Cyclic Counterfactuals under Shift–Scale Interventions

Saptarshi Saha *
Computer Vision and Pattern Recognition Unit
Indian Statistical Institute
Kolkata, West Bengal - 700108, India
saptarshi.saha_r@isical.ac.in

Dhruv Vansraj Rathore
Indian Statistical Institute
Kolkata, West Bengal - 700108, India
cs2306@isical.ac.in

Utpal Garain
Computer Vision and Pattern Recognition Unit
Indian Statistical Institute
Kolkata, West Bengal - 700108, India
utpal@isical.ac.in

Abstract

Most counterfactual inference frameworks traditionally assume acyclic structural causal models (SCMs), i.e. directed acyclic graphs (DAGs). However, many real-world systems (e.g. biological systems) contain feedback loops or cyclic dependencies that violate acyclicity. In this work, we study counterfactual inference in cyclic SCMs under *shift–scale interventions*, i.e., soft, policy-style changes that rescale and/or shift a variable's mechanism.

1 Introduction

Most research on counterfactual reasoning (Pawlowski et al., 2020; Sanchez and Tsaftaris, 2022; Saha and Garain, 2022; Komanduri et al., 2024; Melistas et al., 2024; Wu et al., 2025; Kügelgen et al., 2023; Kladny et al., 2024) assumes that the underlying causal structure among variables can be represented by a Directed Acyclic Graph (DAG). However, this acyclicity assumption is often violated in real-world systems. For instance, gene regulatory networks frequently exhibit feedback loops, leading to cyclic dependencies that DAGs cannot capture. Such cycles are integral to the dynamic behavior of biological systems and are crucial for understanding processes like cell differentiation and immune responses. Given the prevalence of cycles in such systems and the availability of detailed perturbation data, there is a compelling case for extending counterfactual inference frameworks to accommodate cyclic structures.

However, progress in this area has stalled due to the need for new theoretical breakthroughs, as many properties that hold in acyclic models no longer apply when feedback loops are present. A fundamental issue is that a set of structural equations with cycles may not have a unique solution for the endogenous variables. Solvability refers to the existence of at least one solution (equilibrium), and unique solvability means there is exactly one solution (almost surely). If an SCM is not uniquely solvable —i.e., not a simple SCM—it might generate multiple different outcome distributions or undefined behavior under interventions. Although the class of simple SCMs includes acyclic SCMs as a special case, the theory remains much less developed for the cyclic setting.

In causality, a hard (structural) intervention (Pearl's do-operator) (Pearl et al., 2016) sets a variable to a fixed value, severing its dependence on its usual causes. In contrast, shift and scale interventions

* first author

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

are types of soft (parametric) interventions that modify the value of a variable by some function (such as adding or multiplying by a constant) without removing its original input links. In particular, a hard intervention is just a degenerate case of a soft intervention (Massidda et al., 2023). This generality means one can ask nuanced "what-if" questions: What if everyone received 20% more of the drug? What if we lowered each student's class size by 5? Such policies cannot be represented as a simple $do(X = x)$ since they depend on individuals' original X. Soft interventions are strictly more expressive in defining counterfactual worlds. More specifically, a shift can implement dynamic-like policies ("increase dose for those who had high risk") that static *do*-interventions cannot capture. Shift interventions have been used to learn causal cyclic graphs (Rothenhäusler et al., 2015), to match a desired causal state (Zhang et al., 2021). Lorch et al. (2024) use shift-scale intervention in causal modeling with stationary diffusions. However, the theoretical foundations supporting their use remain underdeveloped.

Contributions. In this work, we develop a theoretical framework for counterfactual inference in *cyclic causal models* under *shift–scale interventions*. Our main contributions are as follows:

- We show that *SCMs satisfying a global contraction condition* are *simple*—i.e., uniquely solvable with respect to every subset of variables—even in the presence of cycles.
- We prove that under *shift–scale interventions with bounded scale coefficients* (i.e., $|a_j| \leq 1$), the intervened *twin SCM remains uniquely solvable*, ensuring the well-posedness of counterfactual queries.
- We establish that this class of shift–scale interventions is *closed under composition*, making it algebraically stable for sequential interventional analysis.
- Under an additional *Lipschitz regularity condition in the exogenous noise*, we derive *sub-Gaussian tail bounds* for counterfactual functionals, showing that the distribution of counterfactual outcomes concentrates sharply around their mean.

2 Background and Problem Setup

This section provides a brief overview of SCMs and establishes the notational framework adopted throughout the paper. Our exposition aligns with the formalism presented in Bongers et al. (2021).

Definition 1 (Structural Causal Model). *A* Structural Causal Model *(SCM) is defined as a tuple*

$$\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle,$$

where $\mathcal{I}$ denotes a finite set indexing the endogenous variables*, $\mathcal{J}$ denotes a disjoint finite set indexing the* exogenous variables*, $\mathcal{X} = \prod_{i \in \mathcal{I}} \mathcal{X}_i$ is the joint domain of the endogenous variables, with each $\mathcal{X}_i$ being a standard measurable space, $\mathcal{E} = \prod_{j \in \mathcal{J}} \mathcal{E}_j$ is the joint domain of the exogenous variables, with each $\mathcal{E}_j$ also a standard measurable space, $f : \mathcal{X} \times \mathcal{E} \to \mathcal{X}$ is a measurable function representing the causal mechanisms that determine the values of endogenous variables from both endogenous and exogenous inputs, $\mathbb{P}_{\mathcal{E}} = \prod_{j \in \mathcal{J}} \mathbb{P}_{\mathcal{E}_j}$ is a product probability measure over $\mathcal{E}$, describing the joint distribution of the exogenous variables.*

Definition 2 (Solution of an SCM). *A* solution *of $\mathcal{M}$ is a pair of random variables $(\mathbf{X}, \mathbf{E})$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that:*

(i) $\mathbf{E} : \Omega \to \mathcal{E}$ *has distribution* $\mathbb{P}_{\mathcal{E}}$*;*

(ii) $\mathbf{X} : \Omega \to \mathcal{X}$ *satisfies the structural equations*

$$\mathbf{X} \;=\; f(\mathbf{X}, \mathbf{E}) \quad \mathbb{P}\text{-almost surely.}$$

For convenience, we say that a random variable X is a solution of $\mathcal{M}$ if there exists an exogenous random variable $\mathbf{E}$ such that the pair $(\mathbf{X}, \mathbf{E})$ constitutes a solution of $\mathcal{M}$

Definition 3 (Parent). *For $i \in \mathcal{I}$, an index $k \in \mathcal{I} \cup \mathcal{J}$ is called a* parent *of i iff there does* not *exist a measurable map $\tilde{f}_i : \mathcal{X}_{\setminus k} \times \mathcal{E}_{\setminus k} \to \mathcal{X}_i$ such that for $P_{\mathcal{E}}$-almost every $e \in \mathcal{E}$ and all $x \in \mathcal{X}$,*

$$x_i = f_i(x, e) \iff x_i = \tilde{f}_i(x_{\setminus k}, e_{\setminus k}) \,.$$

Exogenous variables have no parents. We write $\mathrm{pa}(i)$ for the set of parents of i and extend to sets by $\mathrm{pa}(O) := \bigcup_{i \in O} \mathrm{pa}(i)$.

