

A Gentle Introduction into Structural Causal Models

Daniel Saggau • `daniel.saggau@campus.lmu.de`
Department of Statistics, Ludwig Maximilian University Munich, Germany

June 2021

Abstract The interest in understanding relationships of variables beyond co-occurrence has increased the popularity of causal modelling. Probabilistic specifications cast a model based on conditional probabilities. SCMs cast a model based on asymmetric assignments and extend probabilistic models by specifying the entire data generating process rather than solely utilizing conditional probabilities. A SCM is built on a number of different assumptions namely the Independence of Noise, Autonomy of Mechanisms and Causal Sufficiency. I also discuss the role of time in causality, focusing on how differential equations enable us to develop a SCM like a mechanical model. Another difference between the SCM and other model specifications is the ability of SCMs to address different queries such as *predictions*, *interventions* and *counterfactuals*. These queries are part of Pearl's causal hierarchy (2009). Pearl matches these queries with their respective actions namely *observing*, *doing* and *imagining*. I compare the feasibility of addressing these queries. The insights of this paper can be used as a baseline for subsequent research on structural causal models.

Contents

1	Introduction	3
2	Assumptions in Structural Causal Models	4
2.1	Mathematical Components	4
2.2	Independence of Noise and Mechanism	5
3	Directed Acyclic Graphs	7
4	SCMs and Time	8
5	Pearl's Causal Hierachy	10
5.1	Association:	10
5.2	Intervention:	11
5.3	Counterfactuals	13
6	Considerations	14
7	Conclusion	14
8	References	16
9	Appendix	19

1 Introduction

For many research problems, we want to understand the relationship between variables beyond co-occurrence. Algorithmic decision making based on co-occurrence is insufficient in high stake settings (Bareinboim et al. 2020). For many problems we want to understand causal relationships between variables. There are different approaches on how to model causal relationships. The most popular causal model is the structural causal model or in short SCM. The geneticist and statistician Sewall Wright introduced the first ancestor of the SCM, the path analysis in the 1920s (Pearl 2009a; Tarka 2018). Path analysis now falls into the broader class of structural equation models (SEM) for the specific case of using one variable per indicator. SEMs are popular in fields like economics, psychology and sociology (Pearl 2009a). The SCM is the non-parametric counterpart to structural equation models. SCMs specify an underlying data generating process without the computational effort of creating a parametric form.¹ Researchers on SCMs tried to distance themselves from common practices in structural equation modeling (Pearl 2009a; Peters, Janzing, and Schölkopf 2017) by using this different name, although frequently they are used interchangeably.

The SCM is an expressive simulator to estimate causal relationships, accounting for latent factors. Latent factors are unobserved factors (Pearl 2009a). SCMs entail endogenous and exogenous variables. Exogenous variables should not be confused with latent variables. For reference, Peters, Janzing, and Schölkopf (2017) define endogenous and exogenous as follows: *Endogeneous variables are those that the modeler tries to understand, while exogenous ones are determined by factors outside the model, and are taken as given.* The basis of these SCMs is a set of equations, or more precisely assignments, providing functions to derive the conditional probabilities for our model (Hardt and Recht 2021). These assignments describe our variables in our model. Compared to probabilistic models, where we only specify full joint distribution, the SCM actually enables the combination of different sources of knowledge. In association based learning and typical statistical learning we only model based on observational data and conditional probabilities given a specific static dataset. Henceforth, these methods perform poorly in cases of covariate shifts or domain changes. To account for these changes, we can specify an intervention distribution and latent factors. Conditional probabilities alone, cannot represent latent variables because there is no conditional probability in our observational data for unobserved variables (Pearl 2009a).

To accommodate existing literature, this paper provides a gentle introduction to SCMs, focusing on the underlying assumptions we make when building a structural causal model. The aim of this paper is to provide an intuitive understanding of different concepts within causality using elements from core literature on causality. Specifically, this paper focuses on the independence of noise, independence of mechanisms, causal sufficiency (section 2). Section 3 discusses causal graphs and how we derive underlying graphical

¹Observational data is seldom consistent. Scholars often question how informative these estimates are. For further information see Hernán and Taubman (2008) and Pearl (2012a)

representations based on single observational datasets. Section 4 examines causality and the role of time. Section 5 studies the causal hierarchy. I further provide considerations for the usage of SCMs (section 6). Section 6 concludes the paper.

