Standardizing Structural Causal Models

ICLR 2025

Joint work

Weronika Ormaniec

Scott Sussex

Lars Lorch

Bernhard Schölkopf

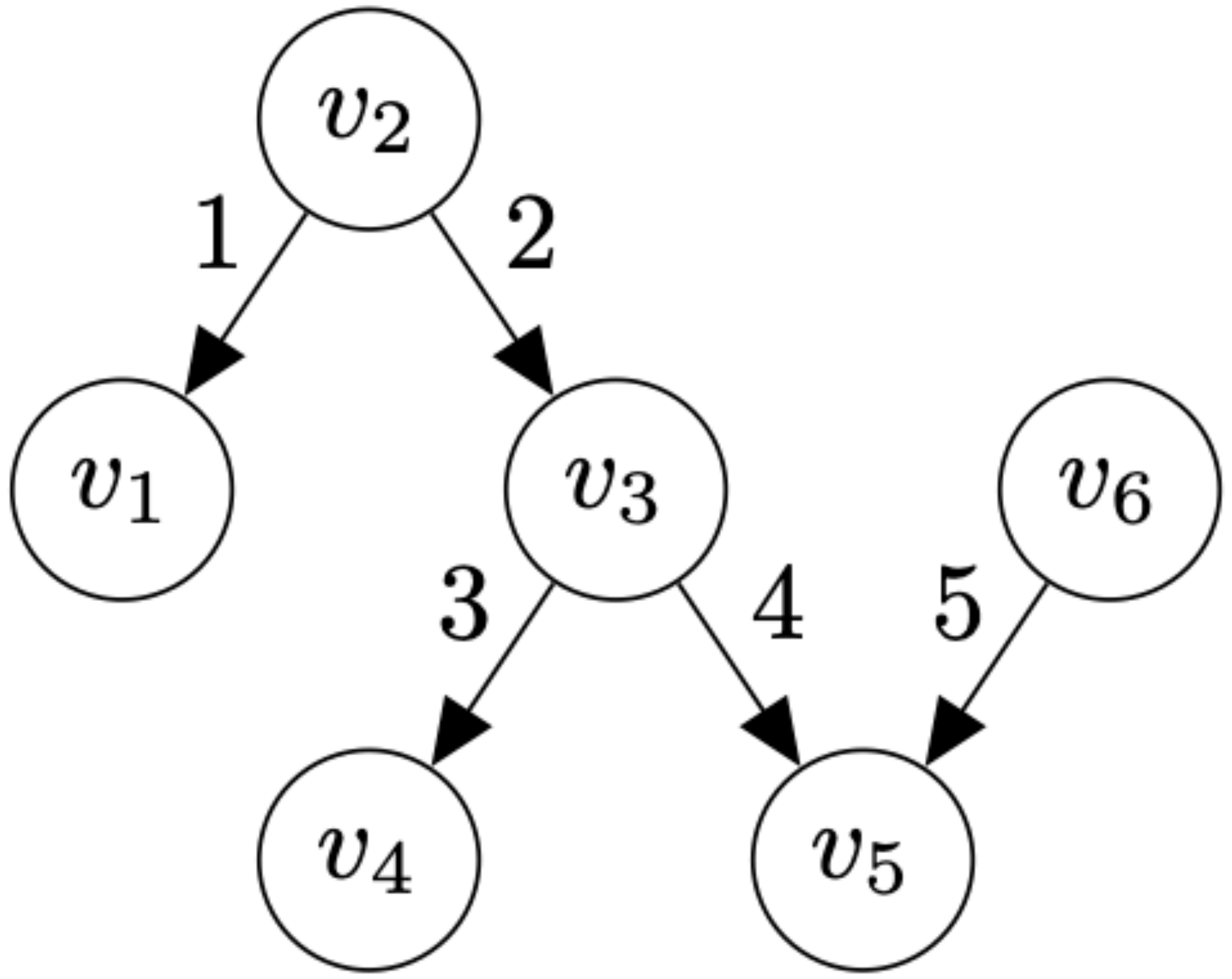
Andreas Krause

ETH zürich

MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

Introduction

SCMs

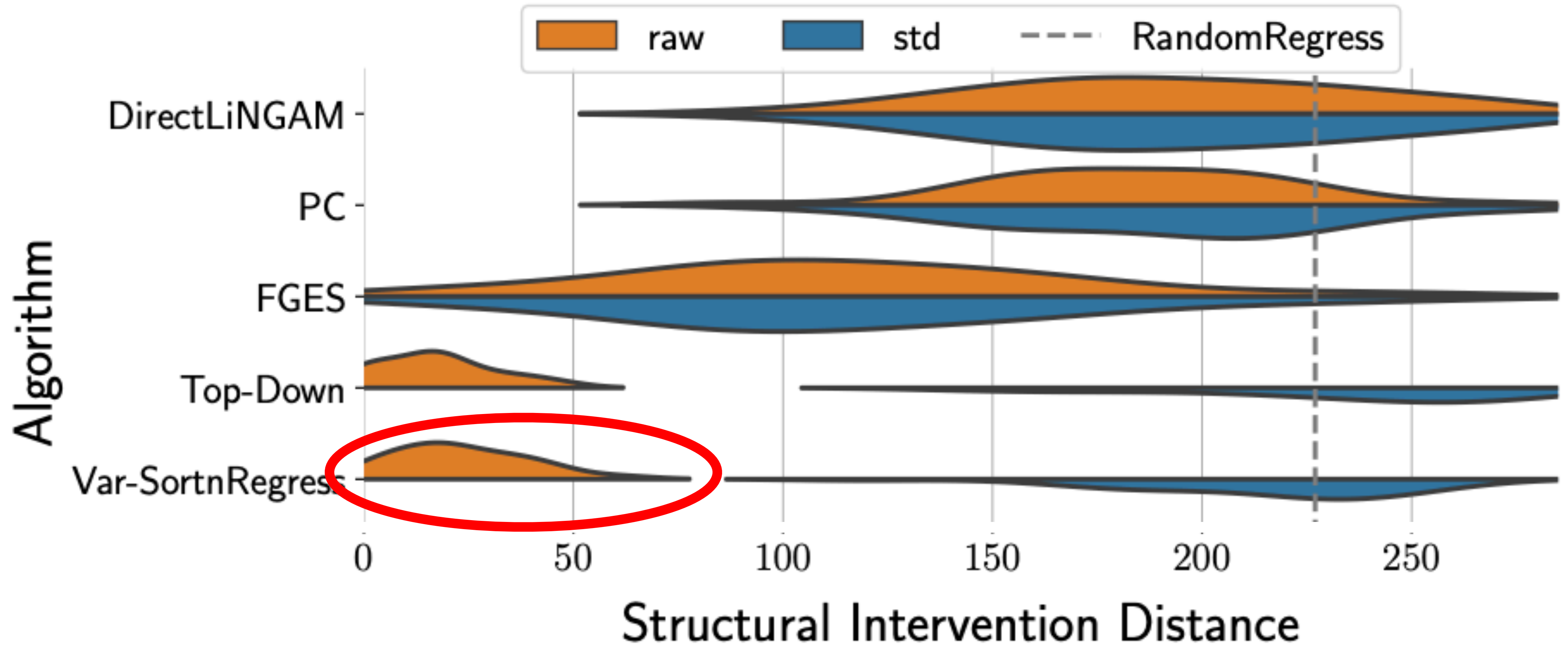


$$x_i := f_i(\mathbf{x}_{\mathrm{pa}(i)}, \varepsilon_i)$$

