Type-II Errors of Independence Tests Can Lead to Arbitrarily Large Errors in Estimated Causal Effects: An Illustrative Example

Nicholas Cornia & Joris M. Mooij
Informatics Institute
University of Amsterdam, The Netherlands
{n.cornia,j.m.mooij}@uva.nl

Abstract

Estimating the strength of causal effects from observational data is a common problem in scientific research. A popular approach is based on exploiting observed conditional independences between variables. It is well-known that this approach relies on the assumption of faithfulness. In our opinion, a more important practical limitation of this approach is that it relies on the ability to distinguish independences from (arbitrarily weak) dependences. We present a simple analysis, based on purely algebraic and geometrical arguments, of how the estimation of the causal effect strength, based on conditional independence tests and background knowledge, can have an arbitrarily large error due to the uncontrollable type II error of a single conditional independence test. The scenario we are studying here is related to the LCD algorithm by Cooper [1] and to the instrumental variable setting that is popular in epidemiology and econometry. It is one of the simplest settings in which causal discovery and prediction methods based on conditional independences arrive at non-trivial conclusions, yet for which the lack of uniform consistency can result in arbitrarily large prediction errors.

Introduction

Inferring causation from observational data is a common problem in several fields, such as biology and economics. To deal with the presence of unmeasured confounders of observed random variables the so-called *instrumental variable* technique [2] has found applications in genetics [3], epidemiology [4, 5] and economics [6]. Given two observable random variables possibly influenced by a hidden confounder, an *instrumental variable* is a third observed variable which is assumed to be independent of the confounder. In practice it is difficult to decide whether the instrumental variable definition is satisfied, and the method has aroused some skepticism [7]. In this paper, we study a setting that is similar in spirit to the instrumental variable model, but where all conditional independence assumptions are directly testable on the observed data. A similar scenario was first studied by Cooper [1] and independently rediscovered in the context of genome biology by Chen *et al.* [8].

An important assumption in causal discovery methods based on conditional independences is *faithfulness*, which means that the observed joint distribution does not contain any additional (conditional) independences beyond those induced by the causal structure. Usually, faithfulness is justified by the assumption that unfaithful distributions are a set of Lebesgue measure zero in the set of the model parameters. By showing that one can create a sequence of faithful distributions which converges to an unfaithful one, Robins *et al.* proved the lack of uniform consistency of causal discovery algorithms [9]. Zhang and Spirtes [10] then introduced the "Strong Faithfulness" assumption to recover the uniform consistency of causal discovery. Using geometric and combinatorial arguments, Uhler *et al.* [11] addressed the question of how restrictive the Strong Faithfulness assumption is in terms of the volume of distributions that do not satisfy this assumption. Even for a modest number of nodes and for sparse graphs, the "not strongly faithful" regions can be surprisingly large, and Uhler *et al.* argue that this result should discourage the use of large scale causal algorithms based on conditional independence tests, such as the PC and FCI algorithms [12].

In this work, we analyse in the context of the LCD setting how an error in a *single* conditional independence test may already lead to arbitrarily large errors in predicted causal effect strengths, even when

the faithfulness assumption is *not* violated. Our results may not be surprising for those familiar with the work of [9], but we believe that the analysis we present here may be easier to understand to those without a background in statistics, as we separate statistical issues (the possibility of type II errors in the conditional independence test from a finite sample) from a rather straightforward analysis of the problem in the population setting. We use an algebraic approach, showing how causal prediction may lead to wrong predictions already in the simple context of linear structural equation models with a multivariate Gaussian distribution.

In Section 1, we begin with a brief description of the problem setting in a formal way, giving the definitions of the causal effect, instrumental variable, LCD algorithm and the toy model we present. We consider three observed random variables (X_1, X_2, X_3), which is the minimal number such that a non-trivial conditional independence test can be obtained. In Section 2, we show how an (arbitrarily weak) conditional dependence that goes undetected can influence our estimation of the causal effect of X_2 on X_3 from the observed covariance matrix, when a confounder between X_2 and X_3 is almost off-set by a direct effect from X_1 to X_3. In fact, we show that this phenomenon can lead to an arbitrarily large error in the estimated causal effect as the noise variance of X_2 approaches zero. We finish with conclusions in Section 3.