2 Assumptions in Structural Causal Models

2.1 Mathematical Components

A SCM contains a (sub-)set of autonomous equations. These equations are asymmetric assignments. Regular equations are bidirectional. Assignments are not bi-directional. Equations in causal models did not always have a concise notation. Treating an equation as a **algebraic equation** led to confusion because those have no causal information. An algebraic equation would imply that $E = F$ and $C = E$ because the order has no concrete meaning in algebraic equations. Cause and effect would be interchangeable which is undesirable for causal modelling. The initial ' $=$ ' sign was replaced with the ' $:=$ ' which is asymmetric and entails causal information. This misconception has caused a lot of challenges.²

There are various definitions for structural causal models. Regardless, there is a consensus that a SCM contains two components namely an underlying causal graph and a set of assignments. (Peters, Janzing, and Schölkopf 2017) define a SCM as follows:

Definition 1: Structural Causal Model:

An SCM C with graph $C \rightarrow E$ consists of two assignments

$$\begin{aligned} C &:= N_C \\ E &:= f_E(C, N_E) \end{aligned}$$

where $N_E \perp\!\!\!\perp N_C$ that is N_E is independent of N_C

In their definition, C is the cause and E is the effect. N_C is the random noise variable for the cause. N_E is the noise variable for the effect variable.

for the subsequent paper i will use a real world example. One common example stems from epidemiology. This example looks at the impact of problem behavior such as smoking on lung cancer. Problem behavior is a conceptual term and latent because we cannot observe pure 'problem behaviour' in observational data. We can use variables that reflect problem behaviour to study problem behaviour. One example of problem behavior is smoking. We can for instance ask participants how frequently they smoke. As we can see, these functions are driven by underlying latent variables. These latent factors are the foundation of the structural causal model (Hardt and Recht 2021; Pearl 2009a, 2012a).

²For more information see Pearl (2009a)

$$S := f_S(U_S) \quad (1)$$

$$C := f_C(S, U_C) \quad (2)$$

where: $\{S\}$ - Frequency of smoking $\{C\}$ - Lung cancer

Every structural causal model contains an underlying graphical model (Hardt and Recht 2021). This is one important feature that differentiates SCMs from other frameworks³.

(Pearl 2009a) describes the SCM a process-based tool, because it enables researchers to reflect on their underlying assumptions. The SCM requires more assumptions and thought. Even for a very minimalistic SCM, we need to define an admissible set of variables, ensure the random noise terms are independent and corroborate that the underlying mechanisms are autonomous. By being forced to think about all these steps, SCMs help to avoid poorly specified probabilistic specifications. Various research has pointed out examples where modeling without DAGs lead to severe mistakes: Hirano and Imbens (2001) suggest a method for covariate selection that according to Pearl (2009b) favours bias-enhancing features in the propensity score. Further Bollen and Pearl (2013) (2013) argue that Rosenbaum (2002) and Rubin (2007) falsely declared that ‘there is no reason to avoid adjustment for a variable describing subjects before treatment.’

2.2 Independence of Noise and Mechanism

we can use the definition provided by (Peters, Janzing, and Schölkopf 2017).

Definition: Independence

“The causal generative process of a system’s variables is composed of autonomous modules that do not inform or influence each other. In the probabilistic case, this means that the conditional distribution of each variable given its causes (i.e., its mechanism) does not inform or influence the other conditional distributions, In case we have only two variables, this reduces to an independence between the cause distribution and the mechanism producing the effect distribution”.

Suppose we have only two variables smoking and cancer. Typically, we can express the full joint probability as either:

$$p(s, c) = p(s|c)p(c) \quad (3)$$

$$p(s, c) = p(c|s)p(s) \quad (4)$$

³e.g. the Potential Outcome framework(Pearl 2009a)

In the first example, we define the probability of smoking given one has lung cancer and in the second we look at the conditional probability of lung cancer given smoking. The central idea is that if we specify the causal effect correctly, we can keep the conditional probability of lung cancer given smoking constant while changing the smoking distribution. Clearly, these interventions need to be sensible and do not work for both cases. Writing the probability as smoking behavior given lung cancer and changing lung cancer is an unreasonable intervention. If we specify the causal structure correctly we can undertake local changes on smoking without changing the conditional probability of lung cancer conditional on smoking. One example is looking at smoking behavior in a different country (domain shift). The independence of our mechanisms means that the mechanisms underlying cancer and smoking are independent. If these terms are independent, we can look at local intervention without rephrasing the entire model and keep the other probabilities invariant.

