(\beta_1 u + e_x\right)^2\left(\beta_2 u + \eta x + e_y\right)^2\right] \\
&\quad - 2\mathbb{E}\left\{\left(\beta_1 u + e_x\right)\left(\beta_2 u + \eta x + e_y\right)\right\}^2 - \mathbb{E}\left[\left(\beta_1 u + e_x\right)^2\right]\mathbb{E}\left[\left(\beta_2 u + \eta x + e_y\right)^2\right] \\
&= \beta_1^2(\beta_2 + \eta\beta_1)^2\mathbb{E}\left[u^4\right] - 3\beta_1^2(\beta_2 + \eta\beta_1)^2\left\{\mathbb{E}\left[u^2\right]\right\}^2 + \eta^2\mathbb{E}\left[(e_x)^4\right] - 3\eta^2\left\{\mathbb{E}\left[(e_x)^2\right]\right\}^2 \\
&= \beta_1^2(\beta_2 + \eta\beta_1)^2 Cum_4(u) + \eta^2 Cum_4(e_x)
\end{aligned}$$

A Proof for (L5)

Repeating the procedure and assumptions used in the proof for (L3), we can obtain

$$\begin{aligned}
Cum(x,y,y,y) &= \mathbb{E}\left[xy^3\right] - 3\mathbb{E}\left[xy\right]\mathbb{E}\left[y^2\right] = \mathbb{E}\left[\left(\beta_1 u + e_x\right)\left(\beta_2 u + \eta x + e_y\right)^3\right] \\
&\quad - 3\mathbb{E}\left[\left(\beta_1 u + e_x\right)\left(\beta_2 u + \eta x + e_y\right)\right]\mathbb{E}\left[\left(\beta_2 u + \eta x + e_y\right)^2\right] \\
&= \beta_1(\beta_2 + \eta\beta_1)^3\mathbb{E}\left[u^4\right] - 3\beta_1(\beta_2 + \eta\beta_1)^3\left\{\mathbb{E}\left[u^2\right]\right\}^2 + \eta^3\mathbb{E}\left[(e_x)^4\right] - 3\eta^3\left\{\mathbb{E}\left[(e_x)^2\right]\right\}^2 \\
&= \beta_1(\beta_2 + \eta\beta_1)^3 Cum_4(u) + \eta^3 Cum_4(e_x).
\end{aligned}$$

Appendix B: A proof for Theorem 2

Theorem 2. *Let* $\Delta_2 = Cum(x,x,y) \times Cum(x,x,y,y) - Cum(x,y,y) \times Cum(x,x,x,y)$. *When no true causal effect from x to y exists (i.e.,* $\eta = 0$*), one obtains* $\Delta_2 = 0$. *When x truly causes y* ($\eta \neq 0$), *with the faithfulness assumption that* $\eta \neq -\frac{\beta_2}{\beta_1}$, *one obtains* $\Delta_2 \neq 0$.

Proof. Given Equations (8)–(11), we obtain

$$\begin{aligned}
\Delta_2 &= Cum(x,x,y) \times Cum(x,x,y,y) - Cum(x,y,y) \times Cum(x,x,x,y) \\
&= \left[\beta_1^2(\beta_2 + \eta\beta_1)\,Cum_3(u) + \eta Cum_3(e_x)\right] \times \left[\beta_1^2(\beta_2 + \eta\beta_1)^2 Cum_4(u) + \eta^2 Cum_4(e_x)\right] \\
&\quad - \left[\beta_1(\beta_2 + \eta\beta_1)^2 Cum_3(u) + \eta^2 Cum_3(e_x)\right] \times \left[\beta_1^3(\beta_2 + \eta\beta_1)\,Cum_4(u) + \eta Cum_4(e_x)\right]
\end{aligned} \tag{B1}$$

By setting $\Delta_1 = 0$ and simplifying the result, we obtain the following quadratic equation for η, assuming $Cum_4(u) \neq 0$, $Cum_4(e_x) \neq 0$, $\beta_1 \neq 0$ and $\beta_2 \neq 0$:

$$\left[Cum_4(u)Cum_3(e_x)\beta_1^3\beta_2 - Cum_3(u)Cum_4(e_x)\beta_1^2\beta_2\right]\eta^2 + \left[Cum_4(u)Cum_3(e_x)\beta_1^2\beta_2^2 - Cum_3(u)Cum_4(e_x)\beta_1\beta_2^2\right]\eta = 0 \tag{B2}$$

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press

Solving Equation (B2) for η, the only two solutions are

$$\eta_1 = 0$$
$$\eta_2 = -\frac{\beta_2}{\beta_1}.$$

Therefore, when $\eta = 0$, it follows that $\Delta_1 = 0$ and, assuming faithfulness holds (i.e., $\eta \neq -\frac{\beta_2}{\beta_1}$), when $\eta \neq 0$, then $\Delta_2 \neq 0$.