1   Problem setting

1.1   LCD algorithm

The model we are interested in arises from the work of Cooper [1], who proposed the "LCD" algorithm for causal discovery in observational databases and the more recent paper of Chen et al.[8], who proposed the "Trigger" algorithm to infer transcriptional regulatory networks among genes. Throughout this section we will assume:

- Acyclicity;
- No Selection Bias.

Definition 1.1. (LCD setting) Given three random variables X_1, X_2, X_3 such that the following statistical properties and prior assumptions are satisfied:
Statistical dependences:

- $X_1 \not\perp\!\!\!\perp X_2$
- $X_2 \not\perp\!\!\!\perp X_3$
- $X_1 \perp\!\!\!\perp X_3 | X_2$

Prior assumptions:

- $An(X_1) \cap \{X_2, X_3\} = \emptyset$
- Faithfulness

where $An(X)$ is the set of the causal ancestors of X (which includes X itself), so this condition means that we assume that X_1 is not caused by the other observed variables X_2, X_3.

Cooper [1] proved that:

Theorem 1.1. *Under the assumptions in Definition 1.1, the causal structure must be a subgraph of:*

$$X_1 \overset{\leftrightarrow}{\to} X_2 \to X_3$$

Here, the directed arrows indicate a direct causal relationship and the bidirected edge denotes an unobserved confounder. □

Our primary interest is to predict $p(X_3|do(X_2))$, the distribution of X_3 after an intervention on X_2. In general, this quantity may differ from $p(X_3|X_2)$, the conditional distribution of X_3 given X_2 [13]. In the linear-Gaussian case, the quantity

$$\frac{\partial \mathbb{E}(X_3|do(X_2))}{\partial X_2}$$

measures the causal effect of X_2 on X_3.

It is easy to show that in the LCD setting, these quantities are equal:

Corollary 1.1. *Under the LCD assumptions in Definition 1.1,*

$$p(X_3|do(X_2)) = p(X_3|X_2).$$

Therefore, in the linear-Gaussian case, the quantity

$$\frac{\partial \mathbb{E}(X_3|do(X_2))}{\partial X_2} = \frac{\partial \mathbb{E}(X_3|X_2)}{\partial X_2} = \frac{\mathbb{Cov}(X_3, X_2)}{\mathbb{Var}(X_2)} \quad (1)$$

is a valid estimator for the causal effect of X_2 on X_3.

1.2   Relationship with instrumental variables

The other model relevant for our discussion is the so called *instrumental variable* model. Following Pearl [13], we define:

Definition 1.2. (Instrumental Variable setting) Given three random variables X_1, X_2, X_3, we call X_1 an *instrumental variable* if the following conditions are satisfied:
Statistical dependences:

- $X_1 \not\perp\!\!\!\perp X_2$

Prior assumptions:

- $X_1 \perp\!\!\!\perp X_3 | do(X_2)$
- Faithfulness

The second assumption says that X_1 and X_3 are independent after an intervention on the variable X_2. In terms of the causal graph, this means that all the unblocked paths between X_1 and X_3 contain an arrow that points to X_2.

Unfortunately the instrumental variable property cannot be directly tested from observed data. The causal graph for the IV setting is a subgraph of:

$$X_1 \overset{\leftrightarrow}{\to} X_2 \overset{\leftrightarrow}{\to} X_3$$

So, a possible confounder between X_2 and X_3 is allowed, in contrast with the LCD setting. Note that the LCD setting is a special case of the IV model.

Lemma 1.1. *Under the IV assumptions in Definition 1.2 and for the linear-Gaussian case, the quantity*

$$\frac{\mathbb{Cov}(X_1, X_3)}{\mathbb{Cov}(X_1, X_2)}$$

is a valid estimator for the causal effect of X_2 on X_3.