$$f_i(\mathbf{x}_{\mathrm{pa}(i)}, \varepsilon_i) = \mathbf{w}_i^\top \mathbf{x}_{\mathrm{pa}(i)} + \varepsilon_i$$

Linear SCMs — a common choice for benchmarking causal structure discovery algorithms

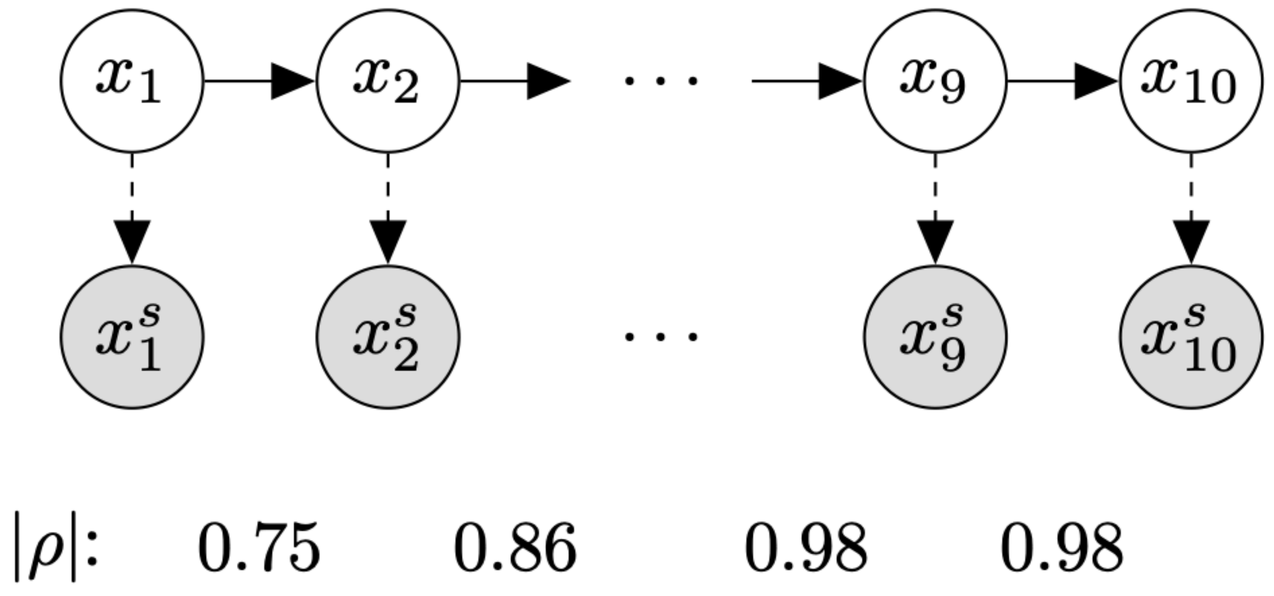
Problem: easy to game by exploiting increasing variance along the causal order



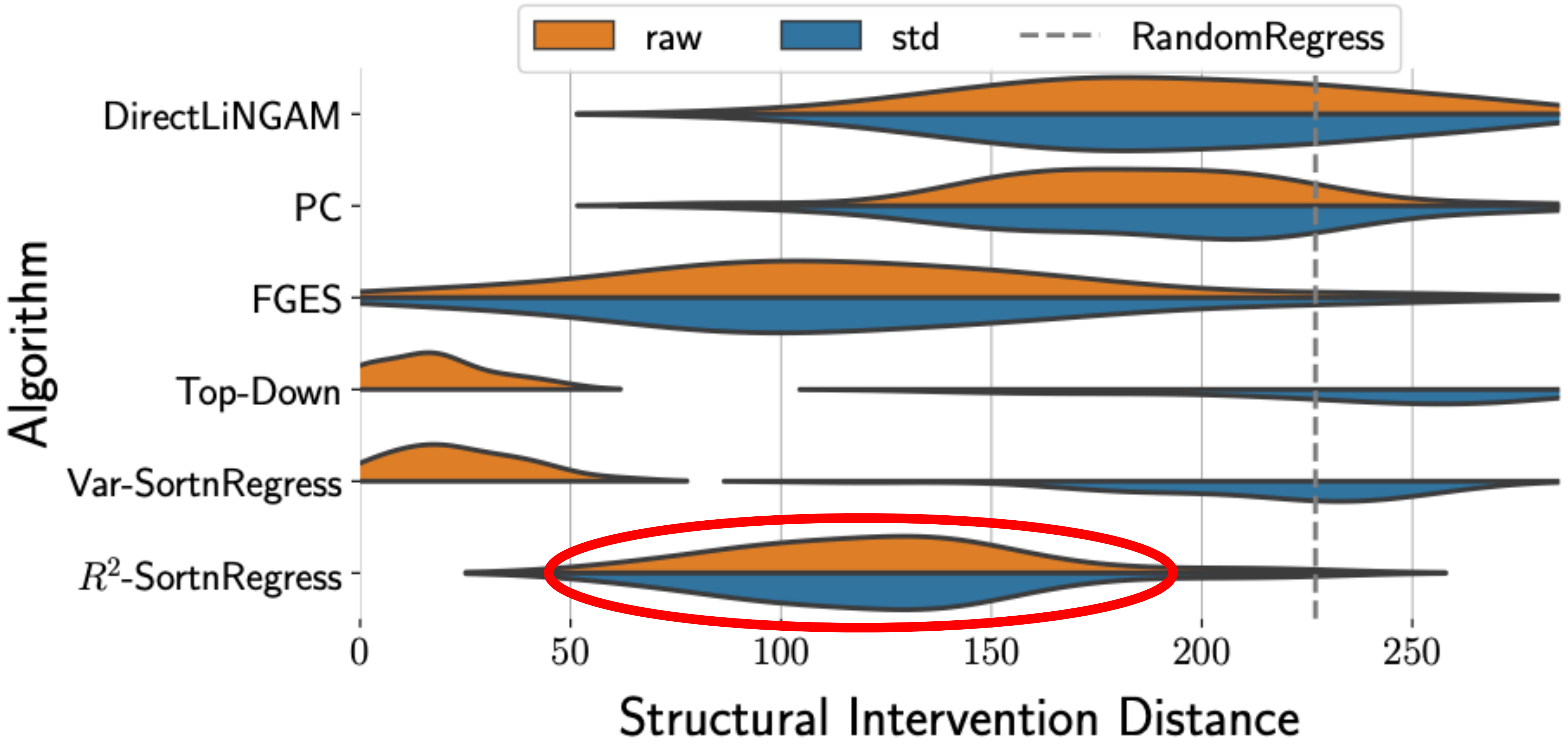
Reisach et al. *A scale-invariant sorting criterion to find a causal order in additive noise models* NeurIPS 2023