Appendix C: A proof for Theorem 3

Theorem 3. *Let* $\Delta_3 = Cum(x,x,y) \times Cum(x,y,y,y) - Cum(x,y,y) \times Cum(x,x,y,y)$. *When no true causal effect from x to y exists (i.e.,* $\eta = 0$*), one obtains* $\Delta_3 = 0$. *When x truly causes y (*$\eta \neq 0$*), with the faithfulness assumption that* $\eta \neq -\frac{\beta_2}{\beta_1}$ *and* $\eta \neq -\frac{Cum_4(u)Cum_3(e_x)\beta_2}{Cum_4(u)Cum_3(e_x)\beta_1 - Cum_3(u)Cum_4(e_x)}$, *one obtains* $\Delta_3 \neq 0$.

Proof. Given Equations (8), (9), (11), and (12), we obtain

$$\begin{aligned}\Delta_3 &= Cum(x,x,y) \times Cum(x,y,y,y) - Cum(x,y,y) \times Cum(x,x,y,y)\\ &= \left[\beta_1^2(\beta_2 + \eta\beta_1)Cum_3(u) + \eta Cum_3(e_x)\right] \times \left[\beta_1(\beta_2 + \eta\beta_1)^3 Cum_4(u) + \eta^3 Cum_4(e_x)\right]\\ &\quad - \left[\beta_1(\beta_2 + \eta\beta_1)^2 Cum_3(u) + \eta^2 Cum_3(e_x)\right] \times \left[\beta_1^2(\beta_2 + \eta\beta_1)^2 Cum_4(u) + \eta^2 Cum_4(e_x)\right] \qquad \text{(C1)}\end{aligned}$$

By setting $\Delta_1 = 0$ and simplifying the result, we obtain the following cubic equation for η, assuming $Cum_4(u) \neq 0$, $Cum_4(e_x) \neq 0$, $\beta_1 \neq 0$ and $\beta_2 \neq 0$:

$$\begin{aligned}&\left[Cum_4(u)Cum_3(e_x)\beta_1^3\beta_2 - Cum_3(u)Cum_4(e_x)\beta_1^2\beta_2\right]\eta^3\\ &\quad + \left[2Cum_4(u)Cum_3(e_x)\beta_1^2\beta_2^2 - Cum_3(u)Cum_4(e_x)\beta_1\beta_2^2\right]\eta^2\\ &\quad + Cum_4(u)Cum_3(e_x)\beta_1\beta_3^3\eta \qquad \text{(C2)}\end{aligned}$$

Solving Equation (C2) for η, the only two solutions are

$$\eta_1 = 0$$
$$\eta_2 = -\frac{\beta_2}{\beta_1}$$
$$\eta_3 = -\frac{Cum_4(u)Cum_3(e_x)\beta_2}{Cum_4(u)Cum_3(e_x)\beta_1 - Cum_3(u)Cum_4(e_x)}.$$

Therefore, when $\eta = 0$, it follows that $\Delta_3 = 0$ and, assuming faithfulness holds (i.e., $\eta \neq -\frac{\beta_2}{\beta_1}$ and $\eta \neq -\frac{Cum_4(u)Cum_3(e_x)\beta_2}{Cum_4(u)Cum_3(e_x)\beta_1 - Cum_3(u)Cum_4(e_x)}$), when $\eta \neq 0$, then $\Delta_3 \neq 0$.

Cite this article: Shi, D., Wiedermann, W., Fairchild, A. J. and Zhang, B. (2026). Does X at Time 1 Cause Y at Time 2? Longitudinal Causal Learning with Hidden Confounders. *Psychometrika*, 1092–1126. https://doi.org/10.1017/psy.2026.10100

https://doi.org/10.1017/psy.2026.10100 Published online by Cambridge University Press