Definition 4 (Unique Solvability)**.** *An SCM $\mathcal{M}$ is* uniquely solvable with respect to *a subset $\mathcal{O} \subseteq \mathcal{I}$ if there exists a measurable function*

$$g_{\mathcal{O}} : \mathcal{X}_{\text{pa}(\mathcal{O})\setminus\mathcal{O}} \times \mathcal{E}_{\text{pa}(\mathcal{O})} \to \mathcal{X}_{\mathcal{O}}$$

such that for all $x \in \mathcal{X}$ *and $\mathbb{P}_{\mathcal{E}}$-almost* every $e \in \mathcal{E}$*,*

$$x_{\mathcal{O}} = g_{\mathcal{O}}(x_{\text{pa}(\mathcal{O})\setminus\mathcal{O}}, e_{\text{pa}(\mathcal{O})}) \quad \Longleftrightarrow \quad x_{\mathcal{O}} = f_{\mathcal{O}}(x, e).$$

Definition 5 (Simple SCM)**.** *An SCM $\mathcal{M}$ is called* simple *if it is uniquely solvable with respect to every subset $\mathcal{O} \subseteq \mathcal{I}$.*

Acyclic SCMs are simple.

Definition 6 (Twin SCM)**.** *Let $\mathcal{M}$ be a structural causal model. The* twin SCM *associated with $\mathcal{M}$ is defined as*

$$\mathcal{M}^{\text{twin}} := \langle \mathcal{I} \cup \mathcal{I}',\ \mathcal{J},\ \mathcal{X} \times \mathcal{X},\ \mathcal{E},\ \tilde{f},\ \mathbb{P}_{\mathcal{E}} \rangle,$$

where $\mathcal{I}' := \{i' : i \in \mathcal{I}\}$ is a disjoint copy of the endogenous index set, and $\tilde{f} : \mathcal{X} \times \mathcal{X} \times \mathcal{E} \to \mathcal{X} \times \mathcal{X}$ is the measurable function defined by

$$\tilde{f}(x, x', e) := \big(f(x, e),\ f(x', e)\big),$$

with $x, x' \in \mathcal{X}$ and $e \in \mathcal{E}$.

How the twin map $\tilde{f}$ is constructed For any noise realisation $e \in \mathcal{E}$ and stacked endogenous state $(x, x') = \big(x_1, \ldots, x_{|\mathcal{I}|},\ x'_1, \ldots, x'_{|\mathcal{I}|}\big) \in \mathcal{X} \times \mathcal{X}$, the twin-SCM mechanism $\tilde{f} : \mathcal{X} \times \mathcal{X} \times \mathcal{E} \longrightarrow \mathcal{X} \times \mathcal{X}$ is defined by $\tilde{f}(x, x', e) := \big(f(x, e),\ f(x', e)\big)$. Written coordinate-wise:

$$\tilde{f}_j(x, x', e) = \begin{cases} f_j(x, e), & j \in \mathcal{I}, \\ f_i(x', e), & j = i' \text{ for some } i \in \mathcal{I} (\text{i.e., } j \in \mathcal{I}'), \end{cases}$$

i.e. the first copy (un-primed) follows the original mechanism $X_j \leftarrow f_j(x, e)$, while the primed copy applies the same function f_i to its own state x'. In compact notation

$$\tilde{f}_j(x, x', e) = \begin{cases} f_j(x, e), & j \in \mathcal{I}, \\ f_i(x', e), & j' \in \mathcal{I}'. \end{cases}$$

Definition 7 (Counterfactual distribution)**.** *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$ be an SCM and let $\mathcal{M}^{\text{twin}}$ be its twin SCM. Consider a perfect intervention*

$$\text{do}\big(\tilde{\mathcal{I}},\ \xi_{\tilde{\mathcal{I}}}\big), \qquad \tilde{\mathcal{I}} \subseteq \mathcal{I} \cup \mathcal{I}', \quad \xi_{\tilde{\mathcal{I}}} \in \mathcal{X}_{\tilde{\mathcal{I}}},$$

applied to $\mathcal{M}^{\text{twin}}$, and denote the intervened model by $\big(\mathcal{M}^{\text{twin}}\big)_{\text{do}(\tilde{\mathcal{I}}, \xi_{\tilde{\mathcal{I}}})}$. If this intervened twin SCM admits a (measurable) solution $(X, X') \in \mathcal{X} \times \mathcal{X}$, then the joint distribution $\mathbb{P}_{(X,X')}$ is called the counterfactual distribution of $\mathcal{M}$ under the perfect intervention $\text{do}(\tilde{\mathcal{I}}, \xi_{\tilde{\mathcal{I}}})$ *associated with the pair of random variables (X, X').*

In the appendix, we delineate a clear correspondence between the twin-network formulation of structural causal models and Pearl's canonical action–abduction–prediction schema for counterfactual inference.

2.1 Shift–Scale intervention

For example, instead of forcing a treatment X to a set value, a shift intervention might increase each individual's natural treatment dose by a fixed amount δ, and a scale intervention might multiply each dose by a factor (e.g. 10% increase) – all while allowing X to remain influenced by its usual causes (confounders, prior variables, etc.). This preserves the causal edges into X but changes the conditional distribution or structural equation of X.

Definition 8 (Shift–Scale intervention)**.** *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$ be an SCM and fix a non-empty subset $\tilde{\mathcal{I}} \subseteq \mathcal{I}$. For each $j \in \tilde{\mathcal{I}}$ choose scale $a_j \in \mathbb{R}$ and shift $b_j \in \mathbb{R}$. The* shift–scale intervention

$$\mathrm{ss}\big(\tilde{\mathcal{I}}, a_{\tilde{\mathcal{I}}}, b_{\tilde{\mathcal{I}}}\big) \quad \big(a_{\tilde{\mathcal{I}}} := (a_j)_{j \in \tilde{\mathcal{I}}},\ b_{\tilde{\mathcal{I}}} := (b_j)_{j \in \tilde{\mathcal{I}}}\big)$$

produces a new SCM

$$\mathcal{M}_{\mathrm{ss}} := \big\langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f^{\mathrm{ss}}, \mathbb{P}_{\mathcal{E}} \big\rangle, \qquad f_i^{\mathrm{ss}}(x, e) := \begin{cases} a_i\, f_i(x, e) + b_i, & i \in \tilde{\mathcal{I}}, \\ f_i(x, e), & i \notin \tilde{\mathcal{I}}. \end{cases}$$

Perfect or do-intervention is recovered as the special case $a_j = 0,\ b_j = \xi_j$.