1.3 Type II errors in LCD

In practice, the confidence on the result of the conditional independence test $X_1 \perp\!\!\!\perp X_3 | X_2$ in the LCD setting depends on the sample size. Indeed, it could be hard to distinguish a weak conditional dependence
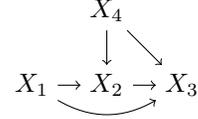
$$X_1 \not\perp\!\!\!\perp X_3 | X_2$$

from a conditional independence using a sample of finite size. Here we study the question of what happens to our prediction of the causal effect of X_2 on X_3 if the conditional independence test encounters a type II error (i.e., erroneously accepts the null hypothesis of independence).

Note that a type I error (i.e., erroneously rejecting the null hypothesis of independence) in the tests $X_1 \not\perp\!\!\!\perp X_2$ and $X_2 \not\perp\!\!\!\perp X_3$ will not be as dangerous as a type II error in the conditional independence test. Indeed, the probability of a type I error can be made arbitrarily small by tuning the significance level appropriately. In addition, a type I error would let the LCD algorithm reject a valid triple, i.e., lower the recall instead of leading to wrong predictions.

For these reasons we study the model described in the following definition, which allows the presence of a hidden confounder X_4, and a direct effect from X_1 on X_3 (not mediated via X_2). We assume that these additional features result in a possible weak conditional dependence between X_1 and X_3 given X_2. For simplicity we consider only the linear-Gaussian case. We also assume no confounders between X_1 and X_2, or

between X_1 and X_3, or between X_1, X_2, X_3. This simplification will not influence the final result of the paper, because we will prove how unboundedness of the causal effect estimation error is already achieved for this special case.

Definition 1.3. We assume that the "true" causal model has the following causal graph:

which is one of the possible causal structures that is compatible with the following conditions:

Statistical dependences:

- $X_1 \not\perp\!\!\!\perp X_2$
- $X_2 \not\perp\!\!\!\perp X_3$
- **A weak conditional dependence**

$$X_1 \not\perp\!\!\!\perp X_3 | X_2$$

Prior assumptions:

- Faithfulness
- $An(X_1) \cap \{X_2, X_3\} = \emptyset$

The observed random variables are X_1, X_2, X_3 while X_4 is a hidden confounder, assumed to be independent from X_1.

The joint distribution of the observed variables is assumed to be a multivariate Gaussian distribution with covariance matrix Σ and zero mean vector. We also assume that the structural equations of the model are linear. Then

$$X = AX + E, \tag{2}$$

where

$$X = (X_1, \ldots, X_4)^T$$

is the vector of the extended system,

$$E = (E_1, \ldots, E_4)^T$$

is the vector of the independent noise terms, such that

$$E \sim \mathcal{N}(0, \Delta) \quad \Delta = diag(\delta_i^2),$$

and $A = (\alpha_{ij}) \in \mathcal{M}_4(\mathbb{R})$ is (up to a permutation of indices) a real upper triangular matrix in the space $\mathcal{M}_4(\mathbb{R})$ of real 4×4 matrices that defines the causal strengths between the random variables of the system.

Remark 1.1. In [14], an implicit representation for the confounder X_4 is used, by using non-zero covariance between the noise variables E_2, E_3. It can be shown that for our purposes, the two representations are equivalent and yield the same conclusions.

In the Gaussian case, a conditional independence is equivalent to a vanishing partial correlation:

Lemma 1.2. *Given a set of three random variables* (X_1, X_2, X_3) *with a multivariate Gaussian distribution the conditional independence*

$$X_1 \perp\!\!\!\perp X_3 \mid X_2$$

is equivalent to a vanishing partial correlation

$$\rho_{13\cdot 2} = \frac{\rho_{13} - \rho_{12}\rho_{23}}{\sqrt{\left(1 - \rho_{12}^2\right)\left(1 - \rho_{23}^2\right)}} = 0 \qquad (3)$$

where ρ_{ij} *is the correlation coefficient of* X_i *and* X_j.