Independence of Noise:

Note, that the way we view noise in SCMs differs from the classical view in regression analysis. Therefore, I will briefly clarify the differences. Noise variables are our latent variables. These unobserved variables are the parent nodes of our child nodes of interest ([Hardt and Recht 2021](#)). In classical regression analysis, such as an ordinary least squares regression model, we employ the Gauss Markov Assumptions to ensure that the estimator is the best linear unbiased estimator (BLUE). The exogeneity assumption, one of the five Gauss Markov Assumptions, suggests that the error terms are uncorrelated with our features ([Wooldridge 2010](#)). For the SCM, the error terms are driving our features and a pivotal component of the model specification ([Hardt and Recht 2021](#)). To ensure that we are correctly specifying mechanisms and our underlying model structure, one needs to ensure independence of these noise terms. As mentioned in the definition for SCMs, we would mathematical notate this independence as follows: $U_C \perp\!\!\!\perp U_E$ where the noise terms of the cause and effect variables are independent. This is pivotal, because we want to derive conditional probabilities for local changes without having to rephrase other variables. There are two related dimensions to consider, namely the informational aspect of independence and implications for modularity. If the independence conditions holds, and the cause and effect variables are independent, the cause and effect variable do not contain information about each other. Modularity ([Pearl 2009a](#)) describes the advantage of being able to treat these variables as distinct modules. This means, even if we have a change in one variable (cause or effect), we can disentangle our variable as unique modular components. Especially looking at machine learning, domain shift and covariate shift are problematic for classical tools. ([Peters, Janzing, and Schölkopf 2017](#))

The central issue of independence of noise variables is that they are latent and henceforth this assumption is untestable. We can only examine this condition in totality but not in isolation⁴. Un-testable assumptions are the reason why many statisticians have distanced themselves from causality in the past ([Pearl 2012a](#)).

⁴via the d-separation criterion ([Hardt and Recht 2021](#))

To accommodate the fact that we can never truly ensure that latent variables are independent, scholarship has proposed the causal sufficiency condition.

Definition: Causal Sufficiency

A set of variables X is usually said to be causally sufficient if there is no hidden common cause $C \notin X$ that is causing more than one variable in X (Peters, Janzing, and Schölkopf 2017; Spirtes 2010)

By ensuring that this condition holds, we can easier disentangle underlying common causes, avoiding a violation of the independence of our noise terms. Independence of noise is one form of causal sufficiency (Peters, Janzing, and Schölkopf 2017).

3 Directed Acyclic Graphs

The most popular graphical model is the directed acyclic graph or in short DAG. Note that a DAG is a specific graphic model but not every SCM has an underlying DAG. Every SCM has an underlying graphical model. A causal graph is a more general graphical model, also including e.g. causal cyclic models. DAGs are predominately used because they are straightforward but simultaneously only apply if our model is truly directed and acyclic. A DAG entails nodes (endogenous and exogenous variables) and edges. Nodes represent our different variables. Edges depict the assignment equations. All edges are directed in the DAG. If we have some edges without arrows, we call that causal graph semi-directed. If we have all edges without arrows, we call that causal graph un-directed. An acyclic graph has no roots that cause itself (directly and indirectly) (Morgan and Winship 2014). This acyclic structure is important for the conditional probabilities (Forré and Mooij 2020). The problem of feedback loops is that we cannot always find unique solutions for the equilibrium state (Peters, Janzing, and Schölkopf 2017).

Figure 1 provides a basic structural causal model. The graph provides an extension of our real world example with another feature namely genetic features. As one can see, genetic features and smoking are autonomous equations and we can derive their conditional distributions in a modular manner. Further, the U_s , U_g and U_c are the expressive noise variables. The square nodes represent the latent variables. The circle nodes represent the observed variables. All edges are directed and there are no variables causing itself. We are dealing with a DAG.

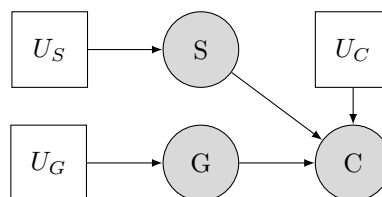


Figure 1: Structural Causal Model

We can obtain a graphical model based on (a single) observational dataset through conditional independence testing. The graphical approach to recover a graph from an observational dataset requires two assumptions namely jointly independent noise terms (which is assessed via the markov condition) and faithfulness.⁵.