Definition 9 (Shift–Scale counterfactual distribution)**.** *Let $\mathcal{M}^{\mathrm{twin}}$ be the twin SCM of $\mathcal{M}$. Apply the shift–scale intervention only to the first copy:*

$$\big(\mathcal{M}^{\mathrm{twin}}\big)_{\mathrm{ss}} := \Big(\mathcal{M}^{\mathrm{twin}}\Big)_{\mathrm{ss}\big(\tilde{\mathcal{I}}, a_{\tilde{\mathcal{I}}}, b_{\tilde{\mathcal{I}}}\big)}.$$

If this intervened twin SCM admits a measurable solution $(X, X') \in \mathcal{X} \times \mathcal{X}$, we call the joint law $\mathbb{P}_{(X,X')}$ the shift–scale counterfactual distribution *of $\mathcal{M}$ under the intervention $\mathrm{ss}\big(\tilde{\mathcal{I}}, a_{\tilde{\mathcal{I}}}, b_{\tilde{\mathcal{I}}}\big)$.*

Given a subset a set of coordinates $\tilde{\mathcal{I}} \subseteq \mathcal{I} \cup \mathcal{I}'$, and parameters $a_{\tilde{\mathcal{I}}}, b_{\tilde{\mathcal{I}}} \in \mathbb{R}^{|\tilde{\mathcal{I}}|}$, we want to replace the structural equation

$$X_j \ = \ \tilde{f}_j(\,\cdot\,) \quad \longrightarrow \quad X_j \ = \ a_j\, \tilde{f}_j(\,\cdot\,) \ + \ b_j,$$

for each $j \in \tilde{\mathcal{I}}$. All other coordinates stay unchanged. Encoding these modifications into a single map

$$\tilde{g}(x, x', e) := \big(g_j(x, x', e)\big)_{j \in \mathcal{I} \cup \mathcal{I}'}, \qquad g_j(x, x', e) := \begin{cases} a_j\, \tilde{f}_j(x, x', e) + b_j, & j \in \tilde{\mathcal{I}}, \\ \tilde{f}_j(x, x', e), & j \notin \tilde{\mathcal{I}}, \end{cases}$$

gives the intervened twin update rule. Specifically

$$\begin{aligned} \text{if } j \in \mathcal{I} \quad &\Rightarrow \quad g_j(x, x', e) = a_j\, f_j(x, e) + b_j, \\ \text{if } j = i' \in \mathcal{I}' \quad &\Rightarrow \quad g_{i'}(x, x', e) = a_{i'}\, f_i(x', e) + b_{i'}. \end{aligned}$$

Thus each copy (un-primed and primed) is modified *only* on the requested coordinates, allowing independent interventions on the two worlds.

2.2 Semantics of Cyclic SCMs

One natural interpretation of cyclic structural equations is by assuming an underlying discrete-time dynamical system, where the equations act as update rules: the value of each variable at time $t+1$ is computed from the values at time t. The system is then analyzed in the limit as $t \to \infty$, focusing on the fixed points to which the dynamics converge. Mooij et al. (2013) demonstrate that an alternative, yet natural, interpretation of SCMs emerges when considering systems of ordinary differential equations (ODEs). By examining the equilibrium (steady-state) solutions of such ODEs, one arrives at a structural causal model that is time-independent, but still retains meaningful causal semantics with respect to interventions. Specifically, the semantics of interventions and counterfactuals (see also Appendix A.1) remain valid and well-defined in this steady-state context, as rigorously formalized by Mooij et al. (2013) and further extended by Bongers et al. (2021). These are by no means the only routes to structural causal models; indeed, SCMs may arise through a variety of alternative constructions and representations, depending on the nature of the system under consideration. Although many physical processes exhibit inertia, static cyclic structural causal models (SCMs) remain appropriate when we focus on equilibrium behavior or sample at a temporal resolution coarser than the fastest feedback loop. Examples include gene-regulatory networks in single-cell genomics (Rohbeck et al., 2024); market-equilibrium models (Bongers et al., 2021); predator–prey ecological systems; Thyroid or reproductive hormone axes exhibit feedback loops (Clarke et al., 2014), etc. Our mathematical framework is agnostic about the interpretation of cycles: we formulate everything directly at the level of structural equations with exogenous noise.

3 Theory

Theorem 1 (Global ℓ^p-contraction $\implies$ simple SCM)**.** *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$ be an SCM whose endogenous index set $\mathcal{I}$ is finite. Assume each coordinate domain ($\mathcal{X}_i \subseteq \mathbb{R}$) is non-empty and* closed*. Fix $p \in [1, \infty]$ and endow every product space with the ℓ^p-norm $\|x\|_p := (\sum_{i \in \mathcal{I}} |x_i|^p)^{1/p}$ (for $p = \infty$ take the maximum norm). Suppose there exists $\kappa \in [0, 1)$ such that*

$$\|f(x, e) - f(y, e)\|_p \;\leq\; \kappa \, \|x - y\|_p \quad \text{for all } x, y \in \mathcal{X},\ e \in \mathcal{E}. \tag{1}$$

Then $\mathcal{M}$ is uniquely solvable with respect to every *subset $\mathcal{O} \subseteq \mathcal{I}$, and hence $\mathcal{M}$ is a* simple *SCM.*

Proof. For any subset $\mathcal{O} \subseteq \mathcal{I}$ let $\mathcal{Q} := \mathrm{pa}(\mathcal{O}) \setminus \mathcal{O}$. Because every $\mathcal{X}_i$ is closed in $\mathbb{R}$ (hence complete) and $\mathcal{O}$ is finite, the product $\mathcal{X}_{\mathcal{O}} = \prod_{i \in \mathcal{O}} \mathcal{X}_i$ is complete under the ℓ^p-metric; the same holds for the full space $\mathcal{X}$.

For each pair $(x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})})$ define

$$h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}} : \mathcal{X}_{\mathcal{O}} \longrightarrow \mathcal{X}_{\mathcal{O}}, \qquad u \mapsto f_{\mathcal{O}}\big(u,\, x_{\mathcal{Q}},\, x_{\mathcal{I} \setminus (\mathcal{O} \cup \mathcal{Q})},\, e_{\mathrm{pa}(\mathcal{O})},\, e_{\mathcal{J} \setminus \mathrm{pa}(\mathcal{O})}\big),$$

where the "dummy" coordinates $x_{\mathcal{I} \setminus (\mathcal{O} \cup \mathcal{Q})}$ and $e_{\mathcal{J} \setminus \mathrm{pa}(\mathcal{O})}$ may be chosen arbitrarily because they do not influence $f_{\mathcal{O}}$.

For $u, v \in \mathcal{X}_{\mathcal{O}}$ define $\tilde{u} := (u, x_{\mathcal{Q}}, x_{\mathcal{I} \setminus (\mathcal{O} \cup \mathcal{Q})}),\ \tilde{v} := (v, x_{\mathcal{Q}}, x_{\mathcal{I} \setminus (\mathcal{O} \cup \mathcal{Q})}) \in \mathcal{X}$. Then $\|\tilde{u} - \tilde{v}\|_p = \|u - v\|_p$ because $\tilde{u}, \tilde{v}$ coincide outside $\mathcal{O}$. Using (1),

$$\|h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}}(u) - h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}}(v)\|_p = \|f_{\mathcal{O}}(\tilde{u}, e) - f_{\mathcal{O}}(\tilde{v}, e)\|_p \leq \|f(\tilde{u}, e) - f(\tilde{v}, e)\|_p \leq \kappa \|u - v\|_p.$$

Thus each map $h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}}$ is a κ-contraction on the *complete* metric space $(\mathcal{X}_{\mathcal{O}}, \|\cdot\|_p)$. By the Banach fixed-point theorem, for every $(x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})})$ there exists a *unique* element $u^*(x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}) \in \mathcal{X}_{\mathcal{O}}$ satisfying $u^* = h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}}(u^*)$.