Standardized SCMs

Problem: still easy to game by exploiting increasing correlations between variables along the causal ordering



$$x_i^s := \frac{x_i - \mathbb{E}[x_i]}{\sqrt{\mathrm{Var}[x_i]}}$$



Reisach et al. *A scale-invariant sorting criterion to find a causal order in additive noise models* NeurIPS 2023

Our Solution

Internally-Standardized SCMs

Algorithm 1 Sampling from an iSCM

Input: DAG $\mathcal{G}$, noise distribution $\mathcal{P}_{\boldsymbol{\varepsilon}}$, functions $\{f_1, ..., f_d\}$

$\pi \leftarrow$ topological ordering of $\mathcal{G}$

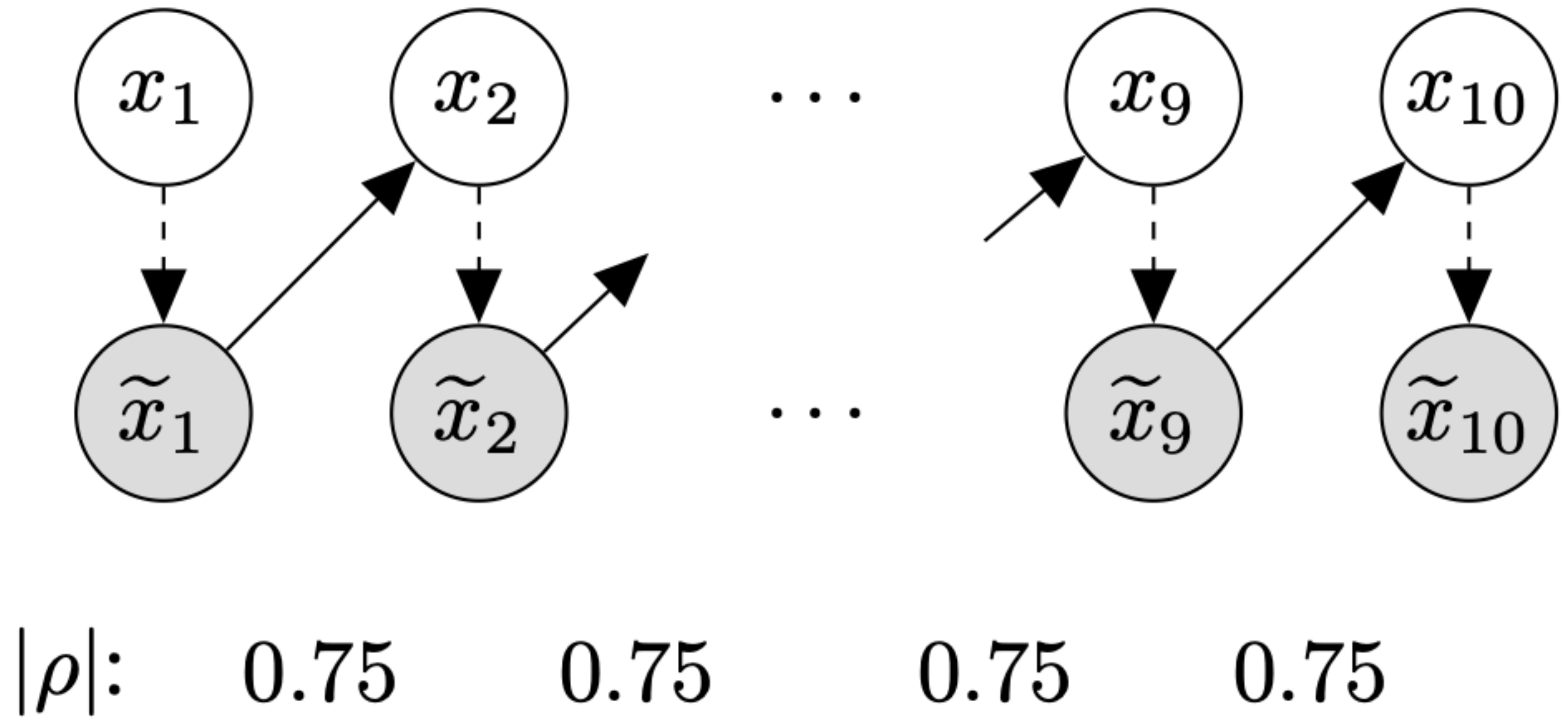
for $i = 1$ to d **do**

$\varepsilon_{\pi_i} \sim \mathcal{P}_{\varepsilon_{\pi_i}}$

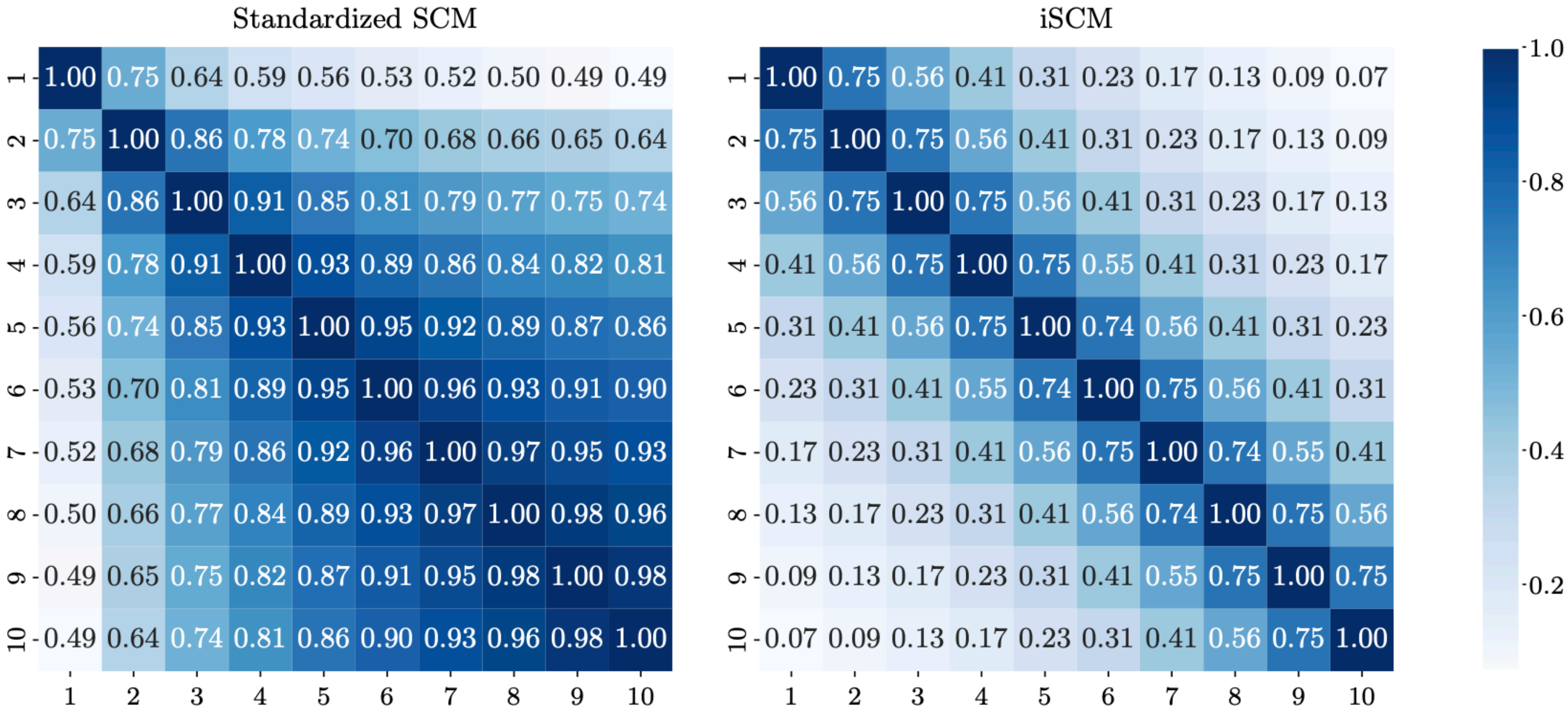
$x_{\pi_i} \leftarrow f_{\pi_i}(\widetilde{\mathbf{x}}_{\mathrm{pa}(\pi_i)}, \varepsilon_{\pi_i})$

$\widetilde{x}_{\pi_i} \leftarrow \dfrac{x_{\pi_i} - \mathbb{E}[x_{\pi_i}]}{\sqrt{\mathrm{Var}[x_{\pi_i}]}}$

return $[\widetilde{x}_1, \ldots, \widetilde{x}_d]$ $\triangleright \in \mathbb{R}^d$