Intuitively speaking, this method examines how noise spreads (Peters, Janzing, and Schölkopf 2017). One of the more prominent algorithm to estimate the underlying DAG structure is the pc-algorithm (Kalisch et al. 2012). Two other methods are ICM⁶ and using an additive noise model⁷. Graphical models developed a mathematical language to assess certain conditions. Two prominent examples are the causal markov condition and the backdoor criterion. The backdoor criterion allow us to select an admissible subset of variables (Pearl 2009a). The causal markov condition allows us to test independence in a graph, ensuring us that we are able to utilize the causal factorization (Peters, Janzing, and Schölkopf 2017).

4 SCMs and Time

In causal modelling we largely ignore time. Most SCMs focus on specifying an equilibrium states post-intervention. Notions of time differ in different disciplines. Time in social sciences is often vague and less exact. Natural sciences such as physics treat time in a concise manner, by explicitly including time in models via differential equations [Peters, Janzing, and Schölkopf (2017);(Mooij, Janzing, and Schölkopf 2013)]. This is beneficial if we study non-equilibrium states because we can actual observe the react at different time points rather than only modelling the post intervention outcome. Dynamical modelling is becoming more relevant in social sciences. E.g. (Creager et al. 2020) suggest that static concepts in fair-algorithm decision making policies perform very differently if we model these policies as dynamic systems.

We can recast structural causal models based on differential equations, too. The first step of casting a SCM with time as a explicit factor is defining an initial state.

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{5}$$

Note, that here the initial state is pivotal because there is a dependence on the prior time points. Subsequently, we need to define our function \mathbf{x} at time point t . We can take a partial derivative with respect to time t of our function \mathbf{x} .

⁵For further information, see (morgan_winship_2014:1?; Peters, Janzing, and Schölkopf 2017)

⁶The intuition here is that the noise terms cultivate the footprint of the assignments. For more information see: Shajarisales et al. (2015)

⁷This method uses regression and conditional independence testing to disentangle graphical structures. For further information see Mooij et al. (2016)

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d \quad (6)$$

Subsequently, we combine these elements and look at time $t + \text{change in } t$:

Value of Function at time $t + dt$:

$$\mathbf{x}(t + dt) = \mathbf{x}(t) + dt \cdot f(\mathbf{x}(t)) \quad (7)$$

Henceforth, there is a dependence on the initial state and then subsequently we combine this effect with the change. In our smoking example, one could perhaps think of a situation where we want to examine whether smoking at certain time points is more harmful (impact of smoking during cancer treatment looking at exact time points).

For further context, (Peters, Janzing, and Schölkopf 2017) provide a taxonomy of different methods (see table 1). This table looks at different modelling approaches and how they perform in different settings.

Table 1 Source: Peters et al. (2017) , Modelling Taxonomy

model	IID setting	changing distributions	counter-factual questions	physical insight
mechanistic model	Y	Y	Y	Y
structural causal model	Y	Y	Y	N
causal graphical model	Y	Y	N	N
statistical model	Y	N	N	N

In the table we can see that traditional statistical models fair well in IID settings, where our observations are independent and identically distributed. But once we encounter a changing distribution such as an covariate shift or a domain shift, these statistical models are not suitable anymore. Statistical models are unable to answer any counterfactual questions about hypothetical settings outside of our observational data and are unable to provide physical insights because we do not model time as an explicitly. Causal graphical models are able to deal with changing distributions and fair well in the IID setting. Nevertheless,

they are unable to answer counterfactual questions on its own and cannot provide physical insights. SCMs are able to deal with the first three queries, but largely ignore time. By commonly making assumptions such as that the effect is acyclic, SCMs simplify time and treat focus on equilibrium states. The mechanistic view, where we model time as an explicit factor, enables us to answer all those queries. Note that the growing complexity makes it very challenging to implement but relevant for natural sciences where time is pivotal to disentangle relationships (Mooij, Janzing, and Schölkopf 2013; Peters, Janzing, and Schölkopf 2017).

5 Pearl’s Causal Hierachy

Pearl (2009a)] introduced the hierarchy of causation to categorize different statistical and causal tools. The hierarchy of causation contains (see table 1) three levels, where the high methods on the hierarchy, the more information the method requires. This section discusses each of these methods and their respective advantages. I also included the usage of different methods and briefly discuss how they relate to the hierarchy based on (Bareinboim et al. 2020).