Measurability. Fix any $\bar{u} \in \mathcal{X}_{\mathcal{O}}$. Defining Picard iterates $u^{(0)} \equiv \bar{u}$ and $u^{(n+1)} := h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}}(u^{(n)})$, one obtains a sequence of measurable functions converging *pointwise* to u^*. Limits of measurable functions are measurable, hence the map

$$g_{\mathcal{O}}(x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}) := u^*(x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})})$$

is measurable.

Equivalence. ($\Rightarrow$) By definition, $g_{\mathcal{O}}(x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})})$ is the unique fixed point u^* of the map

$$h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}}(u) \;:=\; f_{\mathcal{O}}(u, x_{\mathcal{Q}}, x_{\mathcal{I} \setminus (\mathcal{O} \cup \mathcal{Q})}, e_{\mathrm{pa}(\mathcal{O})}, e_{\mathcal{J} \setminus \mathrm{pa}(\mathcal{O})}).$$

Hence, if $x_{\mathcal{O}} = g_{\mathcal{O}}(x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})})$, then $x_{\mathcal{O}} = h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}}(x_{\mathcal{O}}) = f_{\mathcal{O}}(x, e)$.

($\Leftarrow$) Suppose $x_{\mathcal{O}} = f_{\mathcal{O}}(x, e)$. Then $x_{\mathcal{O}}$ satisfies $x_{\mathcal{O}} = f_{\mathcal{O}}(x, e) = h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}}(x_{\mathcal{O}})$, so it is a fixed point of $h_{x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}}$. By Banach's theorem, the fixed point is unique, so

$$x_{\mathcal{O}} = u^* = g_{\mathcal{O}}(x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}).$$

Therefore,

$$x_{\mathcal{O}} = g_{\mathcal{O}}(x_{\mathcal{Q}}, e_{\mathrm{pa}(\mathcal{O})}) \quad \Longleftrightarrow \quad x_{\mathcal{O}} = f_{\mathcal{O}}(x, e),$$

establishing the defining equivalence in the notion of unique solvability holds for all x and for $\mathbb{P}_{\mathcal{E}}$-almost every e.

Because $\mathcal{O} \subseteq \mathcal{I}$ was arbitrary, the same reasoning applies to every subset, establishing that $\mathcal{M}$ is uniquely solvable with respect to all subsets and hence is a simple SCM. □

The global contraction condition implies unique solvability for every subset; hence the models are simple SCMs in the sense of Bongers et al. (2021). This places us directly inside the setting where their closure results for do-interventions, marginalization and twin networks apply.

Theorem 2 (Unique Solvability of Twin SCM under Shift–Scale Interventions)**.** *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$ be an SCM such that the causal mechanism $f : \mathcal{X} \times \mathcal{E} \to \mathcal{X}$ satisfies a global ℓ^p-contraction for some constant $0 \leq \kappa < 1$. Let $\mathcal{M}^{\text{twin}}$ be the associated twin SCM, and fix a subset $\tilde{\mathcal{I}} \subseteq \mathcal{I} \cup \mathcal{I}'$, along with shift–scale coefficients $a_{\tilde{\mathcal{I}}} \in \mathbb{R}^{|\tilde{\mathcal{I}}|}$, $b_{\tilde{\mathcal{I}}} \in \mathbb{R}^{|\tilde{\mathcal{I}}|}$, satisfying:*

$$a_{\max} := \sup_{j \in \tilde{\mathcal{I}}} |a_j| \leq 1.$$

Then the shift–scale intervened twin SCM $\left(\mathcal{M}^{\text{twin}}\right)_{\text{ss}(\tilde{\mathcal{I}}, a_{\tilde{\mathcal{I}}}, b_{\tilde{\mathcal{I}}})}$ is uniquely solvable with respect to every subset of endogenous variables, and is thus a simple SCM. In particular, the associated counterfactual distribution $\mathbb{P}_{(X,X')}$ is well-defined.

Proof. Let $(x, x') \in \mathcal{X} \times \mathcal{X}$ be the endogenous variables of the twin SCM and the twin map $\tilde{f}(x, x', e) := (f(x, e), f(x', e))$. is a global κ-contraction on $\mathcal{X} \times \mathcal{X}$:

$$\|\tilde{f}(x, x', e) - \tilde{f}(y, y', e)\|_p = \left(\|f(x, e) - f(y, e)\|_p^p + \|f(x', e) - f(y', e)\|_p^p\right)^{1/p} \leq \kappa \|(x, x') - (y, y')\|_p.$$

Now define the shift–scale intervened map $\tilde{g}(x, x', e)$ by:

$$\tilde{g}_j(x, x', e) := \begin{cases} a_j \tilde{f}_j(x, x', e) + b_j, & j \in \tilde{\mathcal{I}}, \\ \tilde{f}_j(x, x', e), & j \notin \tilde{\mathcal{I}}. \end{cases}$$

We now show that $\tilde{g}(\cdot, \cdot, e)$ is a global contraction in the ℓ^p-norm with the same constant κ. Let $u := (x, x')$, $v := (y, y')$. Then for any $j \in \mathcal{I} \cup \mathcal{I}'$,

$$|\tilde{g}_j(u, e) - \tilde{g}_j(v, e)| = \begin{cases} |a_j| \cdot |\tilde{f}_j(u, e) - \tilde{f}_j(v, e)| \leq a_{\max} |\tilde{f}_j(u, e) - \tilde{f}_j(v, e)|, & j \in \tilde{\mathcal{I}}, \\ |\tilde{f}_j(u, e) - \tilde{f}_j(v, e)|, & j \notin \tilde{\mathcal{I}}. \end{cases}$$

Therefore,

$$\|\tilde{g}(u, e) - \tilde{g}(v, e)\|_p \leq \left(\sum_{j \in \tilde{\mathcal{I}}} (a_{\max} |\tilde{f}_j(u, e) - \tilde{f}_j(v, e)|)^p + \sum_{j \notin \tilde{\mathcal{I}}} |\tilde{f}_j(u, e) - \tilde{f}_j(v, e)|^p \right)^{1/p}.$$

Since $a_{\max} \leq 1$, we have:

$$\|\tilde{g}(u, e) - \tilde{g}(v, e)\|_p \leq \|\tilde{f}(u, e) - \tilde{f}(v, e)\|_p \leq \kappa \|u - v\|_p.$$

Thus $\tilde{g}(\cdot, e)$ is also a global κ-contraction on $(\mathcal{X} \times \mathcal{X}, \|\cdot\|_p)$, a complete metric space since $\mathcal{X}$ is closed and finite-dimensional.

By Banach's fixed-point theorem (as used in Theorem 1), the shift–scale intervened twin SCM has a unique solution (X, X') for each e, and the solution map is measurable.

Since the SCM is uniquely solvable with respect to all subsets of $\mathcal{I} \cup \mathcal{I}'$, the intervened twin SCM is simple. Pushing forward $\mathbb{P}_{\mathcal{E}}$ through this solution map defines the shift–scale counterfactual distribution $\mathbb{P}_{(X,X')}$. □

Proposition 8.2 in Bongers et al. (2021) shows that the class of simple SCMs is closed under (i) marginalization, (ii) perfect interventions (do), and (iii) the twin construction. It does not give a sufficient condition for simplicity; it presupposes simplicity. Theorem 2 gives sufficient analytic conditions for unique solvability of the twin under a class of soft, parametric shift–scale interventions, thereby ensuring well-posed counterfactuals beyond hard do.