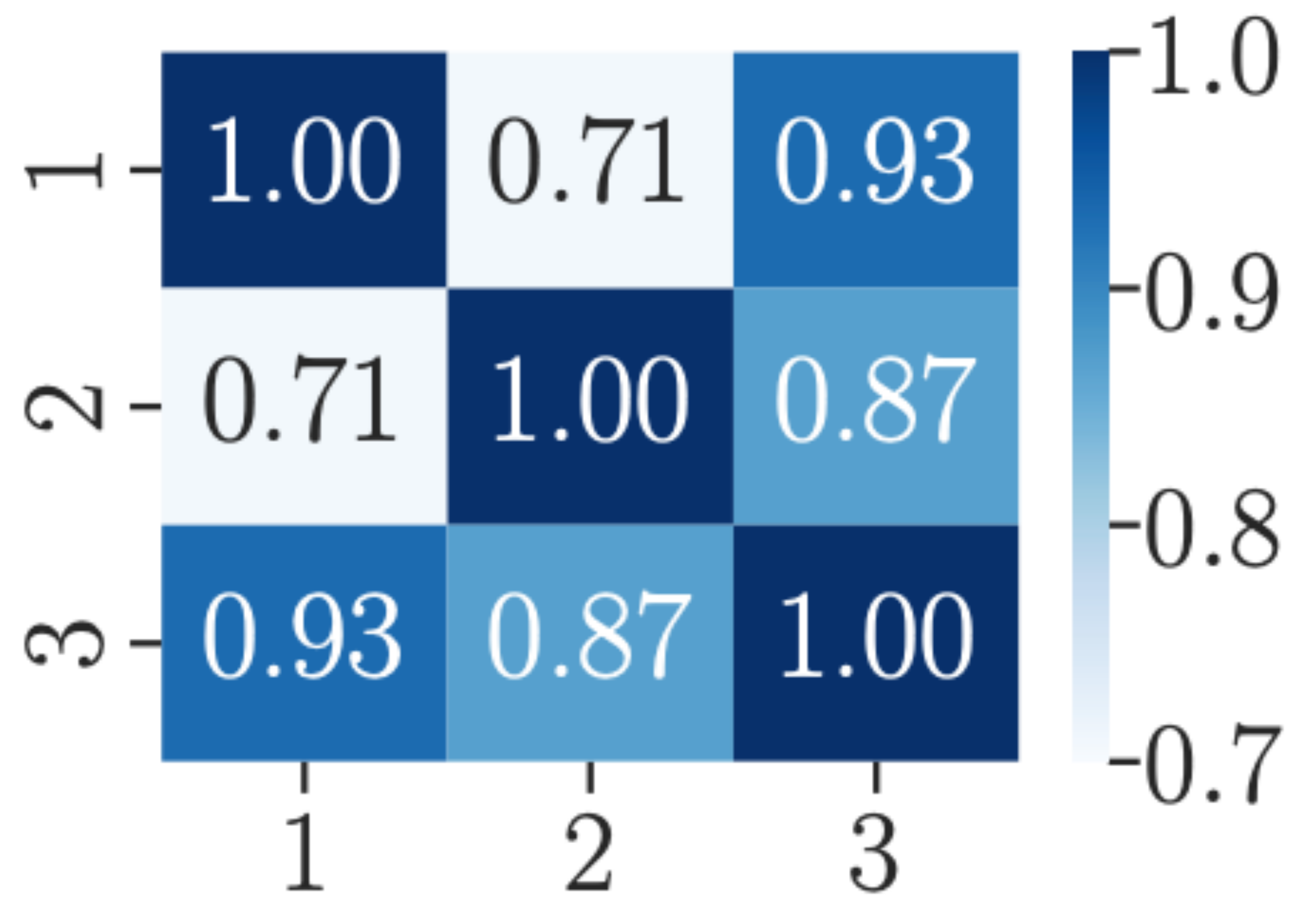
iSCMs (Partially) Remove the Sortability by Correlation

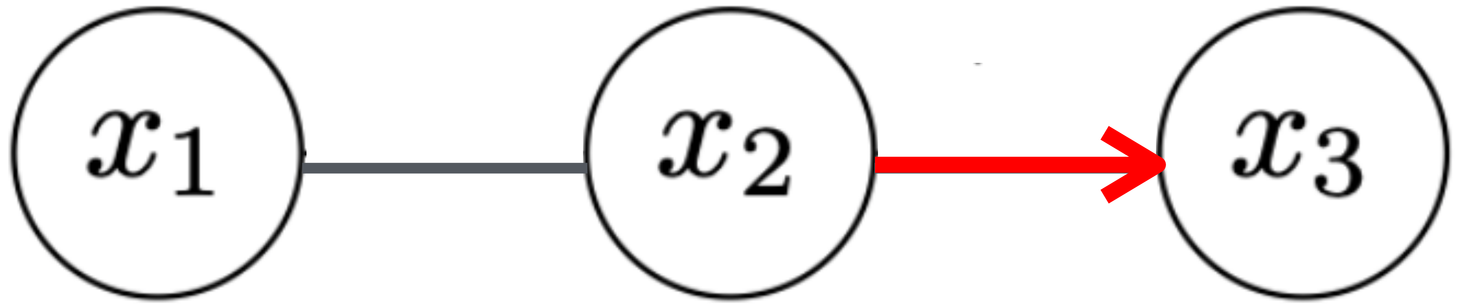


Results

Linear Gaussian stand. SCMs Are Partially Identifiable

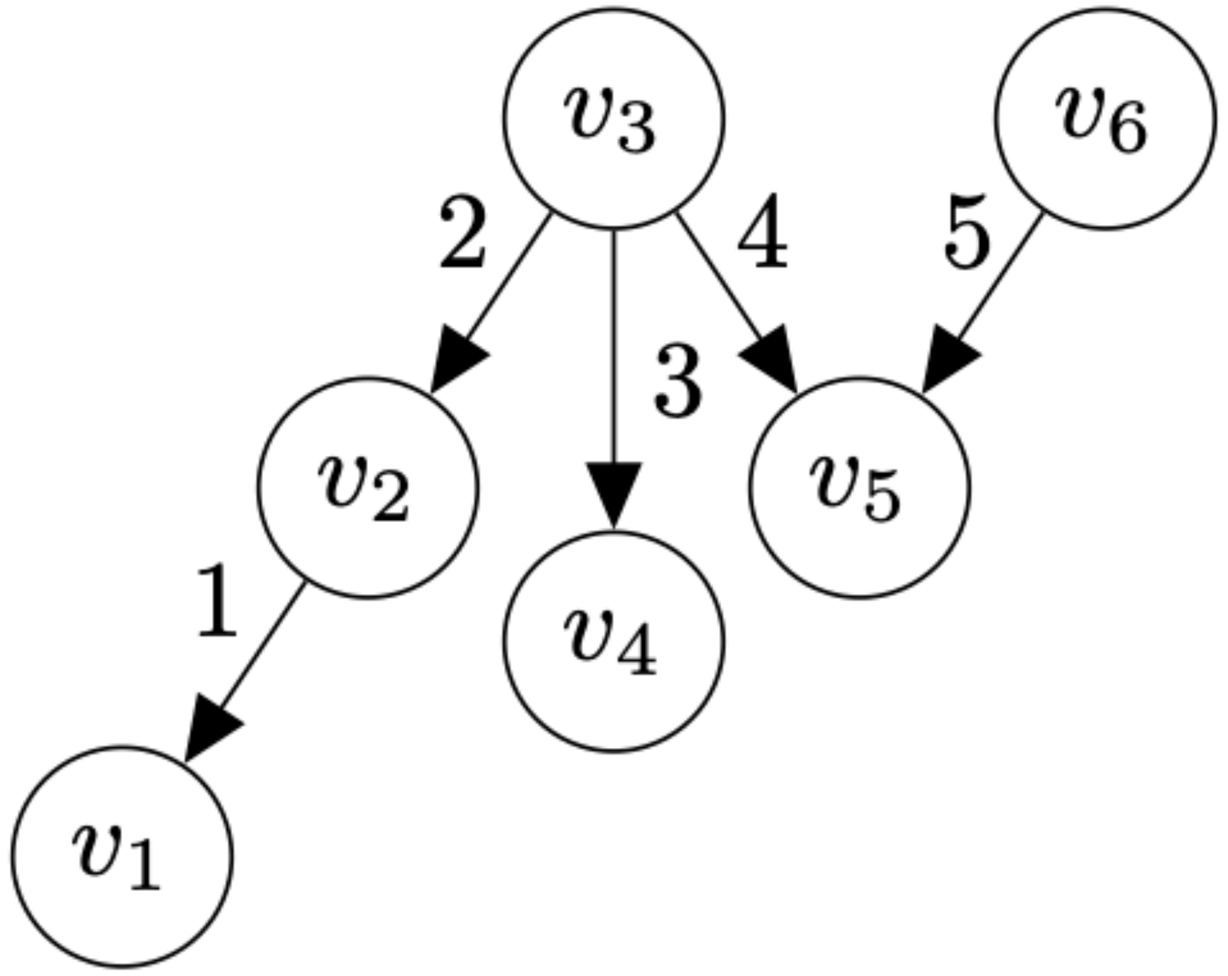
Theorem 3 (Partial identifiability of standardized linear SCMs with forest DAGs). *Let* $\mathbf{x}^s$ *be modeled by a standardized linear SCM* (1) *with forest DAG* $\mathcal{G}$*, additive noise of equal variances* $\mathrm{Var}[\varepsilon_i] = \sigma^2$*, and* $|w_{i,j}| > 1$ *for all* $i \in pa(j)$*. Then, given* $p(\mathbf{x}^s)$ *and the partially directed graph* $\tilde{\mathcal{G}}$ *representing the MEC of* $\mathcal{G}$*, we can identify all but at most one edge of the true DAG* $\mathcal{G}$ *in each undirected connected component of the MEC* $\tilde{\mathcal{G}}$*.*

