ICML
International Conference On Machine Learning

MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

On the Recoverability of Causal Relations from Temporally Aggregated I.I.D. Data

Shunxing Fan[1] Mingming Gong[2,1] Kun Zhang[1,3]

[1]Mohamed bin Zayed University of Artificial Intelligence [2]The University of Melbourne [3]Carnegie Mellon University

THE UNIVERSITY OF MELBOURNE

Carnegie Mellon University

Motivation

What is temporal aggregation? Summation/Averaging/…

$\sum_{t=1}^{k} x_t$, $\frac{1}{k}\sum_{t=1}^{k} x_t$, $\frac{1}{\sqrt{k}}\sum_{t=1}^{k} x_t$ …

Temporal aggregation is so common in real-world observation: Daily/Weekly/Monthly stock prices, hourly/daily/monthly temperature, GDP, electricity consumption…

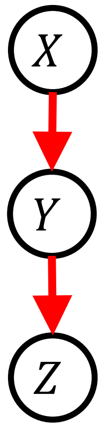
Motivation

Temporal aggregation in non-temporal causal discovery

- Instantaneous causation:

“temporal aggregation is a realistic, plausible, and well-known reason for observing apparent instantaneous causation”

---Granger, C. W. (1988). Some recent development in a concept of causality. *Journal of econometrics*



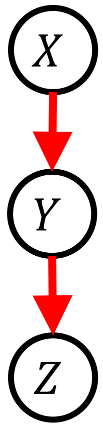
Motivation

Temporal aggregation in Non-temporal causal discovery

- Instantaneous causation:

"temporal aggregation is a realistic, plausible, and well-known reason for observing apparent instantaneous causation"

---Granger, C. W. (1988). Some recent development in a concept of causality. *Journal of econometrics*

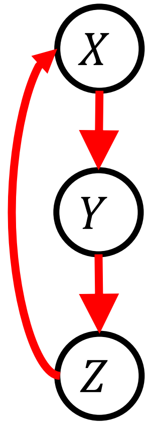


- Cyclic causation:

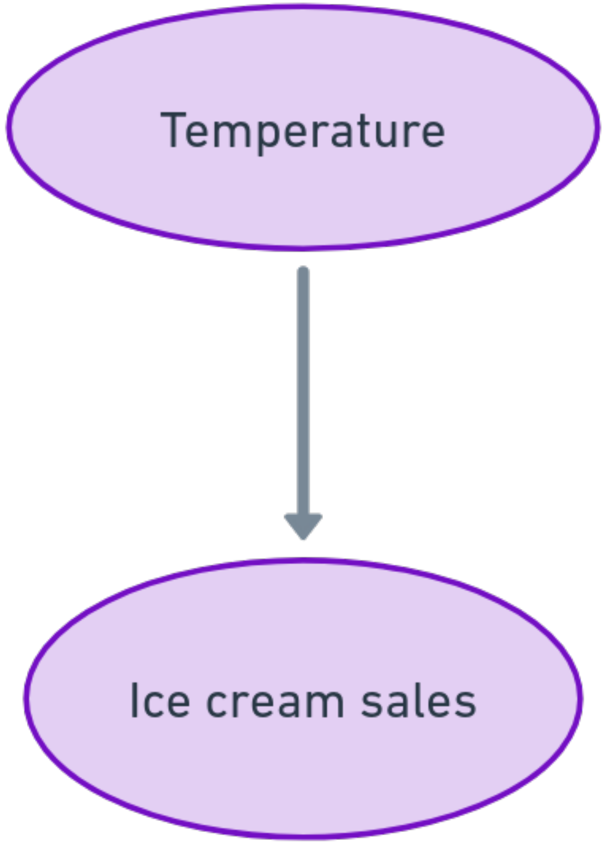
"The observed variable is the vector $\bar{\mathbf{x}}[t]$, defined as the average of $\mathbf{x}$ over the observation period starting at t:"

$$\bar{\mathbf{x}}[t] \equiv \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}[t + k\Delta\theta]$$

