Putting Causal Identification to the Test: Falsification using Multi-Environment Data

Rickard K.A. Karlsson
Delft University of Technology
The Netherlands
r.k.a.karlsson@tudelft.nl

Ștefan Creastă
Delft University of Technology
The Netherlands
creasta@student.tudelft.nl

Jesse H. Krijthe
Delft University of Technology
The Netherlands
j.h.krijthe@tudelft.nl

Abstract

We study the problem of falsifying the assumptions behind a set of broadly applied causal identification strategies: namely back-door adjustment, front-door adjustment, and instrumental variable estimation. While these assumptions are untestable from observational data in general, we show that with access to data coming from multiple heterogeneous environments, there exist novel independence constraints that can be used to falsify the validity of each strategy. Most interestingly, we make no parametric assumptions, instead relying on that changes between environments happen under the principle of independent causal mechanisms.

1 Introduction

A common theme within the field of causal inference has been to study settings with data collected from multiple environments. This type of data often tends to be heterogeneous due to e.g. changing circumstances or time shifts. While data heterogeneity is sometimes seen as an obstacle in data science, it is possible to turn it to one's advantage. For instance, it can allow us to learn invariant predictors that better generalize to unseen environments [Peters et al., 2016, Rothenhäusler et al., 2021], improve causal discovery [Ghassami et al., 2018, Mooij et al., 2020, Huang et al., 2020], and enable new causal effect identification strategies [Bareinboim and Pearl, 2016, Athey et al., 2020]. In this paper, we focus on the last two ideas together.

We study the problem of falsifying a set of broadly applied graphical conditions under the possible presence of latent variables: namely the *1. back-door criterion*, *2. front-door criterion*, and *3. instrumental variable criterion*. These conditions are crucial when we want to estimate the effects of interventions from observational data. Unfortunately, in the most general case, these conditions can not be verified from a single observational dataset alone [Pearl, 2009]. However, we will show that when we have multiple datasets stemming from different environments or clusters – such as different locations, time periods, or studies – some of these conditions can be tested.

Our contribution is to demonstrate that a novel type of independence constraints [Guo et al., 2022, Karlsson and Krijthe, 2023] can be used to falsify the above-mentioned conditions when we have access to multi-environment data under the assumption of independent causal mechanisms [Schölkopf et al., 2012, Peters et al., 2017, Schölkopf et al., 2021]. In particular, we aim to do this without access to interventional data. We believe our findings are of direct interest to those who want to test the validity of their causal identification strategy and have access to multi-environment data. However,

Accepted to the NeurIPS 2023 Workshop on Causal Representation Learning.

the technique we use to obtain our results may be of independent interest to the broader causality community.

2 Related Works

This paper contributes to the growing body of literature on doing causal inference from heterogeneous, multi-environment data [Peters et al., 2016, Bareinboim and Pearl, 2016, Mooij et al., 2020, Huang et al., 2020, Shi et al., 2021, Squires et al., 2023]. Most closely related to our work are Guo et al. [2022] and Karlsson and Krijthe [2023]. Assuming independent causal mechanisms, these works showcase novel independence constraints that can be used for causal discovery in multi-environment settings. Guo et al. [2022] focused on the setting with all variables observed (i.e. having causal sufficiency): they show in this case that we can go beyond the Markov Equivalence Class and uniquely determine the causal DAG from observational data. Meanwhile, in a similar setting, Karlsson and Krijthe [2023] relaxed the causal sufficiency assumption and showed how to detect the presence of latent confounders. The technique used in both works shares similarities to the twin network method for counterfactual reasoning by Balke and Pearl [1994] by looking at independence constraints in a "twinned" graph. In contrast to Balke and Pearl [1994], this "twinning" technique is applied to a setting with different environments having the same causal structure. We build further on these developments, showing new non-parametric identification results for widely applied identification strategies.

In this paper, we explore the possibilities for falsification implied by the independent causal mechanism assumption. There do however also exist other techniques for falsification. For instance, under mild conditions involving an auxiliary variable, Bhattacharya and Nabi [2022] demonstrate testable conditions for the front-door criterion. In addition, there are the well-known instrumental inequalities that sometimes can falsify the validity of instrumental variables [Pearl, 1995, Kédagni and Mourifie, 2017]. We believe our work can be used together with previously proposed tests like the ones mentioned, strengthening our toolbox to (in certain cases) falsify our causal assumptions.

3 Problem setting

We start with some preliminaries of the causal terminology used in this paper.