In the model described in Definition 1.3,

$$\frac{\partial \mathbb{E}(X_3|do(X_2))}{\partial X_2} = \alpha_{23}. \qquad (4)$$

In contrast with the LCD model in Definition 1.1, the equality (1) no longer holds. We are interested in the error in the estimation of the effect of X_2 on X_3 that would be due to a type II error of the conditional independence test in the LCD algorithm. The next section is dedicated to the analysis of the difference between the true value (4) and the estimated one in (1):

$$\left|\mathbb{E}\big(X_3|X_2\big) - \mathbb{E}\big(X_3|do(X_2)\big)\right| = |g\big(A, \Sigma\big)||X_2|,$$

where the "causal effect estimation error" is given by:

$$g\big(A, \Sigma\big) = \frac{\Sigma_{32}}{\Sigma_{22}} - \alpha_{23}. \qquad (5)$$

2 Estimation of the causal effect error from the observed covariance matrix

The following proposition gives a set of equations for the observed covariance matrix Σ, given the model parameters $\big(A, \Delta\big)$ and the linear structural equation model (2).

Proposition 2.1. *The mapping* $\Phi : (A, \Delta) \mapsto \Sigma$ *that maps model parameters* (A, Δ) *to the observed covariance matrix* Σ *according to the model in Definition 1.3 is given by:*

$$\Sigma_{11} = \delta_1^2 \qquad (6)$$

$$\Sigma_{12} = \alpha_{12}\delta_1^2 \qquad (7)$$

$$\Sigma_{13} = \big(\alpha_{13} + \alpha_{23}\alpha_{12}\big)\delta_1^2 \qquad (8)$$

$$\Sigma_{11}\Sigma_{23} = \Sigma_{12}\Sigma_{13}$$
$$+ \Sigma_{11}\big(\delta_2^2\alpha_{23} + \delta_4^2\alpha_{42}\big(\alpha_{43} + \alpha_{23}\alpha_{42}\big)\big) \qquad (9)$$

$$\Sigma_{11}\Sigma_{22} = \Sigma_{12}^2 + \Sigma_{11}\big(\delta_2^2 + \delta_4^2\alpha_{42}^2\big) \qquad (10)$$

$$\Sigma_{11}\Sigma_{33} = \Sigma_{13}^2$$
$$+ \Sigma_{11}\big(\delta_2^2\alpha_{23}^2 + \delta_3^2 + \delta_4^2\big(\alpha_{43} + \alpha_{23}\alpha_{42}\big)^2\big). \qquad (11)$$

Proof. It is possible to express the covariance matrix $\bar{\Sigma}$ of the joint distribution of $X_1, \ldots, X_4$ in terms of the model parameters as follows:

$$\bar{\Sigma} = \big(I - A\big)^{-T}\Delta\big(I - A\big)^{-1}.$$

The individual components in (6)–(11) can now be obtained by straightforward algebraic calculations. $\quad\square$

Remark 2.1. (Instrumental variable estimator) From equation (8) it follows immediately that for $\alpha_{13} = 0$, we have

$$\alpha_{23} = \frac{\Sigma_{13}}{\Sigma_{12}},$$

which corresponds to the usual causal effect estimator in the instrumental variable setting [3].

The lemma we present now reflects the fact that we are always free to choose the scale for the unobserved confounder X_4:

Lemma 2.1. *The equations of proposition 2.1 are invariant under the following transformation*

$$\bar{\alpha}_{4j} = \sqrt{\delta_4^2}\alpha_{4j}, \qquad \bar{\delta}_4^2 = 1$$

for $j \in \{2, 3\}$.

Proof. This invariance follows from the fact that α_{42} and α_{43} always appear in a homogeneous polynomial of degree 2, and they are always coupled with a δ_4^2 term. $\quad\square$

Without loss of generality we can assume from now on that $\delta_4^2 = 1$.

Remark 2.2. (Geometrical Interpretation) From a geometrical point of view the joint system of equations for the observed covariance matrix defines a manifold $\mathcal{M}_\Sigma$ in the space of the model parameters $\mathcal{M}_4(\mathbb{R}) \times D_{\delta^2}$, where $\mathcal{M}_4(\mathbb{R})$ is the space of the possible causal strengths α_{ij} and
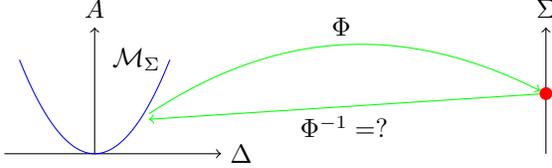
$$D_{\delta^2} = \prod_{i=1}^{3}[0, \Sigma_{ii}]$$

is the compact hypercube of the noise variances. Note that we have used the symmetry $\bar{\Sigma}_{44} = \delta_4^2 = 1$ and that

$$\delta_i^2 \leq \Sigma_{ii}$$

from equations (6), (10) and (11). Note that the map $\Phi : (A, \Delta) \mapsto \Sigma$ is not injective. This means that given an observed covariance matrix Σ, it is not possible to identify the model parameters in a unique way.