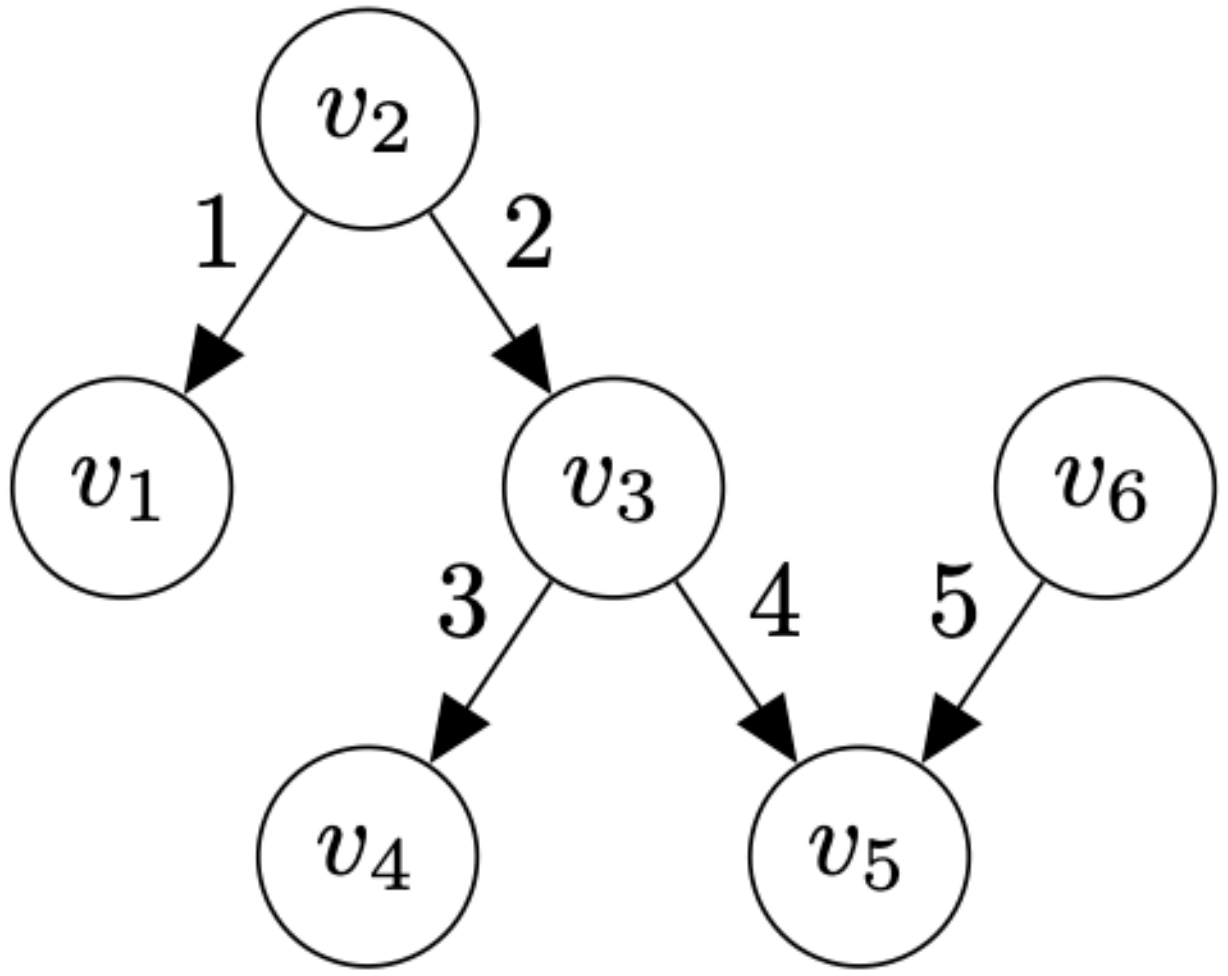


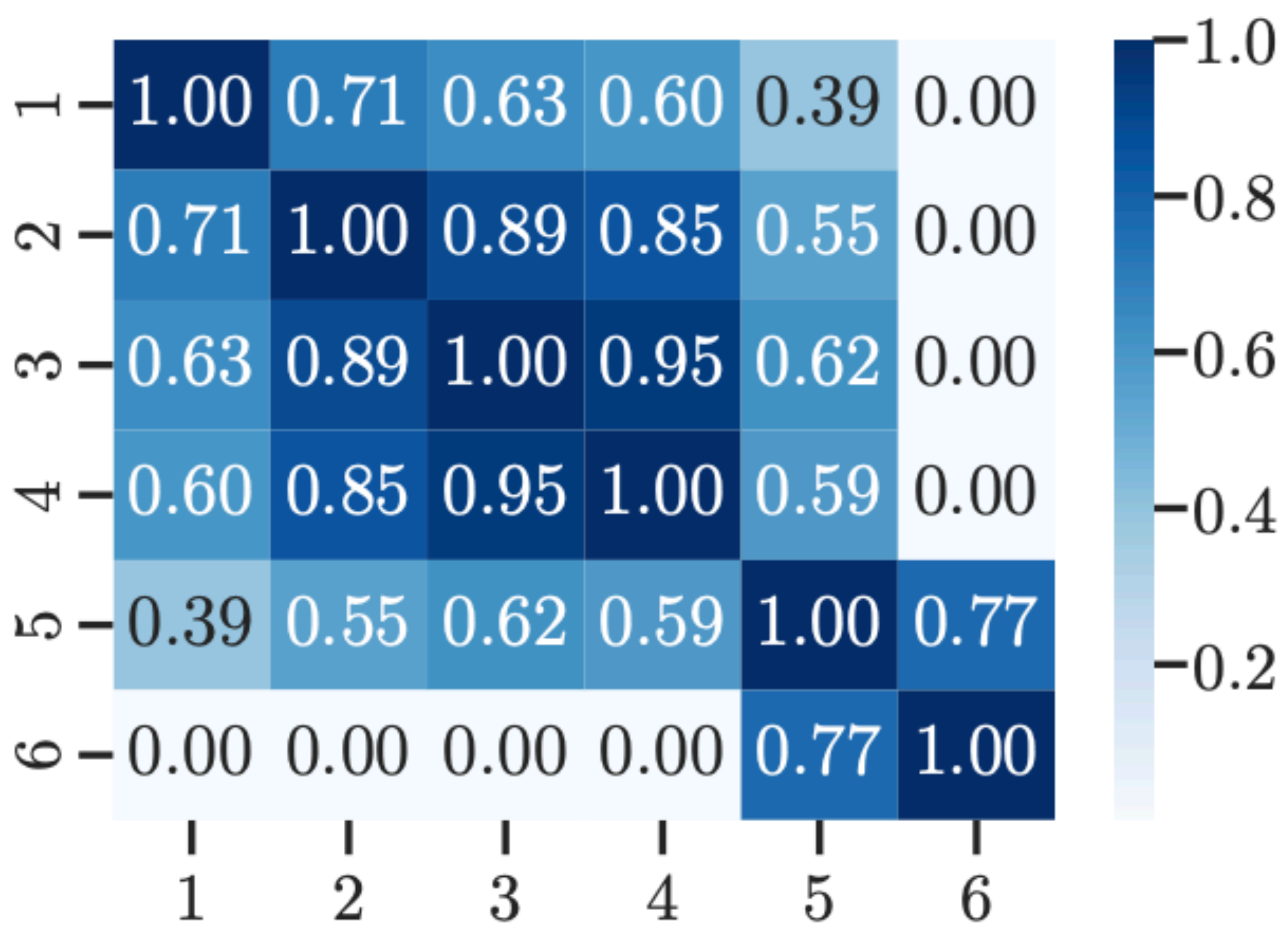


Linear Gaussian iSCMs Are Not Identifiable

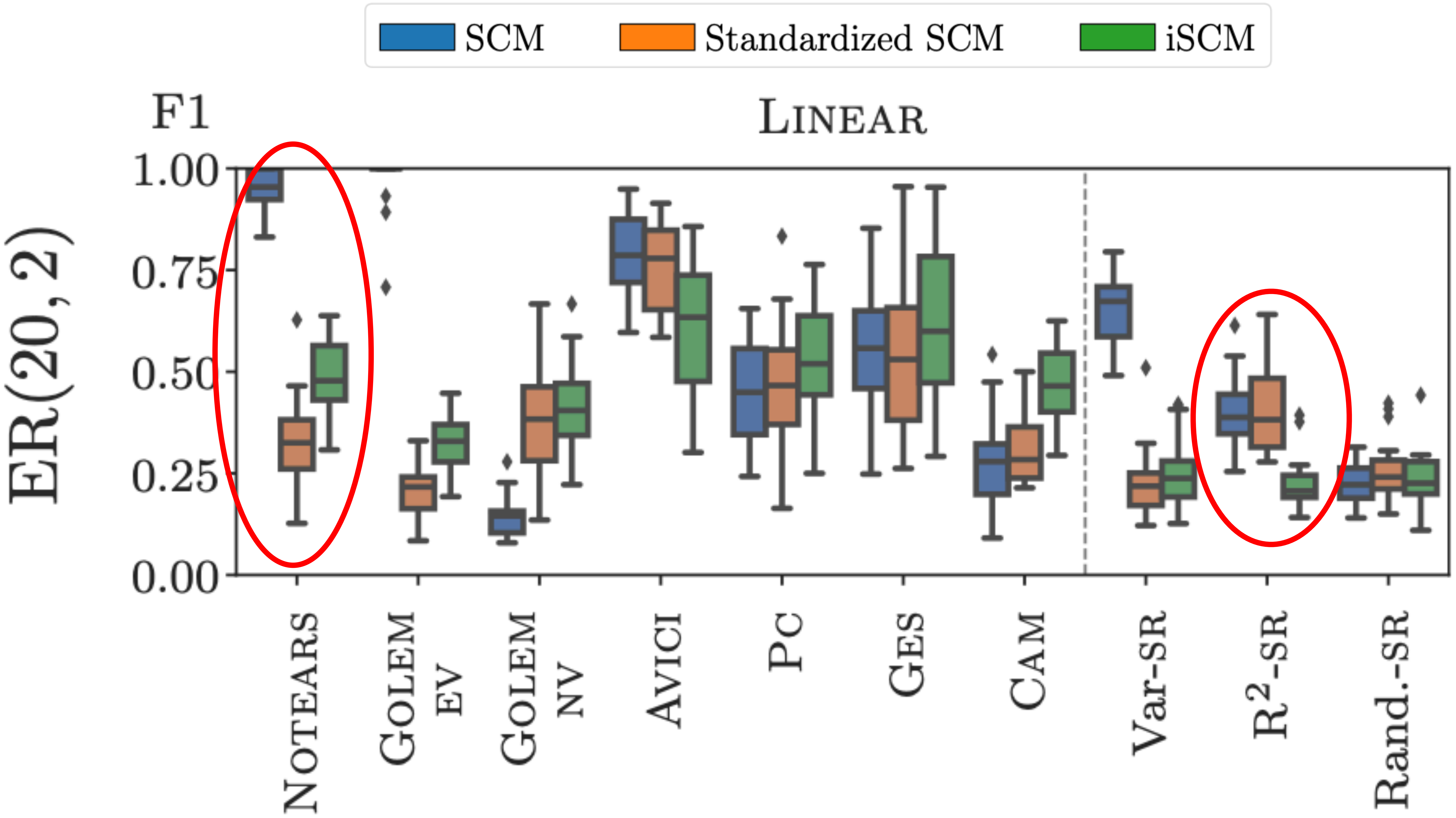
Theorem 4 (Nonidentifiability of linear Gaussian iSCMs with forest DAGs). *Let $\widetilde{\mathbf{x}}$ be modeled by a linear iSCM* (1) *with forest DAG $\mathcal{G}$ and additive Gaussian noise of equal variances* $\mathrm{Var}[\varepsilon_i]$. *Then, for every DAG $\mathcal{G}'$ in the MEC of $\mathcal{G}$, there exists a linear iSCM with DAG $\mathcal{G}'$ that has the same observational distribution as $\widetilde{\mathbf{x}}$, the same noise variances, and the same weights on the corresponding edges in the MEC.*