Definition 1 (Causal Graphical Model (CGM))**.** *A causal graphical model $M = (\mathcal{G}, P)$ over d random variables $\mathbf{V} = (V_1, V_2, \ldots, V_d)$ comprises (i) a directed acyclic graph (DAG) $\mathcal{G}$ with vertices $\mathbf{V}$ and edges $V_j \to V_j'$ iff V_j is a direct cause of V_j', and (ii) a joint distribution P such that it has the following Markov or causal factorization over $\mathcal{G}$:*

$$P(V_1, V_2, \ldots, V_d) = \prod_{j=1}^{d} P(V_j \mid \mathrm{Pa}(V_j)) \tag{1}$$

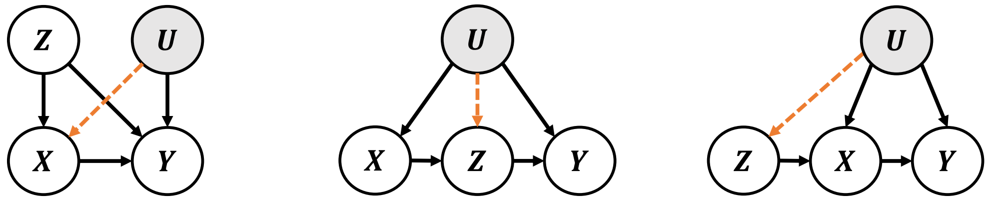
where $\mathrm{Pa}(V_j)$ denotes the parents (direct causes) of V_j in $\mathcal{G}$ and $P(V_j \mid \mathrm{Pa}(V_j))$ is the causal mechanism of V_j.

The DAG $\mathcal{G}$ encodes various conditional independences (or d-separations) between the variables which we write as $\mathbf{A} \perp\!\!\!\perp_d \mathbf{B} \mid \mathbf{C}$ over some disjoint sets of variables $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$. We shall assume that conditional independencies in $\mathcal{G}$ imply the same conditional independencies in P, and vice versa.

Assumption 1 (Faithfulness & Causal Markov Property)**.** *For P and $\mathcal{G}$ we have (i) the faithfulness property that $\mathbf{A} \perp\!\!\!\perp_P \mathbf{B} \mid \mathbf{C} \Rightarrow \mathbf{A} \perp\!\!\!\perp_d \mathbf{B} \mid \mathbf{C}$, and (ii) the causal Markov property that $\mathbf{A} \perp\!\!\!\perp_P \mathbf{B} \mid \mathbf{C} \Leftarrow \mathbf{A} \perp\!\!\!\perp_d \mathbf{B} \mid \mathbf{C}$.*

We will consider a setting with the following variables: We have treatment $X \in \mathcal{X}$ and outcome $Y \in \mathcal{Y}$, as well as an auxiliary variable $Z \in \mathcal{Z}$. In addition, we allow the presence of an unobserved confounder $U \in \mathcal{U}$ between X and Y. We shall further assume that we know that $Y \notin \mathrm{Ancestors}(X)$. This setting comes up when we are interested in estimating the interventional effect of X on Y, denoted as $P(Y \mid \mathrm{do}(X))$ using do-calculus [Pearl, 2009]; here we are often sure X "happens" before Y and we wish to learn if X has an effect on Y.

Depending on how (X, Y) relates with respect to Z, we can use different strategies to estimate the interventional effect from observational data: back-door adjustment if Z fulfills the back-door



(a) Back-door (b) Front-door (c) Instrumental variable

Figure 1: Three common settings where observing Z allows for identification of $P(Y \mid \text{do}(X))$; the shaded variables are unobserved. The addition of the dashed red arrow illustrates one way in which Z becomes insufficient for identification.

criterion; front-door adjustment if Z is a mediator fulfilling the front-door criterion; or instrumental variable estimation if Z is a valid instrument. These different settings are illustrated in Figure 1. While domain knowledge often informs us which strategy to use, no independence constraint exists between (X, Y, Z) that allows us to verify any of these conditions [Pearl, 1995]. We demonstrate that such conditions exist, however, when we have data from multiple environments and assume independent causal mechanisms. We will now formalize this assumption.

3.1 Assumptions for multi-environment data

We have observational datasets from multiple environments e_k, indexed by $k = 1, \ldots, K$. The datasets are sampled as $(X_i^{(k)}, Y_i^{(k)}, Z_i^{(k)}, U_i^{(k)}) \sim P^{(e_k)}(X, Y, Z, U)$ for $i = 1, \ldots, N_k$, where N_k is the number of observations in environment e_k. Note that in what follows, $U_i^{(k)}$ is not observed. We allow each environment to have a different joint distribution $P^{(e_k)}$ but assume they are related to each other through the following assumption:

Assumption 2 (Shared Causal Graph)**.** *All environments share the same causal DAG $\mathcal{G}$.*

Next, we specify how changes in $P^{(e_k)}(X, Y, Z, U)$ arise between the different environments. We shall assume that the conditional probabilities in (1) – which we refer to as causal mechanisms – vary independently per environment. This is known as the independent causal mechanism principle [Peters et al., 2017]. We shall now describe the assumption that operationalizes this.

To model changes between environments with independent causal mechanisms, we parameterize each causal mechanism with a parameter $\mathbf{\Theta} = \{\Theta_V \in \mathcal{O}_V : V \in \{X, Y, Z, U\}\}$.[1] In each environment, these are fixed and determine the distribution $P^{(e_k)}(X, Y, Z, U \mid \mathbf{\Theta}) = \prod_{V \in \{X,Y,Z,U\}} P^{(e_k)}(V \mid \text{Pa}(V), \Theta_V)$. One could see changes in $\mathbf{\Theta}$ as different soft interventions on the causal mechanisms, similar to the settings considered by Huang et al. [2020] and Perry et al. [2022].

Further, we shall assume that environments are randomly sampled from a *distribution over mechanisms* by defining non-degenerate probability measures for each causal mechanism.