Proposition 1 (Closure under (composed) Shift–Scale Interventions)**.** *Let* $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$ *be an SCM such that for some* $p \in [1, \infty]$

$$\|f(x,e) - f(y,e)\|_p \ \leq \ \kappa \, \|x - y\|_p, \qquad \forall\, x, y \in \mathcal{X},\ \forall\, e \in \mathcal{E},$$

with a constant $\kappa \in [0, 1)$*. Consider a* finite sequence *of shift–scale interventions applied successively to* $\mathcal{M}$*:*

$$\mathrm{ss}\big(\tilde{\mathcal{I}}^{(1)}, a^{(1)}, b^{(1)}\big) \ \circ \ \ldots \ \circ \ \mathrm{ss}\big(\tilde{\mathcal{I}}^{(m)}, a^{(m)}, b^{(m)}\big), \qquad m \geq 1,$$

where for every $r = 1, \ldots, m$ *and every* $j \in \tilde{\mathcal{I}}^{(r)}$ *one has* $|a_j^{(r)}| \leq 1$*.*
Then the resulting (composed) intervened model is equivalent to a single shift–scale intervention $\mathrm{ss}\big(\tilde{\mathcal{I}}, a^{\mathrm{comp}}, b^{\mathrm{comp}}\big)$ *with* $|a_j^{\mathrm{comp}}| \leq 1$ *for all* $j \in \tilde{\mathcal{I}}$*, and its structural function* $g(x,e) = a^{\mathrm{comp}} \odot f(x,e) + b^{\mathrm{comp}}$ *satisfies*

$$\|g(x,e) - g(y,e)\|_p \ \leq \ \kappa \, \|x - y\|_p, \qquad \forall\, x, y \in \mathcal{X},\ e \in \mathcal{E}.$$

Consequently the intervened SCM remains κ*-contractive and hence is a* simple *SCM.*

Proof. **(1) Composition reduces to a single affine map.** For any coordinate $j \in \mathcal{I}$ the successive updates act as $x_j \ \mapsto \ a_j^{(m)}\big(a_j^{(m-1)}(\ldots a_j^{(1)} f_j(x,e) + b_j^{(1)} \ldots) + b_j^{(m-1)}\big) + b_j^{(m)}$, which can be written $a_j^{\mathrm{comp}} f_j(x,e) + b_j^{\mathrm{comp}}$ with $a_j^{\mathrm{comp}} := \prod_{r:\, j \in \tilde{\mathcal{I}}^{(r)}} a_j^{(r)}$ and a corresponding affine drift b_j^{comp}. Since each factor satisfies $|a_j^{(r)}| \leq 1$, we have $|a_j^{\mathrm{comp}}| \leq 1$. Hence the final system coincides with $g(x,e) = a^{\mathrm{comp}} \odot f(x,e) + b^{\mathrm{comp}}$.

(2) Contraction constant is preserved. Let $D := \mathrm{diag}(a^{\mathrm{comp}})$ denote the diagonal matrix of multiplicative factors. For any $u \in \mathbb{R}^{|\mathcal{I}|}$, $\|D\,u\|_p \leq \|u\|_p$ because every diagonal entry satisfies $|a_j^{\mathrm{comp}}| \leq 1$. Hence for all x, y and any e

$$\|g(x,e) - g(y,e)\|_p = \|D\big(f(x,e) - f(y,e)\big)\|_p \leq \|f(x,e) - f(y,e)\|_p \leq \kappa \|x - y\|_p.$$

Thus g inherits the same global contraction constant κ.

(3) Simplicity after intervention. Because the intervened mechanism is still κ-contractive with $\kappa < 1$, Theorem 1 applies verbatim: the intervened model is uniquely solvable with respect to every subset of endogenous variables, i.e. it is simple. $\square$

Remark 1 (Scale factors exceeding one)**.** *The proof of Proposition 1 used the bound* $|a_j| \leq 1$ *to conclude that the diagonal scaling* $D = \mathrm{diag}(a_j)$ *obeys* $\|Du\|_p \leq \|u\|_p$ *and therefore does not enlarge the global Lipschitz constant. If some multipliers satisfy* $|a_j| > 1$*, simplicity can still hold, but one must verify an additional criterion:*

Let $\kappa < 1$ *be the original contraction constant of* f *and define*

$$\kappa_{\max} \ := \ \big(\max_{j \in \tilde{\mathcal{I}}} |a_j|\big)\, \kappa.$$

Because $\|D\|_p = \max_j |a_j|$*, the intervened mechanism is globally* $\kappa_{\max}$*-Lipschitz:* $\|Df(x,e) - Df(y,e)\|_p \leq \kappa_{\max} \|x - y\|_p$*. If* $\kappa_{\max} < 1$ *the model remains a contraction and is therefore simple; if* $\kappa_{\max} \geq 1$ *the contraction proof no longer ensures uniqueness and additional analysis is required.*

Proposition 2 (Sub-Gaussian Tails under Gaussian Noise for Shift–Scale Counterfactuals)**.** *Let the SCM* $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$ *satisfy the global* ℓ^2*-contraction condition of Theorem 1 with constant* $\kappa < 1$*. Apply any shift–scale intervention* $\mathrm{ss}(\tilde{\mathcal{I}}, a_{\tilde{\mathcal{I}}}, b_{\tilde{\mathcal{I}}})$ *to the twin SCM, with* $\max_{j \in \tilde{\mathcal{I}}} |a_j| \le 1$*, and let* $(\mathbf{X}, \mathbf{X}')$ *be the unique endogenous solution (Definition 2) of the intervened twin model. Assume*

(a) 1***-Lipschitz in noise (in*** ℓ^2***):*** *For all* $x \in \mathcal{X}$, $e_1, e_2 \in \mathcal{E}$*,*

$$\|f(x, e_1) - f(x, e_2)\|_2 \le \|e_1 - e_2\|_2.$$

(b) ***Gaussian noise:*** $\mathbf{E} \sim \mathcal{N}(\mu, \Sigma)$ *with* $\Sigma \preceq \sigma^2 I$ *for some* $\sigma^2 > 0$*.*

Then for every 1*-Lipschitz functional* $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *(w.r.t.* ℓ^2*),*

$$\mathbb{P}\Big(h(\mathbf{X}, \mathbf{X}') - \mathbb{E}h(\mathbf{X}, \mathbf{X}') \ge t\Big) \le \exp\Big(-\tfrac{t^2}{4(1-\kappa)^{-2}\sigma^2}\Big), \qquad t > 0.$$

Hence $(\mathbf{X}, \mathbf{X}')$ *is sub-Gaussian with proxy* $2(1-\kappa)^{-2}\sigma^2$*.*

Proof. Let $\Phi : \mathcal{E} \to \mathcal{X} \times \mathcal{X}$ denote the unique measurable solution map of the shift–scale–intervened twin SCM. That is,

$$(\mathbf{X}, \mathbf{X}') = \Phi(\mathbf{E}),$$

where $\mathbf{E} \sim \mathbb{P}_{\mathcal{E}}$ is the exogenous random variable. By the global κ-contraction in x and the 1-Lipschitz property in e (both in ℓ^2), Φ is L-Lipschitz with constant $L := \frac{\sqrt{2}}{1-\kappa}$ in ℓ^2:

$$\|\Phi(e_1) - \Phi(e_2)\|_2 \le \frac{\sqrt{2}}{1-\kappa}\|e_1 - e_2\|_2 \quad \text{for all } e_1, e_2 \in \mathcal{E}.$$

Take any $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ that is 1-Lipschitz in ℓ^2. Write $\mathbf{E} = \mu + \Sigma^{1/2}\mathbf{Z}$ with $\mathbf{Z} \sim \mathcal{N}(0, I)$.