Indeed, the number of equations is six, while the number of model parameters is eight. Geometrically, this means that the manifold $\mathcal{M}_\Sigma$ does not reduce to a single point in the space of model parameters. Nevertheless it is still an interesting question whether the function g is a bounded function on $\mathcal{M}_\Sigma$ or not, i.e., whether we can give any guarantees on the estimated causal effect. Indeed, for the instrumental variable case with binary variables, such bounds can be derived (see, e.g., [13]).

The following Theorem and its Corollary are the main results of this paper. We will prove that there still remain degrees of freedom in the noise variances δ_2^2, δ_3^2 and the signs s_1, s_2, given the observed covariance matrix Σ, that will lead to an unbouded causal effect estimation error $g(A, \Sigma)$.

Theorem 2.1. *Given the causal model in Definition 1.3, there exists a map*

$$\Psi : \mathcal{M}_3(\mathbb{R}) \times D(\Sigma) \times \{-1, +1\}^2 \to \mathcal{M}_4(\mathbb{R}) \quad (12)$$

such that for all (A, Δ):

$$\Psi(\Phi(A, \Delta), \delta_2^2, \delta_3^2, s_1, s_2) = A. \quad (13)$$

Here $D(\Sigma) = [0, m/\Sigma_{11}] \times [0, \det \Sigma/m] \subset \mathbb{R}^2$ is the rectangle where the noise variances of X_2 and X_3 live, with m defined below in (19). The map Ψ gives explicit solutions for the causal strengths α_{ij}, given the observed covariance matrix Σ, the noise variances δ_2^2, δ_3^2 and signs $s_i = \pm 1$. The components of Ψ are given by:

$$\alpha_{12} = \frac{\Sigma_{12}}{\Sigma_{11}} \quad (14)$$

$$\alpha_{42} = s_1 \sqrt{\frac{m}{\Sigma_{11}} - \delta_2^2} \quad (15)$$

$$\alpha_{43} = s_2 \frac{\sqrt{\det \Sigma - m\delta_3^2}}{\sqrt{\delta_2^2 \Sigma_{11}}} \quad (16)$$

$$\alpha_{13} = s_1 s_2 \frac{\Sigma_{12} \sqrt{\det \Sigma - m\delta_3^2} \sqrt{m - \Sigma_{11}\delta_2^2}}{m\sqrt{\delta_2^2 \Sigma_{11}}} + \frac{\vartheta}{m}, \quad (17)$$

and the most important one for our purpose:

$$\alpha_{23} = \frac{\gamma}{m} - s_1 s_2 \frac{\sqrt{\det \Sigma - m\delta_3^2} \sqrt{m - \Sigma_{11}\delta_2^2}}{m\sqrt{\delta_2^2}}. \quad (18)$$

Here,

$$m = \Sigma_{11}\Sigma_{22} - \Sigma_{12}^2 > 0 \quad (19)$$

$$\eta = \Sigma_{11}\Sigma_{33} - \Sigma_{13}^2 > 0$$

$$\omega = \Sigma_{22}\Sigma_{33} - \Sigma_{23}^2 > 0$$

$$\vartheta = \Sigma_{13}\Sigma_{22} - \Sigma_{12}\Sigma_{23}$$

$$\gamma = \Sigma_{11}\Sigma_{23} - \Sigma_{12}\Sigma_{13}.$$

Proof. The proof proceeds by explicitly solving the system of equations (6)–(11). Some useful identities are:

$$\alpha_{13} = \frac{\Sigma_{12}\alpha_{42}\alpha_{43}}{m} + \frac{\vartheta}{m},$$

$$\alpha_{42}\alpha_{43} = \frac{\gamma - \alpha_{23}m}{\Sigma_{11}},$$

$$\rho_{13 \cdot 2} = \frac{\vartheta}{\sqrt{\omega m}},$$

$$\eta m - \gamma^2 = \Sigma_{11} \det \Sigma.$$

The signs in the equations are a consequence of the second degree polynomial equations. □