Assumption 3 (Stochastic Independent Causal Mechanisms)**.** *The parameters Θ_V of the causal mechanisms are pair-wise independent random variables with non-degenerate probability measures $P(\Theta_V)$ for all $V \in \{X, Y, Z, U\}$.*

With the above assumption, when we say *independent* causal mechanisms, we refer to statistical independence between them. This independence is a strong assumption to make, which we will see gives us new testable implications in the data.

[1]For our intended purpose, note that we do not have to specify the explicit form of parameterization. This also means that, in principle, we do not specify the dimensionality of these parameters. While it is perhaps easier to imagine what independence between parameters looks like in the finite-dimensional case, one could also consider independence between infinite-dimensional parameters. This concept has been rigorously studied in nonparametric Bayesian inference, where one often constructs a prior over independent parameters [Ghosal and Van der Vaart, 2017].

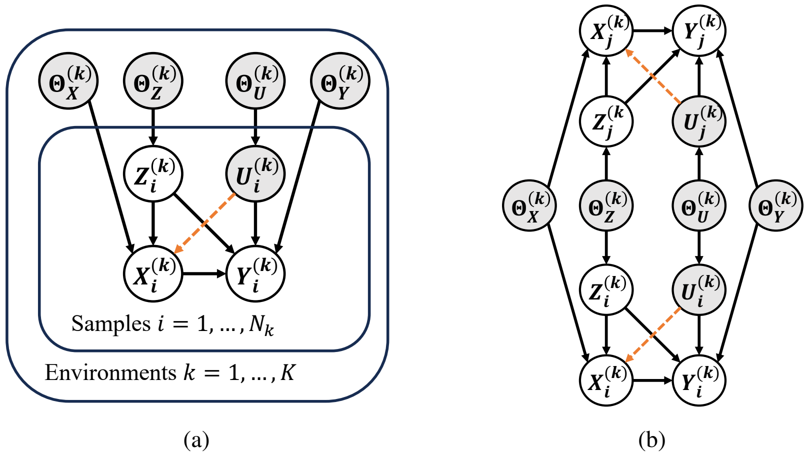


Figure 2: **(a)**: The hierarchical causal graphical model for the DAG from Figure 1a. **(b)**: We unfold the hierarchical causal graphical model to obtain a "twin" structure. This allows us to study the dependency structure between two different observations (i, j) from the same environment k.

4 Testing causal identification strategies with multi-environment data

We are now ready to present the main theoretical tool that we will use: a hierarchical causal graphical model that incorporates the multi-environment structure of the data under our assumptions. Crucially, in contrast to the single-environment causal graphical model, the hierarchical graph encodes additional independence constraints among the observed variables. These can be used to falsify causal identification strategies. We start with the definition of the hierarchical causal graphical model before we go into examples and results using this model.

Definition 2 (Hierarchical Causal Graphical Model)**.** *For a given environment e_k, we have the causal graphical model $M^{(e_k)} = (P^{(e_k)}, \mathcal{G}^{(e_k)})$ with variables $\mathbf{V}^{(k)} = (V_1^{(k)}, V_2^{(k)}, \ldots, V_d^{(k)})$. We define the hierarchical causal graphical model $M^* = (P^*, \mathcal{G}^*)$ as follows: Let $\mathcal{G}^*$ be a DAG containing vertices $\{\mathbf{V}_i^{(k)} : k = 1, \ldots, K\}$ for all observations $i = 1, \ldots, N_k$. It has the edge $V_{i,j}^{(k)} \to V_{i,j'}^{(k)}$ for all i iff the same edge exists in $\mathcal{G}^{(e_k)}$ where $j, j' = 1, \ldots, d$. Furthermore, we posit the causal mechanism parameters $\mathbf{\Theta}^{(k)} = (\Theta_{V_1}^{(k)}, \Theta_{V_2}^{(k)}, \ldots, \Theta_{V_d}^{(k)})$ to $\mathcal{G}^*$ so that $\Theta_{V_j}^{(k)} \to V_{i,j}^{(k)}$ for every i, j and k. The joint distribution P^* over all variables in $\mathcal{G}^*$ factorizes as*

$$\prod_{k=1}^{K} \prod_{i=1}^{N_k} \prod_{j=1}^{d} P^*(V_{i,j}^{(k)} \mid \mathrm{Pa}(V_{i,j}^{(k)}), \Theta_{V_j}^{(k)}) P^*(\Theta_{V_j}^{(k)}) \tag{2}$$

where $P^*(V_{i,j}^{(k)} \mid \mathrm{Pa}(V_{i,j}^{(k)}), \Theta_{V_j}) = P^{(e_k)}(V_{i,j}^{(k)} \mid \mathrm{Pa}(V_{i,j}^{(k)}), \Theta_{V_j})$.

To illustrate why the hierarchical causal graphical model is helpful, we first revisit a result from Karlsson and Krijthe [2023], showing how it can be used to falsify the back-door criterion.

4.1 Testing the back-door criterion

Let all environments share the graph $\mathcal{G}$ from Figure 1a and construct its corresponding hierarchical DAG $\mathcal{G}^*$, seen in Figure 2a. For the original DAG $\mathcal{G}$, it is well-known that there exist no independence constraints between the observed variables for testing the presence of the unobserved confounder U [Pearl, 1995]. For the hierarchical $\mathcal{G}^*$, however, we will see that such constraints exist.