Then $g(z) := h\big(\Phi(\mu + \Sigma^{1/2}z)\big)$ is $L\|\Sigma^{1/2}\|_2$-Lipschitz in ℓ^2. Since $\Sigma \preceq \sigma^2 I$, we have $\|\Sigma^{1/2}\|_2 \le \sigma$. By Gaussian concentration for Lipschitz functions (e.g., (Vershynin, 2018, Thm. 5.2.3)),

$$\mathbb{P}\Big(h(\mathbf{X}, \mathbf{X}') - \mathbb{E}[h(\mathbf{X}, \mathbf{X}')] \ge t\Big) \le \exp\left(-\frac{t^2}{2L^2\sigma^2}\right) = \exp\left(-\frac{t^2}{4(1-\kappa)^{-2}\sigma^2}\right).$$

This proves the stated concentration inequality. □

Remark 2 (ℓ^p version of Proposition 2)**.** *Let all Lipschitz and contraction conditions in Proposition 2 be measured in* ℓ^p $(1 \le p \le \infty)$*, and keep the Gaussian noise assumption* $\mathbf{E} \sim \mathcal{N}(\mu, \Sigma)$*. Then for every* 1*-Lipschitz* $h : (\mathcal{X} \times \mathcal{X}, \|\cdot\|_p) \to \mathbb{R}$ *and all* $t > 0$*,*

$$\mathbb{P}\Big(h(\mathbf{X}, \mathbf{X}') - \mathbb{E}h(\mathbf{X}, \mathbf{X}') \ge t\Big) \;\le\; \exp\Big(-\frac{t^2}{2^{1+\frac{2}{p}}\,(1-\kappa)^{-2}\,\|\Sigma^{1/2}\|_{2\to p}^2}\Big),$$

where $\|A\|_{2\to p} := \sup_{x \ne 0} \frac{\|Ax\|_p}{\|x\|_2}$*. In particular, if* $\Sigma \preceq \sigma^2 I$ *and* $d := \dim(\mathcal{E})$*, then*

$$\|\Sigma^{1/2}\|_{2\to p} \le \sigma\,\|I\|_{2\to p} \quad \textit{with} \quad \|I\|_{2\to p} = \begin{cases} 1, & p \ge 2, \\ d^{\frac{1}{p}-\frac{1}{2}}, & 1 \le p < 2, \end{cases}$$

so the proxy equals $2^{\frac{2}{p}}(1-\kappa)^{-2}\sigma^2$ *for* $p \ge 2$*, and* $2^{\frac{2}{p}}(1-\kappa)^{-2}\sigma^2\, d^{2(\frac{1}{p}-\frac{1}{2})}$ *for* $1 \le p < 2$*.* Proof sketch. *Write* $\mathbf{E} = \mu + \Sigma^{1/2}\mathbf{Z}$*,* $\mathbf{Z} \sim \mathcal{N}(0, I)$*; then* $z \mapsto h(\Phi(\mu + \Sigma^{1/2}z))$ *is* $2^{\frac{1}{p}}(1-\kappa)^{-1}\|\Sigma^{1/2}\|_{2\to p}$*-Lipschitz in* ℓ^2*, and Gaussian Lipschitz concentration applies.*

4 Example

Consider a linear cyclic SCM with mutual dependence between consumption (C) and income (I):

$$\begin{aligned} C &= 0.50\, I + 1 + E_C, \\ I &= 0.40\, C + 0.50 + E_I, \end{aligned} \qquad (E_C, E_I)^\top \sim \mathcal{N}(\mathbf{0},\, 0.04\,\mathbf{I}_2). \tag{2}$$

Written in matrix form $X = (C, I)^\top$, $X = AX + b + E$ with

$$A = \begin{pmatrix} 0 & 0.50 \\ 0.40 & 0 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}, \qquad E = (E_C, E_I)^\top.$$

Because the spectral norm is $\|A\|_2 \le \|A\|_F = \sqrt{0.25 + 0.16} = 0.6403 < 1$, the model is *globally contractive* and therefore *simple* in the sense of Definition 5.

For any 2×2 matrix $A = \left[\begin{smallmatrix} 0 & a \\ b & 0 \end{smallmatrix}\right]$ with $ab < 1$, $(I - A)^{-1} = \frac{1}{1-ab}\left[\begin{smallmatrix} 1 & a \\ b & 1 \end{smallmatrix}\right]$. With $(a, b) = (0.50, 0.40)$ we have $(I - A)^{-1} = \left[\begin{smallmatrix} 1.25 & 0.625 \\ 0.50 & 1.25 \end{smallmatrix}\right]$ and therefore

$$\mathbb{E}[X] = (I - A)^{-1} b = (1.5625,\ 1.125)^\top, \tag{3}$$

$$\mathrm{Cov}(X) = (I - A)^{-1}(0.04\,\mathbf{I}_2)(I - A)^{-\top} = \begin{pmatrix} 0.0781 & 0.0563 \\ 0.0563 & 0.0725 \end{pmatrix}. \tag{4}$$

Thus $X_{\text{obs}} \sim \mathcal{N}\big(\mu, \Sigma\big)$ with μ and Σ given in (3)–(4). The correlation between C and I is $\rho = 0.75$.

Shift–Scale intervention on I. Suppose a fiscal policy reform dampens the effect of both consumption and random shocks on income by a factor $\alpha = 0.8$, and provides a fixed income supplement of $\beta = 1.0$ units:

$$\mathrm{ss}(I, \alpha, \beta) \ :\ I \ \leftarrow\ \alpha\big(0.40\, C + 0.50 + E_I\big) + \beta, \quad \alpha = 0.8,\ \beta = 1.0.$$

The intervened structural parameters are

$$A' = \begin{pmatrix} 0 & 0.50 \\ 0.32 & 0 \end{pmatrix}, \qquad b' = \begin{pmatrix} 1 \\ 1.4 \end{pmatrix}, \quad \text{so} \quad \|A'\|_2 \le \|A'\|_F = \sqrt{0.25 + 0.1024} = 0.5936 < 1.$$

Scaling also affects the exogenous term: $E' = (E_C,\, \alpha E_I)^\top$, $\Sigma_{E'} = \mathrm{diag}(0.04, \alpha^2 \cdot 0.04) = \mathrm{diag}(0.04, 0.0256)$. Contractivity is preserved, hence a unique *interventional* equilibrium exists:

$$\mathbb{E}[X_{\text{int}}] = (I - A')^{-1} b' = (2.024,\ 2.048)^\top, \tag{5}$$

$$\mathrm{Cov}(X_{\text{int}}) = (I - A')^{-1}\Sigma_{E'}(I - A')^{-\top} = \begin{pmatrix} 0.0658 & 0.0363 \\ 0.0363 & 0.0421 \end{pmatrix}, \qquad \rho_{\text{int}} \approx 0.69. \tag{6}$$

Consumption rises by $\sim 29\,\%$ and income by $\sim 82\,\%$, while the C–I correlation falls from 0.75 to 0.69.