Corollary 2.1. *It is possible to express the error in the estimated causal effect as*

$$g(\Psi(\Sigma, \delta_2^2, \delta_3^2, s_1, s_2), \Sigma) = \frac{\vartheta \Sigma_{12}}{m\Sigma_{22}} +$$
$$s_1 s_2 \frac{\sqrt{\det \Sigma - m\delta_3^2}\sqrt{m - \Sigma_{11}\delta_2^2}}{m\sqrt{\delta_2^2}}. \quad (20)$$

By optimizing over δ_3^2 we get:

$$\alpha_{23} \in [b_-, b_+] \subset \mathbb{R},$$

with

$$b_\pm(\delta_2^2) = \frac{\gamma}{m} \pm \frac{\sqrt{\det \Sigma}\sqrt{m - \Sigma_{11}\delta_2^2}}{m\sqrt{\delta_2^2}}. \quad (21)$$

The length of the interval $[b_-, b_+]$ is a function of (Σ, δ_2^2) and satisfies

$$\frac{\partial|b_+ - b_-|}{\partial \delta_2^2} < 0.$$

Proof. Equation (20) follows from (18) and:

$$\frac{\Sigma_{23}}{\Sigma_{22}} = \frac{\gamma}{m} + \frac{\vartheta \Sigma_{12}}{m\Sigma_{22}}.$$

From equation (11), combined with the results of Theorem 2.1, we can obtain the following inequality, using also the fact that $\delta_3^2 \Sigma_{11} > 0$:

$$m\alpha_{23}^2 - 2\gamma\alpha_{23} + \eta - \Sigma_{11}\alpha_{43}^2 \geq 0.$$

The two solutions of the inequality define the interval $[b_-, b_+]$. Its length is a decreasing function of δ_2^2. □

Unfortunately, the causal effect strength α_{23} in equation (18) is unbounded. This means that for all the choices of the observed covariance matrix Σ that are in accordance with the model assumptions in Definition 1.3, the set of model parameters $(A, \Delta) \in \mathcal{M}_\Sigma$ that would explain Σ leads to an unbounded error g.

Indeed, a singularity is reached in the hyperplane $\delta_2^2 = 0$, which corresponds to making the random variable X_2 deterministic with respect to its parents X_1, X_4. Figure 1 shows the singularity of the function $|g(\Sigma, \delta_2^2, \delta_3^2)|$ in the limit $\delta_2^2 \to 0$. The rate of growth is proportional to the inverse of the standard deviation of the noise variable E_2:

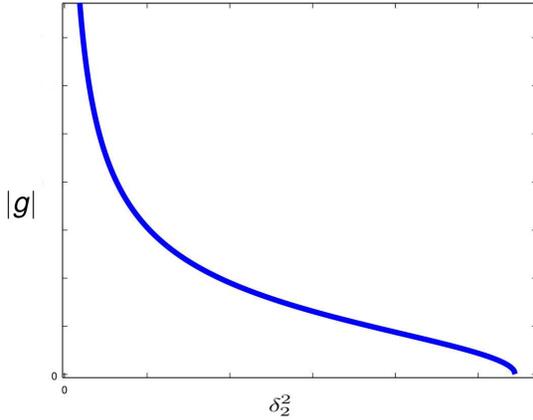$$|g| \propto \frac{1}{\delta_2} \quad \text{as } \delta_2 \to 0. \tag{22}$$

Figure 1: Causal effect estimation error $|g|$ as a function of δ_2^2, for fixed δ_3^2, Σ and $s_1 s_2 = 1$.