Table 2 Source: Pearl (2009) , Hierarchy of Causation

Method	Action	Example	Usage
Association $P(a b)$	Co-occurrence	What happened...	(Un-)Supervised ML, BN, Reg.
Intervention $P(a do(b), c)$	Do-manipulation	What happens if ...	CBN,MDP,RL
Counterfactual $P(a_b a', b')$	Hypotheticals	What would have happened if...	SCM ,PO

5.1 Association:

The first level, association, requires the least information. This query deals with questions like ‘what happen?’ In our smoking example, a possible question is e.g. ‘what was the impact of smoking on lung cancer?’ Association based methods are most prevalent and contain the largest class of methods. Standard statistical tools such as regression analysis, supervised and unsupervised learning and Bayesian Networks all fall into this category (Bareinboim et al. 2020). The underlying action for association is co-occurrence. As prominently criticized by Bender et al. (2021), this reduces the number of questions we can answer, because methods are heavily dependent on the observational data. In the context of deep learning and the advancement of natural language processing, Bender et al. (2021) suggest that many association-based methods results in stochastic parrots as opposed to natural language understanding. Association-based methods ignore external changes outside of our data. The interventional distribution has information on these external changes. Note, that the intervention distribution is only defined in high order methods. Reiterating

our lung cancer example, the full joint distribution looks as follows:

$$P(c, s) = P(s) \times P(c|s)$$

If there are any changes in our distribution, e.g. smoking changes its distribution from s to s_{new} we are unable to accommodate these changes (Peters, Janzing, and Schölkopf 2017). These conditional probabilities are typically not derived from functions $do(S = s)$ in these methods. They are based on observational data for the given methods. As mentioned in section 1, the amount of things we can derive based on a specific observational dataset is limited (Hernán and Robins 2020; Hernán and Taubman 2008). Association-based methods are not equipped to accommodate such modifications.

There are ways to also graphically depict associations-based models. One example is the Bayesian Network (Pearl 2009a). In the probabilistic representation (see figure 2), we ignore latent factors (Creager et al. 2020; Pearl 2009a). We can see, that this model is technically speaking also a directed acyclic graph.

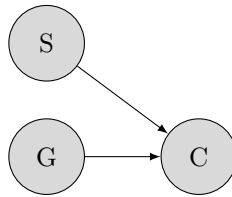


Figure 2: Probabilistic Model

While in this simple 2 feature case, the graphical illustration is rather straight forward. But especially in a high dimensional space, these specifications become convoluted.

5.2 Intervention:

Intervention deals with questions like ‘what happens if.’ E.g. in our smoking example, we could look at e.g. ‘what happens if people smoke less?’ or, ‘what happens if people smoke more?’ This higher order methods opens many different applications and is predominately possible in Causal Bayesian Networks, Reinforcement Learning or Markov Decision Processes. For interventions, we can use Pearl (2009a) do-calculus. The do-calculus enables us to study the manipulation of parent nodes. Instead of merely seeing the co-occurrence of variables, we can actively manipulate the conditional distribution of one variable. Commonly, one could use this method to evaluate different policies (Creager et al. 2020). In our cancer example one can actively set the smoking behavior to a fixed value or a different conditional probability. The post intervention joint distribution would look as follows: $P_{S=s}(c, g) = P(c|S = s, g) \times P(g)$ where $S = s$ is the new probability or atomic value. There are various types of intervention. I will illustrate atomic intervention and policy intervention based on our smoking example.

In **atomic intervention**, we set a variable to a constant value. As one can see in figure 3, c is constant that is

not dependent on the latent factor, because in atomic intervention we do not derive the value of c based on the function S .

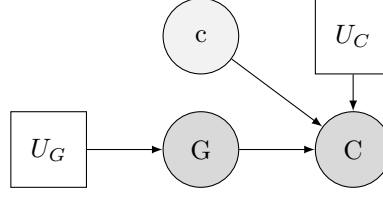


Figure 3: Atomic Intervention

Further, in mathematical notation the model would change as follows:

$$S := s \quad (8)$$

$$G := f_G(U_G) \quad (9)$$

$$C := f_C(S, G, U_C) \quad (10)$$

In **policy intervention** (see figure 4) we specify a different conditional probability $do(S = s)$ for an equation. We can derive s from S , because we include information on the intervention distribution. This information cannot be obtained, if we directly specify our model as conditional probabilities.