In graph $\mathcal{G}^*$, we can study dependencies between two different samples (i, j) within an environment k: that is, $(X_i^{(k)}, Y_i^{(k)}, Z_i^{(k)})$ and $(X_j^{(k)}, Y_j^{(k)}, Z_j^{(k)})$ where $i \neq j$. Interestingly, if we do not condition on the environment – or conversely, the causal mechanism parameters $\mathbf{\Theta}$ in $\mathcal{G}^*$ – these two samples are dependent as they share parents in $\mathcal{G}^*$. This is illustrated in Figure 2b, where we unfold the hierarchical structure; or, one could say that we have created a "twin" of the original graph.

Now, one can verify graphically that if the dashed arrow is absent in the graph in Figure 2b then

$$X_i^{(k)} \perp\!\!\!\perp_{P^*} Y_j^{(k)} \mid X_j^{(k)}, Z_i^{(k)}, Z_j^{(k)} \,. \tag{3}$$

But if the dashed arrow is present, such that Z becomes an invalid back-door adjustment set as we have an open backdoor path between X and Y, then the independence in (3) is violated. As (3) only contains observed variables, the back-door criterion has testable implications according to the hierarchical model. In fact, this statement is true even if we consider a larger set of possible graphs.

Theorem 1 (Karlsson and Krijthe [2023])**.** *Consider assumptions 1, 2 and 3 where $Y \notin$ Ancestors(X) and that there is no selection bias. Let $\mathcal{G}$ be the shared causal DAG across environments and $\mathcal{G}^*$ its corresponding hierarchical DAG. Then, we have that* (3) *holds for any k and $i \neq j$ in $\mathcal{G}^*$ iff Z blocks every back-door path between X and Y in G.*

How do we test this independence? We have shown here that opening a back-door path leads to the violation of a new independence constraint in the observed data distribution. For the rest of the paper, our goal is to provide more of these identification results, while constructing efficient tests for these dependencies is outside the scope of our paper. This problem has been studied by both Guo et al. [2022] and Karlsson and Krijthe [2023]. For the interested reader, however, we provide an explanation of how to test independencies such as (3) in the Appendix.

4.2 Testing the front-door criterion

The next graphical condition we will explore is the front-door criterion where Z is a mediator between X and Y, as demonstrated in Figure 1b and which is defined as follows:

Definition 3 (Front-door criterion [Pearl, 2009])**.** *A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) in a DAG $\mathcal{G}$ if: (i) Z intercepts all directed paths from X to Y; (ii) there is no unblocked back-door path from X to Z; and (iii) all back-door paths from Z to Y are blocked by X.*

If we know the causal ordering of (X, Z, Y), then we see that Theorem 1 can be directly applied to construct testable implications for both (ii) and (iii) in Definition 3.

Corollary 1. *Consider assumption 1, 2 and 3 with $Y \notin$ Ancestors(Z), $Z \notin$ Ancestors(X), and no selection bias, let $\mathcal{G}$ be the shared causal DAG across environments and $\mathcal{G}^*$ its corresponding hierarchical DAG. Then, for any k and $i \neq j$,*

$$X_i^{(k)} \perp\!\!\!\perp_{P^*} Z_j^{(k)} \mid X_j^{(k)} \;\; \textit{and} \;\; Z_i^{(k)} \perp\!\!\!\perp_{P^*} Y_j^{(k)} \mid Z_j^{(k)}, X_i^{(k)}, X_j^{(k)} \tag{4}$$

iff condition (ii) and (iii) in Definition 3 hold true for $\mathcal{G}$.

Proof. We apply Theorem 1 twice, noting that conditions (ii) and (iii) concern that there exist no unblocked back-door paths. For (ii), we need to check that there is no open back-door path between the ordered pair (X, Z) with an empty adjustment set; this results in the first independence. Similarly, for (iii), we get the second independence by having to check whether X is sufficient to block any back-door path between the ordered pair (Z, Y). □

Starting on a positive note, we have shown that it is in fact possible to verify two out of three conditions in the front-door criterion. While the "twinning" technique is very suitable to detect open back-door paths, we will see now that testing the remaining condition – whether Z intercepts all directed paths between (X, Y) – is more difficult to test; in fact, it is impossible to do it with this technique.

Theorem 2. *Consider the same assumptions as in Corollary 1 with $\mathcal{G}$ being the shared causal DAG across environments and $\mathcal{G}^*$ its corresponding hierarchical DAG. then there exist no independence constraints in $\mathcal{G}^*$ that imply whether condition (i) in Definition 3 holds.*

A proof of the theorem is provided in the Appendix.

4.3 Testing the instrumental variable criterion

Now, we turn our attention to the third identification strategy to see whether it is possible to reject the validity of an instrumental variable. As before, we start with the graphical definition of an instrumental variable.