Remark 2.3. (Lower bound for δ_2^2) Corollary 2.1 is the main result of our analysis. The right hand term in (20) consists of two terms: the first one, through ϑ, represents the contribution of the partial correlation, and is small if $\rho_{13 \cdot 2}$ is small. The second term is a fundamental, intrinsic quantity not controllable from the conditional independence test and the sample size. However, in situations where one is willing to assume a lower bound on δ_2^2:

$$\delta_2^2 \geq \hat{\delta}_2^2,$$

it is possible to give a confidence interval $[b_+, b_-]$ for the function g, depending on the choice of the lower bound $\hat{\delta}_2^2$.

Remark 2.4. (IV estimation error)
In the instrumental variable literature the IV estimator is used, presented in Lemma 1.1. Unfortunately, this estimator and its error function

$$h(\Sigma, A) = \frac{\Sigma_{13}}{\Sigma_{12}} - \alpha_{23} \tag{23}$$

is proportional to α_{13} and from (17) one can deduce a similar growing rate of the function h in terms of the variance of the noise term E_2:

$$|h| \propto \frac{1}{\delta_2} \quad \text{as } \delta_2 \to 0. \tag{24}$$

Remark 2.5. (Singularity analysis)
Figure 2 shows a contour plot of $|g|$ on the rectangle $D(\Sigma) \ni (\delta_2^2, \delta_3^2)$. The singularity in the causal effect
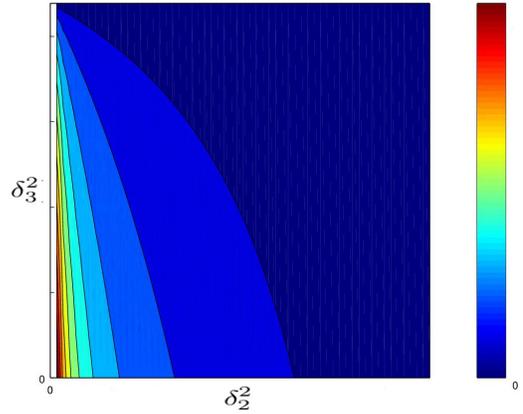

Figure 2: The function $|g|$ has a singularity in the hyperplane $\delta_2^2 = 0$.

function g is reached in the degenerate case, when the conditional distribution of X_2 given X_1 and X_4 approaches a Dirac delta function. This cannot be detected empirically, as we can still have well-defined covariance matrices Σ of the observed system even if the covariance matrix $\bar{\Sigma}$ of the extended one is degenerate.

Let us investigate in detail the limit for $\delta_2^2 \to 0$ from the point of view of the causal model. This proposition will show a simple example of how the causal strengths can be arbitrarily large, keeping the entries of the observed covariance matrix Σ_{ij} finite.

Proposition 2.2. *Assume that the observed covariance matrix Σ is positive-definite. Then, for the limit $\delta_2^2 \to 0$ we have the following scenario for the causal strength parameters:*

$$\begin{cases} \alpha_{23} \approx \pm \delta_2^{-1} \\ \alpha_{43} \approx \mp \operatorname{sgn}(\alpha_{42})\, \delta_2^{-1} \\ \alpha_{13} \approx \mp \operatorname{sgn}(\alpha_{12})\, \delta_2^{-1}. \end{cases}$$

This limit, in which our error in the estimated causal effect strength of X_2 on X_3 diverges, is illustrated in Figure 3.
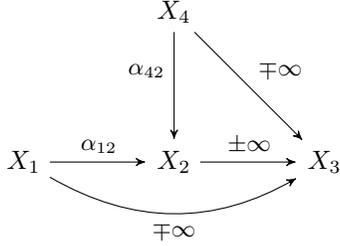
Figure 3: Scenarios in which the error in the causal effect strength of X_2 on X_3 based on the LCD algorithm may become infinitely large.

3 Conclusions and future work

Corollary 2.1 shows how the causal effect estimation error can be extremely sensitive to small perturbations of our model assumptions. Equation (20) holds for any value of ϑ (which is proportional to the partial correlation $\rho_{13 \cdot 2}$) and the second term vanishes when the confounder is not present. This shows that with a finite sample, a type II error in the conditional independence test may lead to an arbitrarily large error in the estimated causal effect. Even in the infinite sample limit, this error could be arbitrarily large if faithfulness is violated. The result is in agreement with the results in [9], and it shows in a clear algebraic way how type II errors of conditional independence tests can lead to wrong conclusions.