From intervention to *counterfactual* via the twin SCM. Equations (5)–(6) describe the *interventional* distribution $P(X_{\text{int}})$, i.e. what we would observe *if* the shift–scale policy were enacted *for the whole population*. To answer individual–level *counterfactual* queries ("what would *this* household's consumption be had the policy applied?") we follow the *twin network* construction. Let (c, i) denote the *actually observed* consumption and income for one household. The twin-SCM duplicates every endogenous variable and shares the same exogenous noise:

$$\boxed{\begin{aligned} C &= 0.50\, I + 1 + E_C, & C' &= 0.50\, I' + 1 + E_C, \\ I &= 0.40\, C + 0.50 + E_I, & I' &= 0.8\big(0.40\, C' + 0.50 + E_I\big) + 1. \end{aligned}} \tag{Twin SCM}$$

where the primed copy encodes the shift–scale intervention $\mathrm{ss}(I, \alpha{=}0.8, \beta{=}1.0)$ and the unprimed copy remains factual. From the factual equations, we can solve

$$E_C = c - 0.50\, i - 1, \qquad E_I = i - 0.40\, c - 0.50. \tag{7}$$

Eliminating C' from the primed equations of the twin SCM yields $0.84\, I' = 1.72 + 0.8\, E_I + 0.32\, E_C$. Substituting 7 gives the counterfactual income $I'(c, i) = \frac{25}{21} + \frac{16}{21} i = 1.190476 + 0.761905\, i$. Back-substitution furnishes the *counterfactual consumption* $C'(c, i) = c + \frac{25}{42} - \frac{5}{42} i = c + 0.595238 - 0.119048\, i$. Above equations give the *counterfactual response mapping* $(c, i) \mapsto (C', I')$. Because the mapping is affine and the factual distribution is Gaussian, the *marginal counterfactual* (C', I') is again Gaussian with $\mathbb{E}[(C', I')^\top] = (2.024, 2.048)^\top$ and covariance exactly matching (6). Here, the twin-SCM reconciles individual counterfactual semantics $(C'(c, i), I'(c, i))$ with the *population-level* interventional distribution already reported,while remaining in closed form due to the linearity and contractiveness of the model.

5 Limitations

The κ–contraction condition must hold *uniformly* over the entire state space; many realistic feedback systems may violate this in certain regimes, even though they still admit unique equilibria. Our concentration result further relies on Gaussianity of the exogenous variables. For heavy-tailed or merely bounded-moment noise, one typically obtains polynomial rather than exponential concentration, which we do not analyze here. We also invoke Banach's theorem on closed (and hence complete) coordinate domains; models whose natural domains are open or lie on manifolds require additional care. From a practical standpoint, certifying global Lipschitz constants of black-box simulators is challenging; data-driven or local contraction diagnostics may be more feasible in applications. Finally, our closure results explicitly cover only shift–scale maps with bounded gains ($|a_j| \leq 1$); interventions with larger multiplicative factors, stochastic policies, or more general functional forms are not yet addressed.

At present, our analysis is restricted to *shift–scale interventions*. This class, however, already strictly generalizes hard (do-) interventions and provides a principled foundation for broader extensions. The key theoretical principle underpinning our guarantees is the preservation of *global contractivity*—that is, the global Lipschitz constant of the intervened system remains below one. This property ensures unique solvability of the intervened SCM, even in the presence of cycles, and is abstract enough to accommodate richer intervention types, provided they do not destroy contractivity. Consequently, our results extend directly to any intervention family (including nonlinear, stochastic, or more complex parametric changes) that preserves the contraction property. An explicit analysis of broader classes of interventions, and sufficient conditions under which they preserve contractivity, is a promising direction for future work

6 Conclusion and Future Work

We have established a principled foundation for counterfactual inference in cyclic structural causal models under shift–scale interventions. Leveraging a global contraction assumption, we proved that such models are simple, ensuring unique solvability even in the presence of feedback. We further showed that shift–scale interventions preserve solvability, are closed under composition, and admit sub-Gaussian tail bounds for counterfactual functionals under natural regularity assumptions. These results demonstrate that contraction-based SCMs offer a mathematically tractable yet expressive class for reasoning about interventions and counterfactuals in cyclic settings. Future research may focus on developing deep generative models for cyclic counterfactuals, leveraging the theoretical foundations established here.

Acknowledgments and Disclosure of Funding

This research was funded in part by the Indo-French Centre for the Promotion of Advanced Research (IFCPAR/CEFIPRA) through project number CSRP 6702-2.

References

Bongers, S., Forré, P., Peters, J., and Mooij, J. M. (2021). Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5):2885 – 2915.

Clarke, B., Leuridan, B., and Williamson, J. (2014). Modelling mechanisms with causal cycles. *Synthese*, 191(8):1–31.

Kladny, K.-R., von Kügelgen, J., Schölkopf, B., and Muehlebach, M. (2024). Deep backtracking counterfactuals for causally compliant explanations. *Transactions on Machine Learning Research*.

Komanduri, A., Zhao, C., Chen, F., and Wu, X. (2024). Causal diffusion autoencoders: Toward counterfactual generation via diffusion probabilistic models. In *Proceedings of the 27th European Conference on Artificial Intelligence*.

Kügelgen, J. V., Mohamed, A., and Beckers, S. (2023). Backtracking counterfactuals. In van der Schaar, M., Zhang, C., and Janzing, D., editors, *Proceedings of the Second Conference on Causal*

Learning and Reasoning, volume 213 of *Proceedings of Machine Learning Research*, pages 177–196. PMLR.

Lorch, L., Krause, A., and Schölkopf, B. (2024). Causal modeling with stationary diffusions. In *International Conference on Artificial Intelligence and Statistics*, pages 1927–1935. PMLR.

Massidda, R., Geiger, A., Icard, T., and Bacciu, D. (2023). Causal abstraction with soft interventions. In van der Schaar, M., Zhang, C., and Janzing, D., editors, *Proceedings of the Second Conference on Causal Learning and Reasoning*, volume 213 of *Proceedings of Machine Learning Research*, pages 68–87. PMLR.

Melistas, T., Spyrou, N., Gkouti, N., Sanchez, P., Vlontzos, A., Panagakis, Y., Papanastasiou, G., and Tsaftaris, S. A. (2024). Benchmarking counterfactual image generation. In Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., and Zhang, C., editors, *Advances in Neural Information Processing Systems*, volume 37, pages 133207–133230. Curran Associates, Inc.

Mooij, J. M., Janzing, D., and Schölkopf, B. (2013). From ordinary differential equations to structural causal models: the deterministic case. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, UAI'13, page 440–448, Arlington, Virginia, USA. AUAI Press.

Pawlowski, N., de Castro, D. C., and Glocker, B. (2020). Deep structural causal models for tractable counterfactual inference. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*.

Pearl, J., Glymour, M., and Jewell, N. (2016). *Causal Inference in Statistics: A Primer*. Wiley.

