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 thes 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

For many research problems, we want to understand the relationship between variables beyond co-occurrence. The most popular causal model is the structural causal model or in short SCM. The geneticist and statistican 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 model. Many scholars challenge the parameter-specification in a SEM (Pearl 2012a).[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 interchargably.

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 conditional probabilities, 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. 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. Specifically, this paper focuses on the independence of noise, independence of mechanisms, causal sufficiency, the role of time and the Causal Markov Condition. Additionally, this paper provides an overview of various other statistical and causal methods within the grand scheme of pearls causal hierarchy. I accommodate this outlook with various considerations for the usage of SCMs. The aim of this paper is to provide an intuitive understanding of different concepts within causality using elements from core literature on causality.

The rest of the paper is structured as follows:

Subsection 2.1 introduces a baseline example for both a simple set of assignments and the underlying

---

[1] Inconsistency refers to inherent nature of creating parameters based of observational data. Observational data is seldom consistent. Henceforth respective estimates are often questioned. For further information see Hernán and Taubman (2008)

causal graph. Subsection 2.2 provides further context to assumptions in causal modelling. Section 3 discusses causal graphs and how we derive underlying graphical representations based on single observational datasets. Section 4 examines causality and the role of time. Section 5 studies the causal hierachy. Section 6 suggests instructions on when and when not to employ a structural causal model.

2   Assumptions in Structural Causal Models

A SCM contains a (sub-)set of autonomous equations. These equations are asymmetric assignments, because they are not bi-directional. Equations in causal models did not always have a concise notation (Pearl, 2009). First, there was no sign to express the assignment equation and people used the '=' and one would e.g. write $A = B$. Treating an equation as a **algebraic equation** led to confusion because those have no causal information. This algebraic equation would imply that $B = A$ because the order has no concrete meaning in algebraic equations. The problem is that the equation is symmetric. The initial $'='$ sign was replaced with the ':=' which is asymmetric (Pearl, 2009) and called an **assignment**. This misconception has caused a lot of challenges.[2]. We cannot change the sides of the equation as one can do with regular equations (Pearl 2012a). (Peters, Janzing, and Schölkopf 2017) define a SCM as follows:

Definition 1: Structural Causal Model:

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

$$C := N_C$$
$$E := f_E\left(C, N_E\right)$$

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

In their definition, C is the cause and E is the effect.

2.1   Mathematical Components

We can set up an epidemiological example with three variables, looking at the impact of problem behavior and genetic code on lung cancer. Problem behavior is a latent variable, but we can use observational data to characterize problem behavior. One example of problem behavior is smoking. Henceforth,to estimate problem behaviour, we can for instance ask participants how frequently they smoke. For genetic code, we could look for specific genes and and examine whether the presence of specific genes has an impact on getting cancer. 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).

---

[2]For more information see Pearl (2009a)

$$S := f_S(U_S) \tag{1}$$

$$G := f_G(U_G) \tag{2}$$

$$C := f_C(S, G, U_C) \tag{3}$$

where: $\{S\}$ - Frequency of smoking $\{G\}$ - Presence of specific genes $\{C\}$ - 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 frameworks[3].

(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 minimalisitic 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

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" (Peters, Janzing, and Schölkopf 2017).

If we specify the causal structure correctly:

- possible to undertake local intervention -> change f(smoking), regardless of f(cancer | smoking)
- these components are autonomous objects -> set of autonomous assignments

Independence of Noise:

The way we view noise in SCMs differs from the classical view in regression analysis. Therefore, I will briefly clarify essential differences. Noise variables are our latent variables which are the parent nodes of

---

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

our child nodes of interest (Hardt and Recht 2021). In classical regression analysis, such as an ordinary least squares regression, we employ the Gauss Markov Assumptions. 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: $N_F \perp\!\!\!\perp N_C$

Scholarship in causality established 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).

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) By using these independence assumptions in SCMs, we can accommodate these settings.

Looking at our smoking example:

$$P(Cancer|Smoking) = P(Smoking|Cancer) \times P(Cancer)$$

Mechanism

$$\begin{aligned} p(a,t) &= p(a \mid t)p(t) \\ &= p(t \mid a)p(a) \end{aligned} \qquad (4)$$

$$p^{o}(a,t) = p(t \mid a)p^{o}(a) \text{ and } p^{s}(a,t) = p(t \mid a)p^{s}(a). \qquad (5)$$

$$P_x(v) = \prod_{\{i|V_i \notin X\}} P\left(v_i \mid pa_i\right)$$

for all v consistent with x

3 Directed Acyclic Graphs

The most popular graphical model is the directed acyclic graph or in short DAG. A causal graph for a SCM contains endogenous and exogenous variables. A DAG entails nodes and edges.