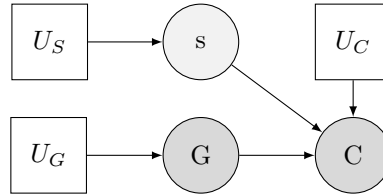


Figure 4: Policy Intervention

where s is our new conditional probability. One way of the most prominent ways to estimate casual effects of our treatment is the **average treatment effect**. The average treatment effect looks at the difference between treatment and control group. For the case of simplicity in our example, we want to look at whether smoking at all has an impact on lung cancer. Mathematically, this could as follows:

$$f(C|S = 1) - f(C|S = 0) \quad (11)$$

where $S = 1$ means the participant smokes and $S = 0$ means the participant does not smoke. Furthermore, it is crucial that we hold all other factors constant (e.g. age, gender, genetic code).

Confounders

Confounding is the circumstance where the observational information (conditional probabilities in our data) and intervention are different. Essentially, we want to express our do-intervention based on conditional probabilities, but have no direct information in our data. To still estimate a conditional probability, we can use a mixture of existing conditional probabilities and estimate the desired probability based on the adjustment formula. For further information, see (Barocas, Hardt, and Narayanan 2019).

5.3 Counterfactuals

Counterfactuals deal with hypothetical situations. Typically questions are: ‘what would have happened if.’ In our smoking case, suppose we want to see what would have happened if we wanted to include a different problem behavior that was not included in the observational data. Philosophically speaking, we interpret these outcomes as actual outcomes⁸. The two most prominent models that can obtain counterfactuals are the structural causal model and the potential outcome framework. This section focuses on counterfactuals in structural causal models. For information on how to obtain counterfactuals in the potential outcome framework, see Cunningham (2021).

SCMs derive Counterfactuals. We can describe the process as follows:

- (a) Abduction: Cast probability $P(u_s)$ as conditional probability $P(u_s|\epsilon)$
- (b) Action: Exchange ($S = s'$)
- (c) Prediction: Compute ($C = c'$)

In the first step, the abduction step, we cast the new conditional probability for our latent factor. This probabilities is based on the events in our data which are denoted as ϵ . As one can see, we can only cast the probability for latent factors if we specify them in our model. Henceforth, models such as the Bayesian Causal Network or a Markov Decision Process cannot provide counterfactuals because they do not include information on these factors. In the second step, we exchange S for s' . In the third step, we compute the outcome c' for this hypothetical setting.

Regardless of whether dealing with counterfactuals in SCMs or in the potential-outcome-Framework, we need to ensure three assumptions. Here, I am using the definitions by (Hardt and Recht 2021).

Stable Unit Treatment Value Assumption {SUTVA}: ‘The treatment that one unit receives does not change the effect of treatment for any other unit.’

Consistency: The outcome Y agrees with the potential outcome corresponding to the treatment indicator.’

The first two conditions hold for counterfactuals in structural causal models. As suggested by (Hardt and Recht 2021), the third condition ensures that we are dealing with a perfect randomized controlled trial.

⁸In other frameworks such as the Potential outcome framework we interpret them as potential outcomes rather than true outcomes (Hardt and Recht 2021)

Ignorability: The potential outcomes are conditionally independent of treatment given some set of de-confounding variables.

We can never truly ensure that we are undertaking an experiment under perfect conditions, because we cannot ensure that we include all pivotal variables. Therefore, the third condition is an un-testable assumption. Due to the backdoor criterion, we can nevertheless verify, that our model specification is consistent with this assumption ([Hardt and Recht 2021](#)). Note, that this is merely consistency and not a test that allows us to corroborate this assumptions entirely. The backdoor criterion is unique to causal graphs. E.g. the potential outcome framework cannot utilize this test because they neglect graphical implementations of their underlying mathematical models.

6 Considerations

Structural causal models are data driven models ([Hernan 2015](#)). There is a strong dependence on data quality. In a situation, in which we are certain that the quality of observational data is low, using a SCM is not advisable. Alternatively, if we have a very strong theoretical understanding, we might be more interested in other modelling approaches such as Markov Decision Processes ([Hernan 2015](#)). Evaluating different sources of information into causal models accordingly is an open problem in causal modelling ([Spirites 2010](#)). If our insights based on data are limited, or perhaps we are not really able to clearly define function relationships beyond conditional probabilities, structural causal models can be problematic. Furthermore, sometimes association-based knowledge is sufficient. inevitably, the entry barrier to higher order causal models is higher in terms of required knowledge. Not every situation accommodates these conditions. Nevertheless, especially in a situation where we have high-stake repercussions based on our model, causal modelling has been a complementary tool. When dealing with policy implementations, that might be controversial or unethical in the real world, simulations are pivotal. Recently, research has fairness has adopted various concepts and developed new characteristics such as counterfactual fairness ([Kusner et al. 2018](#)).