We believe that this conclusion holds more generally: even when we increase the complexity and the number of observed variables, the influence of confounders will still remain hidden, mixing their contribution with the visible parameters, thereby potentially leading to arbitrarily large errors. This means that for individual cases, we cannot give any guarantees on the error in the estimation without making further assumptions. An interesting question for future research is whether this negative worst-case analysis can be supplemented with more positive average-case analysis of the estimation error. Indeed, this is what one would hope if Occam's razor can be of any use for causal inference problems.
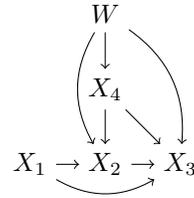
Other possible directions for future work are:

- **Study more complex models, in terms of the number of nodes, edges and cycles.**

- **Bayesian model selection:** We hope that the Bayesian approach will automatically prefer a simpler model that excludes a possible weak conditional dependence even though the partial correlation from the data is not exactly zero.

- **Bayesian Information Criterion:** We could directly assign a score based on the likelihood function of the data given the model parameters (A, Δ) and the model complexity, without assuming any prior distribution for the model parameters.

- **Nonlinear structural causal equations:** To deal with nonlinearity it is possible to consider Spearman's correlation instead of the usual one, using the following relationships:

$$m = \Sigma_{11}\Sigma_{22}(1 - \rho_{12}^2)$$
$$\eta = \Sigma_{11}\Sigma_{33}(1 - \rho_{13}^2)$$
$$\omega = \Sigma_{22}\Sigma_{33}(1 - \rho_{23}^2)$$
$$\gamma = \Sigma_{11}\sqrt{\Sigma_{22}\Sigma_{33}}(\rho_{23} - \rho_{12}\rho_{13})$$
$$\vartheta = \Sigma_{22}\sqrt{\Sigma_{11}\Sigma_{33}}(\rho_{13} - \rho_{12}\rho_{23})$$

- **"Environment" variable:** In many applications in biology, for example where X_1 is genotype, X_2 gene expression and X_3 phenotype, the observed random variables X_2 and X_3 are strongly dependent on the environmental conditions of the experiment. It might be reasonable to assume that most of the external variability is carried by the covariance between the environment variable W and the other measured ones, including possible confounders. This leads to the following graphical model, which could be useful in deriving some type of guarantees for this scenario:

References

[1] G. F. Cooper. A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Mining and Knowledge Discovery*, 1:203–224, 1997.

[2] R.J. Bowden and D.A. Turkington. *Instrumental Variables*. Cambridge University Press, 1984.

[3] V. Didelez and N. Sheehan. Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, 16:309–330, 2007.

[4] S. Greenland. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, 29:722–729, 2000.

[5] D. A. Lawlor, R. M. Harbord, J. A. C. Sterne, N. Timpson, and G. D. Smith. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, 27:1133–1163, 2008.

[6] J.D. Angrista, W. G. Imbens, and D.B. Rubinc. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444–455, 1996.

[7] D. A. Jaeger J. Bound and R. M. Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association*, 90:443–450, 1995.

[8] L. S. Chen, F. Emmert-Streib, and J. D. Storey. Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biology*, 8, 2007.

[9] R. Scheines J. M. Robins, P. Spirtes, and L. Wasserman. Uniform consistency in causal inference. *Biometrika*, 90:491–515, 2003.

[10] J. Zhang and P. Spirtes. Strong faithfulness and uniform consistency in causal inference. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence (UAI 2003)*, pages 632–639, 2003.

[11] C. Uhler, G. Raskutti, P. Bühlmann, and B. Yu. Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, 41:436–463, 2013.

[12] P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, prediction, and search*. The MIT Press, 2000.

[13] J. Pearl. *Causality: models, reasoning and inference*. Cambridge University Press, 2000.

[14] M. Drton, R. Foygel, and S. Sullivan. Global identifiability of linear structural equation models. *The Annals of Statistics*, 39:865–886, 2011.