Numerical Results



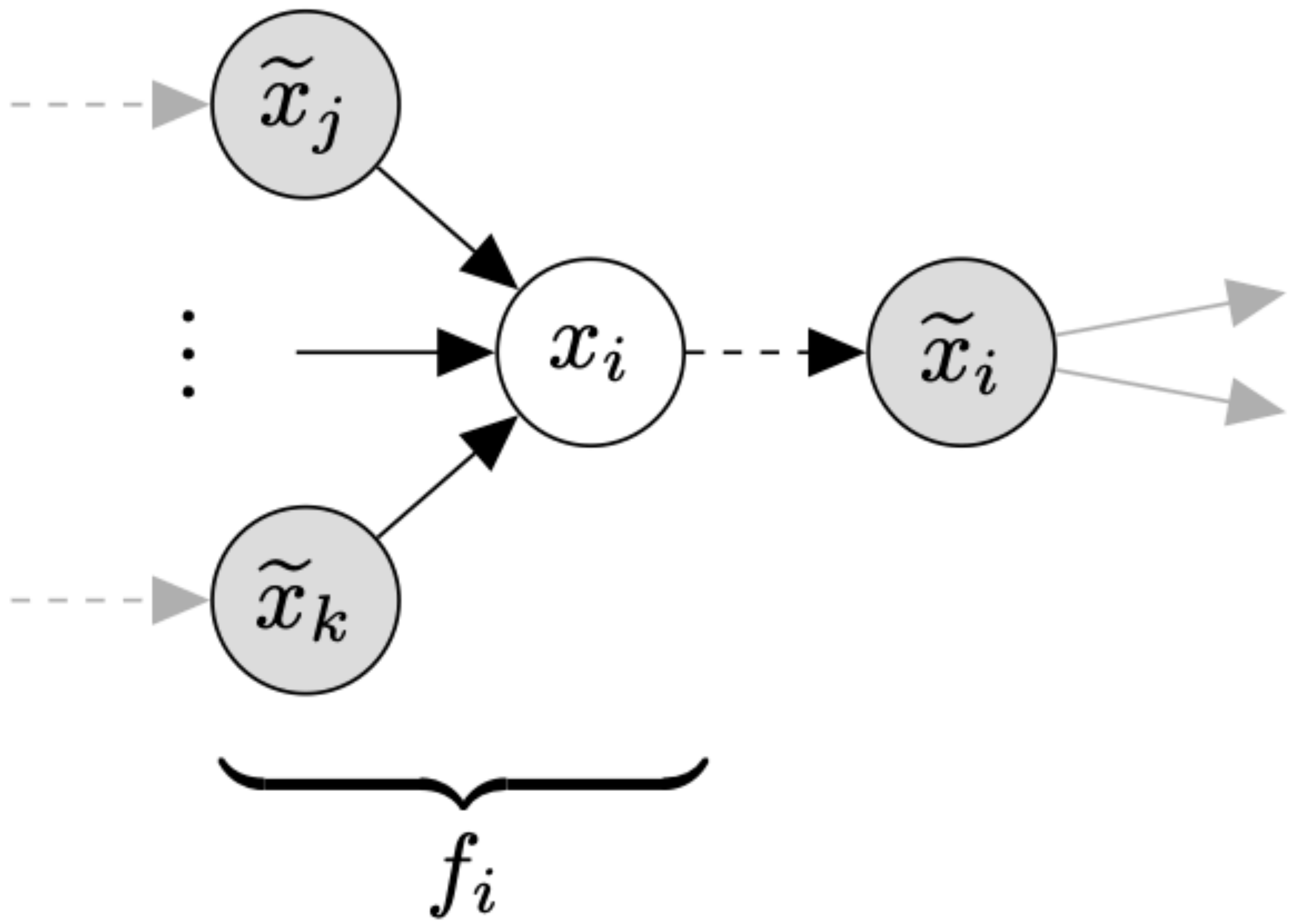
Summary

We propose a simple modification of SCMs that stabilizes the data-generating process and thereby removes exploitable covariance artifacts.

In the paper you will find:

- theoretical analysis of linear iSCMs
- empirical verification of $\mathbf{R}^2 - sortability$ properties
- benchmark of popular causal discovery methods (linear + nonlinear)

Thank you!



Standardizing Structural Causal Models

Weronika Ormaniec
Scott Sussex
Lars Lorch
Bernhard Schölkopf
Andreas Krause

Contact: wormaniec@ethz.ch