7 Conclusion

Structural causal models are flexible simulators to disentangle causality for higher order queries such as interventions and counterfactuals. We make a number of different assumptions when building a structural causal model. The most important assumption is the independence of mechanisms and independence of noise terms which are the core of our simulation. There are many advantages of structural causal models: We are able to model latent factors forcing us to reconsider existing assumptions about the relationship in our data. Further, we get a underlying graphical representation including a mathematical language for this graphical systems to test causal assumptions that are otherwise untestable. Additionally, one is able

to model queries beyond mere association going as far as dealing with hypothetical situations. Apart from the assumptions, causal graphics can help researchers avoid mistakes. To determine this causal graph. We can use conditional independence testing and assess assumptions about faithfulness and jointly independent noise terms in the graphical approach. Looking at the causal hierarchy, structural causal modelling enables us to compute interventions efficiently given certain conditions hold. These conditions do not always hold true in every instance. The entry barriers to causal modeling are high and involve considerable investments in terms of data and time.

8 References

- Bareinboim, Elias, JD Correa, Duligur Ibeling, and Thomas Icard. 2020. "On Pearl's Hierarchy and the Foundations of Causal Inference." *ACM Special Volume in Honor of Judea Pearl (Provisional Title)*.
- Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2019. *Fairness and Machine Learning*. fairmlbook.org.
- Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23.
- Bollen, Kenneth A, and Judea Pearl. 2013. "Eight Myths about Causality and Structural Equation Models." In *Handbook of Causal Analysis for Social Research*, 301–28. Springer.
- Chiappa, Silvia, and William S. Isaac. 2019. "A Causal Bayesian Networks Viewpoint on Fairness." *arXiv:1907.06430 [Cs, Stat]* 547: 3–20. https://doi.org/10.1007/978-3-030-16744-8_1.
- Cortés, Efrén Cruz, and Debashis Ghosh. 2019. "A Simulation Based Dynamic Evaluation Framework for System-Wide Algorithmic Fairness." *arXiv:1903.09209 [Cs]*, March. <http://arxiv.org/abs/1903.09209>.
- Creager, Elliot, David Madras, Toniann Pitassi, and Richard Zemel. 2020. "Causal Modeling for Fairness in Dynamical Systems." In *Proceedings of the 37th International Conference on Machine Learning*, edited by Hal Daumé III and Aarti Singh, 119:2185–95. Proceedings of Machine Learning Research. PMLR.<http://proceedings.mlr.press/v119/creager20a.html>.
- Cunningham, Scott. 2021. *Causal Inference: The Mixtape*. Yale University Press.
- Forré, Patrick, and Joris M Mooij. 2020. "Causal Calculus in the Presence of Cycles, Latent Confounders and Selection Bias." In *Uncertainty in Artificial Intelligence*, 71–80. PMLR.
- Hardt, Moritz, and Benjamin Recht. 2021. *Patterns, Predictions, and Actions: A Story about Machine Learning*. <https://mlstory.org>. <http://arxiv.org/abs/2102.05242>.
- Hernan, M. A. 2015. "Invited Commentary: Agent-Based Models for Causal Inference–Reweight Data and Theory in Epidemiology." *American Journal of Epidemiology* 181 (2): 103–5. <https://doi.org/10.1093/aje/kwu272>.
- Hernán, Miguel A, and James M Robins. 2020. "Causal Inference: What If," 311.
- Hernán, Miguel A, and Sarah L Taubman. 2008. "Does Obesity Shorten Life? The Importance of Well-Defined Interventions to Answer Causal Questions." *International Journal of Obesity* 32 (3): S8–14.
- Hirano, Keisuke, and Guido W Imbens. 2001. "Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catheterization." *Health Services and Outcomes Research Methodology* 2 (3): 259–78.