Definition 4 (Graphical criterion for instrumental variable Pearl [2009])**.** *A variable Z is an instrument relative to the total effect of X on Y if (i) $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$ and (ii) $(Z \not\perp\!\!\!\perp_d X)_{\mathcal{G}}$. Here $\mathcal{G}_{\bar{X}}$ refers to the causal graph $\mathcal{G}$ where all incoming edges into X have been removed.*

We note that condition (ii) in the above definition is already a testable independence constraint. Thus, we put our attention on whether we can test condition (i) – which does not have any observable independence constraints in $\mathcal{G}$ – using the "twinning" technique. We start with observing a problematic special case.

Theorem 3. *Consider assumption 1, 2 and 3 with $Y \notin Ancestors(X)$, $X \notin Ancestors(Z)$, and no selection bias, let $\mathcal{G}$ be the shared causal DAG across environments where the (testable) condition $(Z \not\perp\!\!\!\perp_d Y)$ holds and $\mathcal{G}^*$ its corresponding hierarchical DAG. Then, there exist no independence constraints in $\mathcal{G}^*$ for whether the edge $Z \to Y$ is present or not.*

We provide proof in the Appendix. The consequence of this theorem is that without further assumptions, we can not find an independence constraint in $\mathcal{G}^*$ that implies $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$. The reason is that the presence of the edge $Z \to Y$ implies that $(Z \not\perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$. That Z may not have a direct effect on Y is also referred to as the exclusion restriction for instrumental variables [Angrist et al., 1996]. The result itself might not come as a surprise, as the impossibility result we proved for the front-door criterion in Theorem 2 also relates to the presence of such direct edges. Despite this, we can still in some scenarios falsify if $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$ is true.

Theorem 4. *Consider the same assumptions as in Theorem 3, let $\mathcal{G}$ be the shared causal graph across environments and $\mathcal{G}^*$ its corresponding hierarchical DAG. Then, for any k and $i \neq j$, we have $Z_i^{(k)} \not\perp\!\!\!\perp_{P^*} Y_j^{(k)} \mid Z_j^{(k)} \Rightarrow (Z \not\perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$.*

The theorem presents an approach to falsify the validity of an instrument. As shown in the proof of the theorem, which is found in the Appendix, falsification is possible when the unobserved confounder U is a cause of Z. In literature, this relates to the necessary condition that Z must be independent of any exogenous variable between X and Y [Angrist et al., 1996]. We note however that falsification is not possible if it is the other way around, i.e. $Z \to U$. This means that if one would conclude that $Z_i^{(k)} \perp\!\!\!\perp_{P^*} Y_j^{(k)} \mid Z_j^{(k)}$, one still needs to think carefully about the assumptions that have been made.

5 Discussion

In this paper, we have studied a new type of hierarchical causal model for data from multiple environments and its use in deriving testable implications of violations of common identification strategies. We learned that there exist independence constraints in this new class of DAGs that can be used to falsify (parts of) three common identification strategies in causal inference: the back-door, front-door, and instrumental variable criterion. If one of the testable conditions we have presented is violated, this could be informative to us that not all of our assumptions are valid for identification.

It is important to note that, although these hierarchical models expand the possibilities of testing assumptions, they are not a silver bullet. Firstly, our theory relies on a new untestable assumption: the independent causal mechanisms varying across environments. This assumption should not be taken for granted, yet a more conservative interpretation of the tests presented in this paper would be that they are a joint test to detect either a violation in the identification assumptions or that the mechanisms are dependent. Secondly, we demonstrated some limits of using the "twinning" technique with the hierarchical models. In particular, we learned that we can not test for the presence of a direct edge in the front-door and the instrumental variable setting. Still, we believe that showing we can test parts of these conditions constitutes important progress in the falsification of causal assumptions.

The hierarchical causal graphical model was a useful model in this setting that may be insightful in other causal inference settings as well. Interesting directions in this regard are investigating other identification strategies or combining this model with traditional independence-based causal discovery.

References

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

Susan Athey, Raj Chetty, and Guido Imbens. Combining experimental and observational data to estimate treatment effects on long term outcomes. *arXiv preprint arXiv:2006.09676*, 2020.

Alexander Balke and Judea Pearl. Counterfactual probabilities: Computational methods, bounds and applications. In *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*, UAI'94, page 46–54, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc. ISBN 1558603328.

Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

Rohit Bhattacharya and Razieh Nabi. On testability of the front-door model via verma constraints. In *Uncertainty in Artificial Intelligence*, pages 202–212. PMLR, 2022.

AmirEmad Ghassami, Negar Kiyavash, Biwei Huang, and Kun Zhang. Multi-domain causal structure learning in linear systems. *Advances in neural information processing systems*, 31, 2018.

Subhashis Ghosal and Aad Van der Vaart. *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press, 2017.

Siyuan Guo, Viktor Tóth, Bernhard Schölkopf, and Ferenc Huszár. Causal de finetti: On the identification of invariant causal structure in exchangeable data. *arXiv preprint arXiv:2203.15756*, 2022.

Biwei Huang, Kun Zhang, Jiji Zhang, Joseph Ramsey, Ruben Sanchez-Romero, Clark Glymour, and Bernhard Schölkopf. Causal discovery from heterogeneous/nonstationary data. *The Journal of Machine Learning Research*, 21(1):3482–3534, 2020.

Rickard K.A. Karlsson and Jesse H. Krijthe. Detecting hidden confounding in observational data using multiple environments. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Désiré Kédagni and Ismael Mourifie. Generalized instrumental inequalities: Testing the iv independence assumption. *Available at SSRN 2692274*, 2017.