Rohbeck, Martin Clarke, B., Mikulik, K., Pettet, A., Stegle, O., and Ueltzhöffer, K. (2024). Bicycle: Intervention-based causal discovery with cycles. In *Conference on Causal Learning and Reasoning*. PMLR.

Rothenhäusler, D., Heinze, C., Peters, J., and Meinshausen, N. (2015). Backshift: Learning causal cyclic graphs from unknown shift interventions. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Saha, S. and Garain, U. (2022). On noise abduction for answering counterfactual queries: A practical outlook. *Transactions on Machine Learning Research*.

Sanchez, P. and Tsaftaris, S. A. (2022). Diffusion causal models for counterfactual estimation. In *First Conference on Causal Learning and Reasoning*.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wu, Y., McConnell, L., and Iriondo, C. (2025). Counterfactual generative modeling with variational causal inference. In *The Thirteenth International Conference on Learning Representations*.

Zhang, J., Squires, C., and Uhler, C. (2021). Matching a desired causal state via shift interventions. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA. Curran Associates Inc.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We study counterfactuals in cyclic causal graph under shift scale intervention.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Added limitations section. Please refer to section 5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Clearly stated.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification: The research work is theoretical in nature.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: The research work is theoretical in nature.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The research work is theoretical in nature.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The research work is theoretical in nature.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The research work is theoretical in nature.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: Our research fully complies with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: We may have such impacts but it is not very immediate.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The research work is theoretical in nature.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The research work is theoretical in nature.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The research work is theoretical in nature.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The research work is theoretical in nature.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The research work is theoretical in nature.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLMs were used only for polishing the work, not for core research.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.

A Supplementary Material

A.1 Equivalence of Twin SCM and Pearl's Abduction-Action-Prediction (AAP) Procedure

Theorem 3. *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$ be a structural causal model that is* simple *(i.e., uniquely solvable with respect to every subset of endogenous variables). Assume:*

(i) the exogenous vector $\mathbf{E} \sim \mathbb{P}_{\mathcal{E}}$ has a joint density and mutually independent components;

(ii) the counterfactual query is defined by a perfect intervention $\mathrm{do}(X_{\mathcal{A}} := \tilde{x}_{\mathcal{A}})$ for some $\mathcal{A} \subseteq \mathcal{I}$.

Let $\mathbf{X} \sim \mathcal{M}$ denote the factual world, and define the counterfactual outcome via:

- ***Twin SCM definition:*** *Solve the intervened twin SCM where the do-intervention is applied to the primed copy, yielding random variables $(\mathbf{X}, \mathbf{X}')$.*
- ***Pearl's AAP (Pearl, 2009) procedure:*** *Sample $\mathbf{E} \sim \mathbb{P}_{\mathcal{E}}$ conditioned on $\mathbf{X} = x^{\mathrm{obs}}$, and solve the intervened SCM with this fixed noise to obtain $\mathbf{X}^{cf}$.*

Then, the counterfactual distribution derived from the ***twin SCM*** *coincides with the distribution obtained from* ***Pearl's abduction–action–prediction (AAP)*** *procedure:*

$$\mathbb{P}_{\mathbf{X}'|\mathbf{X}=x^{\mathrm{obs}}} = \mathbb{P}_{\mathbf{X}^{\mathrm{cf}}|\mathbf{X}=x^{\mathrm{obs}}}.$$

Proof. Since $\mathcal{M}$ is simple, by definition there exists a unique measurable function $g : \mathcal{E} \to \mathcal{X}$ such that for every $e \in \mathcal{E}$,

$$g(e) = f(g(e), e), \quad \text{so that } \mathbf{X} = g(\mathbf{E}) \text{ a.s.}$$

Similarly, after the intervention $\mathrm{do}(X_{\mathcal{A}} := \tilde{x}_{\mathcal{A}})$, the intervened SCM admits a unique measurable solution map $g^{\mathrm{do}} : \mathcal{E} \to \mathcal{X}$ (Proposition 3.8, Bongers et al. (2021)).

In the twin SCM construction, both copies (factual and counterfactual) share the same exogenous vector $\mathbf{E}$, and the intervention is applied only to the first copy. Solving the twin SCM yields:

$$(\mathbf{X}, \mathbf{X}') = (g(\mathbf{E}),\ g^{\mathrm{do}}(\mathbf{E})).$$

Let $x^{\mathrm{obs}} \in \mathcal{X}$ be the observed factual outcome. Pearl's procedure involves:

- *Abduction:* condition on the observation $\mathbf{X} = x^{\mathrm{obs}}$, which corresponds to conditioning on the set $\{e \in \mathcal{E} : g(e) = x^{\mathrm{obs}}\}$;
- *Action:* apply the intervention to obtain the new function g^{do};
- *Prediction:* evaluate $\mathbf{X}^{\mathrm{cf}} = g^{\mathrm{do}}(\mathbf{E})$ using the same noise, but now drawn from the posterior $\mathbb{P}_{\mathcal{E}|g(\mathbf{E})=x^{\mathrm{obs}}}$.

Define the posterior distribution on exogenous noise:

$$\mu_{x^{\mathrm{obs}}}(A) := \mathbb{P}\big(\mathbf{E} \in A \,\big|\, g(\mathbf{E}) = x^{\mathrm{obs}}\big), \qquad \text{for all measurable } A \subseteq \mathcal{E}.$$

This measure is supported on the set

$$\{e \in \mathcal{E} : g(e) = x^{\mathrm{obs}}\}.$$

Equality of counterfactual laws. By construction, $\mathbf{X}' = g^{\mathrm{do}}(\mathbf{E})$, so the counterfactual distribution from the twin SCM (given $\mathbf{X} = x^{\mathrm{obs}}$) is:

$$\mathbb{P}_{\mathbf{X}'|\mathbf{X}=x^{\mathrm{obs}}}(B) = \mu_{x^{\mathrm{obs}}}\big(e : g^{\mathrm{do}}(e) \in B\big), \qquad B \subseteq \mathcal{X}.$$

Pearl's method also conditions on $\{\mathbf{X} = x^{\mathrm{obs}}\}$, inducing the same posterior $\mu_{x^{\mathrm{obs}}}$ over exogenous variables. Then, the counterfactual outcome is $\mathbf{X}^{\mathrm{cf}} = g^{\mathrm{do}}(\mathbf{E})$ with $\mathbf{E} \sim \mu_{x^{\mathrm{obs}}}$. Thus,

$$\mathbb{P}_{\mathbf{X}^{\mathrm{cf}}|\mathbf{X}=x^{\mathrm{obs}}}(B) = \mu_{x^{\mathrm{obs}}}\big(e : g^{\mathrm{do}}(e) \in B\big),$$

which is the same expression as for the twin SCM. □

The counterfactual law derived from the twin SCM coincides with the one obtained from Pearl's abduction–action–prediction procedure under unique solvability.

Remark 3 (Notation $\mathbf{X} \sim \mathcal{M}$)**.** *The shorthand* $\mathbf{X} \sim \mathcal{M}$ *means that the random vector* $\mathbf{X}$ *is a* solution *of the SCM* $\mathcal{M}$.

References

Bongers, S., Forré, P., Peters, J., and Mooij, J. M. (2021). Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5):2885 – 2915.

Pearl, J. (2009). *Causality*. Cambridge University Press, 2 edition.