- Kalisch, Markus, Martin Mächler, Diego Colombo, Marloes H Maathuis, and Peter Bühlmann. 2012. "Causal Inference Using Graphical Models with the r Package Pcalg." *Journal of Statistical Software* 47 (11): 1–26.
- Karimi, Amir-Hossein, Bernhard Schölkopf, and Isabel Valera. 2020. "Algorithmic Recourse: From Counterfactual Explanations to Interventions." *arXiv:2002.06278 [Cs, Stat]*, October. <http://arxiv.org/abs/2002.06278>.
- Kilbertus, Niki, Mateo Rojas-Carulla, Giambattista Parascandolo, Moritz Hardt, Dominik Janzing, and Bernhard Schölkopf. 2018. "Avoiding Discrimination Through Causal Reasoning." *arXiv:1706.02744 [Cs, Stat]*, January. <http://arxiv.org/abs/1706.02744>.
- Kusner, Matt J., Joshua R. Loftus, Chris Russell, and Ricardo Silva. 2018. "Counterfactual Fairness." *arXiv:1703.06856 [Cs, Stat]*, March. <http://arxiv.org/abs/1703.06856>.
- Kügelgen, Julius von, Amir-Hossein Karimi, Umang Bhatt, Isabel Valera, Adrian Weller, and Bernhard Schölkopf. 2021. "On the Fairness of Causal Algorithmic Recourse." *arXiv:2010.06529 [Cs, Stat]*, February. <http://arxiv.org/abs/2010.06529>.
- Mooij, Joris M, Dominik Janzing, and Bernhard Schölkopf. 2013. "From Ordinary Differential Equations to Structural Causal Models: The Deterministic Case." *arXiv Preprint arXiv:1304.7920*.
- Mooij, Joris M, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. 2016. "Distinguishing Cause from Effect Using Observational Data: Methods and Benchmarks." *The Journal of Machine Learning Research* 17 (1): 1103–204.
- Morgan, Stephen L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. Analytical Methods for Social Research. Cambridge University Press. <https://doi.org/10.1017/CBO9781107587991>.
- Ness, Robert Osazuwa, Kaushal Paneri, and Olga Vitek. 2019. "Integrating Markov Processes with Structural Causal Modeling Enables Counterfactual Inference in Complex Systems." *arXiv:1911.02175 [Cs, q-Bio, Stat]*, November. <http://arxiv.org/abs/1911.02175>.
- Oberst, Michael, and David Sontag. 2019. "Counterfactual Off-Policy Evaluation with Gumbel-Max Structural Causal Models." In *International Conference on Machine Learning*, 4881–90. PMLR.
- Pearl, Judea. 2009a. *Causality*. Cambridge university press.
- . 2009b. "Myth, Confusion, and Science in Causal Analysis."
- . 2010. "An Introduction to Causal Inference." *The International Journal of Biostatistics* 6 (2): 1–62.
- . 2012a. "The Causal Foundations of Structural Equation Modeling." California Univ Los Angeles Dept of Computer Science.

- . 2012b. “The Do-Calculus Revisited.” *arXiv Preprint arXiv:1210.4852*.
- Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press.
- Rosenbaum, Paul R. 2002. “Overt Bias in Observational Studies.” In *Observational Studies*, 71–104. Springer.
- Rubin, Donald B. 2007. “The Design Versus the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials.” *Statistics in Medicine* 26 (1): 20–36.
- Shajarisales, Naji, Dominik Janzing, Bernhard Schoelkopf, and Michel Besserve. 2015. “Telling Cause from Effect in Deterministic Linear Dynamical Systems.” In *International Conference on Machine Learning*, 285–94. PMLR.
- Spirtes, Peter. 2010. “Introduction to Causal Inference.” *Journal of Machine Learning Research* 11 (5).
- Tarka, Piotr. 2018. “An Overview of Structural Equation Modeling: Its Beginnings, Historical Development, Usefulness and Controversies in the Social Sciences.” *Quality & Quantity* 52 (1): 313–54.
- Turbin, Mark S, Richard Jessor, and Frances M Costa. 2000. “Adolescent Cigarette Smoking: Health-Related Behavior or Normative Transgression?” *Prevention Science* 1 (3): 115–24.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.

9 Appendix

Table 2: Source: Pearl (2009)

Method	CBN	SCM
Prediction	<ul style="list-style-type: none"> · Unstable · Volatile to parameter changes · Re-Estimate entire model 	<ul style="list-style-type: none"> · Stable · More Natural Specification · Only estimate ΔCM
Intervention	<ul style="list-style-type: none"> · Costly for Non-Markovian Models · Unstable(Nature Conditional prob.) · Only generic estimates(ΔCP) 	<ul style="list-style-type: none"> · Pot. Cyclic Representation · Stable(Nature Equation) · Context specific(Invariance of Equation)
Counterfactuals	<ul style="list-style-type: none"> · Impossible · No information on latent factors(ϵ) 	<ul style="list-style-type: none"> · Possible · Inclusion of latent factors