Joris M. Mooij, Sara Magliacane, and Tom Claassen. Joint causal inference from multiple contexts. *Journal of Machine Learning Research*, 21(99):1–108, 2020.

Judea Pearl. On the testability of causal models with latent and instrumental variables. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, page 435–443, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1558603859.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Ronan Perry, Julius Von Kügelgen, and Bernhard Schölkopf. Causal discovery in heterogeneous environments under the sparse mechanism shift hypothesis. *Advances in Neural Information Processing Systems*, 35:10904–10917, 2022.

Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5):947–1012, 2016.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 1st edition, 2017.

Dominik Rothenhäusler, Nicolai Meinshausen, Peter Bühlmann, and Jonas Peters. Anchor regression: Heterogeneous data meet causality. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(2):215–246, 2021.

Bernhard Schölkopf, Dominik Janzing, Jonas Peters, Eleni Sgouritsa, Kun Zhang, and Joris Mooij. On causal and anticausal learning. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pages 459–466, 2012.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

Claudia Shi, Victor Veitch, and David M Blei. Invariant representation learning for treatment effect estimation. In *Uncertainty in Artificial Intelligence*, pages 1546–1555. PMLR, 2021.

Chandler Squires, Anna Seigal, Salil Bhate, and Caroline Uhler. Linear causal disentanglement via interventions, 2023.

A Practical testing of independence constraints

In this section, we outline the procedure for testing a conditional independence relationship like $X_i^{(k)} \perp\!\!\!\perp_{P^*} Y_j^{(k)} \mid X_j^{(k)}, Z_i^{(k)}, Z_j^{(k)}$ or those in (4), utilizing multi-environment data. We denote this data with $\{x_i^{(k)}, y_i^{(k)}, z_i^{(k)}\}_{i=1}^{N_k}$ with $k = 1, \ldots, K$.

To test such independencies, we want to simulate sampling from the joint distribution $P^*(X_i^{(k)}, Y_i^{(k)}, Z_i^{(k)}, X_j^{(k)}, Y_j^{(k)}, Z_j^{(k)})$ for some $i \neq j$. It is worth noting here that we do not condition on the environment, because otherwise the sample pair (i, j) would always be independent. Here's the approach we follow:

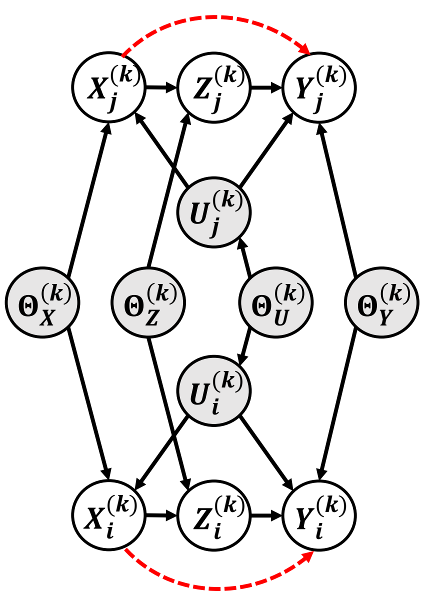
1. We select two distinct observations, denoted as i and j, from all environments. This selection yields vectors of observed treatments $x_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(K)})$, outcomes $y_i = (y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(K)})$, and so on for the vectors for z_i, x_j, y_j and z_j.

2. Subsequently, we apply a suitable conditional independence testing method, using the data points in $(x_i, y_i, z_i, x_j, y_j, z_j)$ as samples of each respective random variable.

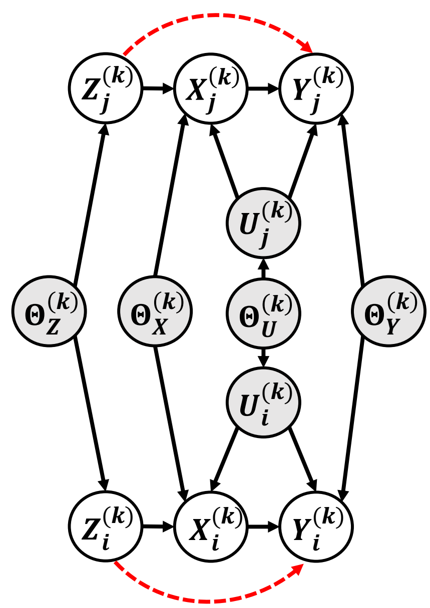
It's important to note that the choice of observations within each environment is arbitrary, as long as we avoid selecting the same observation for both i and j. This flexibility arises from the assumption that observations are independent and identically distributed within each environment.

We see that, in principle, we only need two observations per environment to perform this independence test. The "sample size" of the test is the number of environments. However, it is possible to construct a procedure that uses all available data by combining multiple independence tests using Fisher's method, as long as we select different observations for each test [Karlsson and Krijthe, 2023].

B Proofs



(a) Front-door setting



(b) Instrumental variable setting

Figure 3: Graphs to illustrate the claims by Theorem 2 and Theorem 3. We compare the independence constraints in the hierarchical DAG with either the red dashed edge present or absent; this corresponds to a violation of either the front-door or instrumental variable criterion respectively.