Nodes represent our different variables. The square nodes represent the latent variables. The circle nodes represent the observed variables. Edges depict the assignment equations. All edges are directed in the DAG. If we have some edges without arrows, we call that graph semi-directed. If we have all edges without arrows, we call that 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). Variables have incoming paths from their parent nodes.
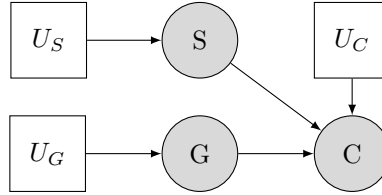

Figure 1: Structural Causal Model

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 and faithfulness. 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 ICM[4] and using an additive noise model[5].

Graphical models developed a mathematical language to assess certain conditions.

---

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

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

4 SCMs and Time

largely ingore time

time in mechanical modeling crucial (Peters, Janzing, and Schölkopf 2017).

- Time in Social Sciences: Often Vague
- Time in Physical Sciences: Mechanical via **Differential equations**
- dependence on prior time point and change in time contribute to the value at time point

Initial Value:

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{6}$$

Derivative of function x with respect to time t:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d \tag{7}$$

Value of Function at time t + dt:

$$\mathbf{x}(t + dt) = \mathbf{x}(t) + dt \cdot f(\mathbf{x}(t)) \tag{8}$$

(Peters, Janzing, and Schölkopf 2017) provide a taxonomy of different methods (see table 1)

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 (Bareinboim et al. 2020; Pearl 2009a) introduced the hierachy of causation to categorize different statistical and causal tools. The hierachy 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.

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. Association based methods are most prevalent and contain the largest class of methods. Standard statistical tools such as regression analysis, supervised and unsupervsied learning and Bayesian Networks all fall into this category.(Bareinboim et al. 2020) The underlying action for association is co-occurence. 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 langauge 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:

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

If there are any changes in our distribution, e.g. smoking changes its distribution from s to s_{new} we are unable to accomodate these changes (Peters, Janzing, and Schölkopf 2017). Association-based methods are not equipped to accomodate such modifications.

There are ways to also graphically depict associational models. One example is the Bayesian Network, which uses conditional probabilities instead of functions to describe the relationship between variables (Pearl 2009a) In the probabilistic representation, we ignore latent factors (Creager et al. 2020; Pearl 2009a).
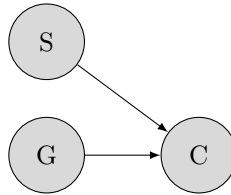

Figure 2: Probabilistic Model

While in this simple 3 feature case, the graphical illustration is rather straight forward. But especially in a high dimensional space, the graphical models become convoluted. Henceforth, finding the correct specification without reconsidering all modelling assumptions can led to mistakes as discussed in section 2.

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-occurence 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.
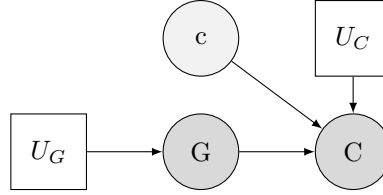

Figure 3: Atomic Intervention

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

$$S := s \tag{9}$$

$$G := f_G(U_G) \tag{10}$$

$$C := f_C(S, G, U_C) \tag{11}$$

In **policy intervention** (see figure 4) we specify a different conditional probability $(S = s)$ 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.
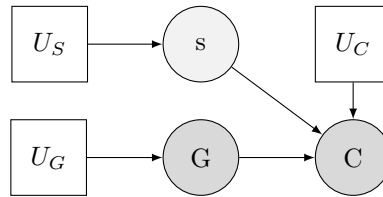

Figure 4: Policy Intervention

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:

$$f(C|S = 1) - f(C|S = 0) \tag{12}$$

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 so looking at our example, we would compare people with the same genetic features to truely isolate the effect of smoking. The ATE is a population level effect (Hardt and Recht 2021).

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. Philosophically speaking, we interpret these outcomes as actual outcomes[6]. 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).

We can describe the process as follows:

  (a) Abduction: Cast probability $P(u)$ as conditional probability $P(u|\epsilon)$
  (b) Action: Exchange $(X = x)$
  (c) Prediction: Compute $(Y = y)$

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.

Stable Unit Treatment Value Assumption {SUTVA}

'The treatment that one unit receives does not change the effect of treatment for any other unit.'

Consistency

---

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

The outcome Y agrees with the potential outcome corresponding to the treatment indicator.'

Ignorability

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

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

6 Considerations

Structural causal models are data driven models (Hernan 2015). There is a dependence on data quality. Evaluating different sources of information into causal models accordingly is an open problem in causal modelling (Spirtes 2010). Henceforth, 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 accomodates 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 manifold different 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 indepedence 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 hyptoethical situations.

Apart from the assumptions, causal graphics can help reseachers avoiding mistakes. To determine this causal graph, we can use conditional independence testing and assess assumptions about faithfulness and jointly independent noise terms.

Looking at the causal hierachy, structural causal modelling enables us to compute interventions efficiently given certain condition hold. These conditions defintely not always hold true in every instance. The entry barriers to causal modeling are high and in 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–Reweighting 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. n.d. "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, Massachuestts: 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.