B.1 Proof of Theorem 2

Proof. We will show that there exists no independence constraint in the hierarchical graph $\mathcal{G}^*$, illustrated in Figure 3a, that is affected by the presence or absence of the red dashed edge. This edge corresponds to a violation of the fact that Z must intercept all directed paths between X and Y according to the front-door criterion (see Defintion 3). We construct our proof by showing that for the graph $\mathcal{G}^*$ in Figure 3a, the presence or absence of a red dashed arrow changes no independence constraint in $\mathcal{G}^*$.

First, we note that because of Assumption 2, any independence constraint in $\mathcal{G}^*$ holds for all environments k. Secondly, because of Assumption 3, there exist open paths between the sample pair (i, j). Thirdly, Assumption 1 allows us to connect d-separation in G^* with the independence statement in the data distribution P^*. Finally, to show our claim, we only have to consider independence constraints between (i, j) samples, in contrast to for instance $(X_i^{(k)} \perp\!\!\!\perp_d Z_i^{(k)})_{\mathcal{G}^*}$, since $\mathcal{G}^*$ in this case otherwise does not provide anything extra compared to the corresponding non-hierarchical DAG $\mathcal{G}$.

We start by considering the independencies of the form $X_i^{(k)} \perp\!\!\!\perp_d Y_j^{(k)} \mid S$, where S is a set of the other observed variables. We note there always is a path between $X_i^{(k)}$ and $Y_j^{(k)}$ that traverses through $(U_i^{(k)}, U_j^{(k)})$, regardless of S. Thus, this type of independence does not change based on the presence of the edge $X \to Y$.

Next, we look at the independencies of the form $X_i^{(k)} \perp\!\!\!\perp_d Z_j^{(k)} \mid S$. We note that there is a path between $X_i^{(k)}$ and $Z_j^{(k)}$ through X_j that does not depend on the edge $X \to Y$, thus we always $X_j^{(k)} \in S$ to block this path. The only way to unblock the path between $X_i^{(k)}$ and $Z_j^{(k)}$ is to let $Y_j^{(k)} \in S$. However, this path does not go through the direct edge $X \to Y$ either. Thus, no independence of the form $X_i^{(k)} \perp\!\!\!\perp_d Z_j^{(k)} \mid S$ can detect the presence of this direct edge.

We look at the final form of independencies: that is $Y_i^{(k)} \perp\!\!\!\perp_d Z_j^{(k)} \mid S$. Using similar reasoning as above, it is clear that regardless of S, no path between $Y_i^{(k)}$ and $Z_j^{(k)}$ depends on the presence of the edge $X \to Y$.

As we have considered all possible types of independence constraints between observed variables, we see that no independence in $\mathcal{G}^*$ will change because of the presence of the direct edge $X \to Y$. This means that we can not test whether Z intercepts all directed paths between X and Y using this "twinning" technique. □

B.2 Proof of Theorem 3

Proof. We will show that there exists an observable independence constraint in the hierarchical graph $\mathcal{G}^*$, illustrated in Figure 3b, that depends on the presence of the red dashed edge. This edge corresponds to a violation of the fact that $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$ must hold for Z to be a valid instrument (see Definition 4). Here $\mathcal{G}_{\bar{X}}$ refers to the causal graph $\mathcal{G}$ where all incoming edges into X have been removed. We construct our proof by showing that for the graph $\mathcal{G}^*$ in Figure 3b, the presence or absence of a red dashed arrow changes no independence constraint in $\mathcal{G}^*$.

We use the same arguments as in the proof of Theorem 2 to conclude that we may look at independence constraint for any k and that there exist open paths between different pairs of samples (i, j). Once again, we will check all relevant independence constraints in the hierarchical DAG $\mathcal{G}^*$ and see if they would change if the red dashed edge is present or absent.

First, we look at the independencies of the form $X_j^{(k)} \perp\!\!\!\perp_d Y_i^{(k)} \mid S$ with S comprising the other observed variables. We note that this independence will always be violated, i.e. $X_j^{(k)} \not\!\perp\!\!\!\perp_d Y_i^{(k)} \mid S$ for any S. This is because we can always reach $(U_i^{(k)}, U_j^{(k)})$ without traversing $Z \to Y$.

Secondly, we look at the independencies of the form $X_i^{(k)} \perp\!\!\!\perp_d Z_j^{(k)} \mid S$. If $Z_i^{(k)} \notin S$, then we always have $X_i^{(k)} \not\!\perp\!\!\!\perp_d Z_j^{(k)} \mid S$. So we only have to consider $Z_i^{(k)} \in S$. In that case we have that $X_i^{(k)} \not\perp\!\!\!\perp_d Z_j^{(k)} \mid S$ holds whenever $X_j^{(k)}$ is also being conditioned on. In case $X_j^{(k)} \notin S$, $X_i^{(k)} \not\perp\!\!\!\perp_d Z_j^{(k)} \mid S$ will still be true if either $Y_i^{(k)}$ and/or $Y_j^{(k)}$ are being conditioned on. Therefore, the outcome of this type of independence test does not change based on the presence of the edge $Z \to Y$.

Finally, we look at $Y_i^{(k)} \perp\!\!\!\perp_d Z_j^{(k)} \mid S$. $Y_i^{(k)} \not\perp\!\!\!\perp_d Z_j^{(k)} \mid S$ will always hold if $Z_i^{(k)} \notin S$. In case $Z_i^{(k)} \in S$, we have $Y_i^{(k)} \not\perp\!\!\!\perp_d Z_j^{(k)} \mid S$ if $X_j^{(k)} \in S$ and/or $Y_j^{(k)} \in S$. This is because they both open a collider path through $X_j^{(k)}$. If $\{X_i^{(k)}, X_j^{(k)}\} \in S$, then the confounder association can be traversed. However, if $S = \{Z_i^{(k)}\}$ or $S = \{X_i^{(k)}, Z_i^{(k)}\}$, then $Y_i^{(k)}$ and $Z_j^{(k)}$ are independent. In none of these cases the edge $Z \to Y$ was used.

As we have considered all possible types of independence constraints between observed variables, we see that no independence in $\mathcal{G}^*$ will change because of the presence of the direct edge $Z \to Y$ in $\mathcal{G}$. This means that we cannot test whether $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$ using this "twinning" technique in general. In Theorem 4, however, we show that there are still cases where we can detect a violation of $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$. □

B.3 Proof of Theorem 4

Proof. For this proof, we use a computational approach to iterate over different DAGs G while simultaneously searching for independence constraints in the corresponding hierarchical DAG G^* that can discriminate whether $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$ holds or not. Compared to the proofs of Theorem 2 and 3, where we only considered two graphs, we now must consider a much larger set of graphs to check whether an independence constraint gives the same value as $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$.

This approach consists of two parts: First, we iterate over a list of DAGs $\mathcal{G}$ that respect the following properties:

- U is always a confounder between X and Y, i.e. $U \to X$ and $U \to Y$ must be present;
- edges $X \to Y$, $Z \to Y$ and/or $Z \to X$ are present or absent (as we assume to know the causal ordering);
- and $U \to Z$ can either be present, absent or reversed;

This gives us a total of 24 graphs. In these graphs, the independence constraint $Z_j^{(k)} \perp\!\!\!\perp_d Y_i^{(k)} \mid Z_i^{(k)}$ often has the same values as $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$. As we illustrate in Table 1, we see that $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$ always holds if also $Z_j^{(k)} \perp\!\!\!\perp_d Y_i^{(k)} \mid Z_i^{(k)}$ is true, but not vice versa. We also see that $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$ is violated for graphs 0-11 as these have the direct edge between Z to Y, thus the most interesting cases are for graphs 12-23.

As $(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}} \Rightarrow (Z_j^{(k)} \perp\!\!\!\perp_d Y_i^{(k)} \mid Z_i^{(k)})_{\mathcal{G}^*}$, then must $(Z \not\perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}} \Leftarrow (Z_j^{(k)} \not\perp\!\!\!\perp_d Y_i^{(k)} \mid Z_i^{(k)})_{\mathcal{G}^*}$. Due to faithfulness (Assumption 1) we have that

$$(Z \not\perp\!\!\!\perp_P Y)_{\mathcal{G}_{\bar{X}}} \Leftarrow Z_j^{(k)} \not\perp\!\!\!\perp_{P^*} Y_i^{(k)} \mid Z_i^{(k)} .$$

□

Table 1: Each row corresponds to a different graph $\mathcal{G}$ considered in the proof for Theorem 4. The second column depicts the necessary conditions for the validity of Z being an instrument, while the third presents the independence constraint in $\mathcal{G}^*$ – a checkmark (✓) indicates that an independence hold. The remaining columns show the edges we change in the graphs.

Graph	$(Z \perp\!\!\!\perp_d Y)_{\mathcal{G}_{\bar{X}}}$	$Z_j \perp\!\!\!\perp_d Y_i \mid Z_i$	Z, X	Z, U	X, Y	Z, Y
0		✓	$Z \to X$	$Z \to U$	$X \to Y$	$Z \to Y$
1		✓	$Z \to X$		$X \to Y$	$Z \to Y$
2			$Z \to X$	$U \to Z$	$X \to Y$	$Z \to Y$
3		✓		$Z \to U$	$X \to Y$	$Z \to Y$
4		✓			$X \to Y$	$Z \to Y$
5				$U \to Z$	$X \to Y$	$Z \to Y$
6		✓	$Z \to X$	$Z \to U$		$Z \to Y$
7		✓	$Z \to X$			$Z \to Y$
8			$Z \to X$	$U \to Z$		$Z \to Y$
9		✓		$Z \to U$		$Z \to Y$
10		✓				$Z \to Y$
11				$U \to Z$		$Z \to Y$
12		✓	$Z \to X$	$Z \to U$	$X \to Y$	
13	✓	✓	$Z \to X$		$X \to Y$	
14			$Z \to X$	$U \to Z$	$X \to Y$	
15		✓		$Z \to U$	$X \to Y$	
16	✓	✓			$X \to Y$	
17				$U \to Z$	$X \to Y$	
18		✓	$Z \to X$	$Z \to U$		
19	✓	✓	$Z \to X$			
20			$Z \to X$	$U \to Z$		
21		✓		$Z \to U$		
22	✓	✓				
23				$U \to Z$		