---Lacerda, G., Spirtes, P. L., Ramsey, J., & Hoyer, P. O. (2008). Discovering cyclic causal models by independent components analysis. UAI 2008

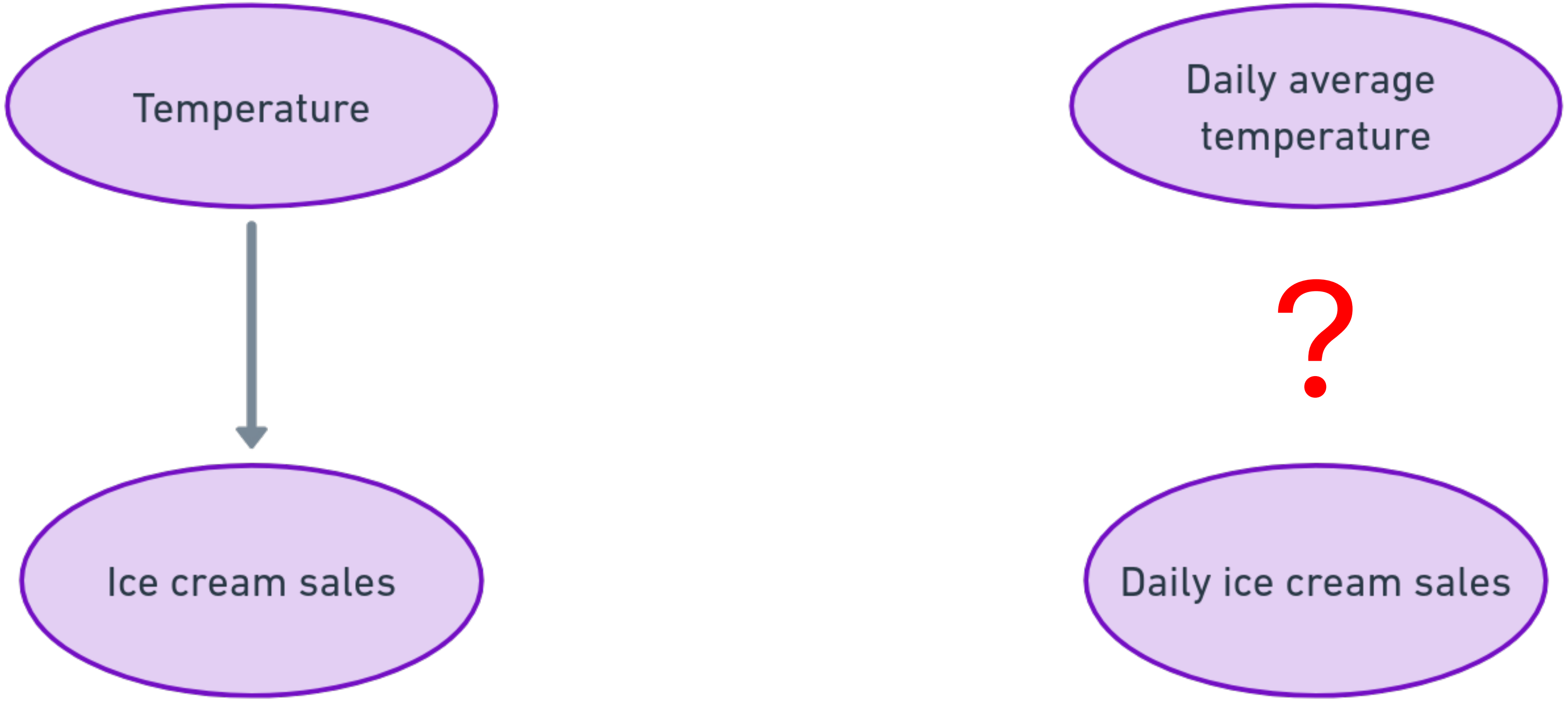


Motivation



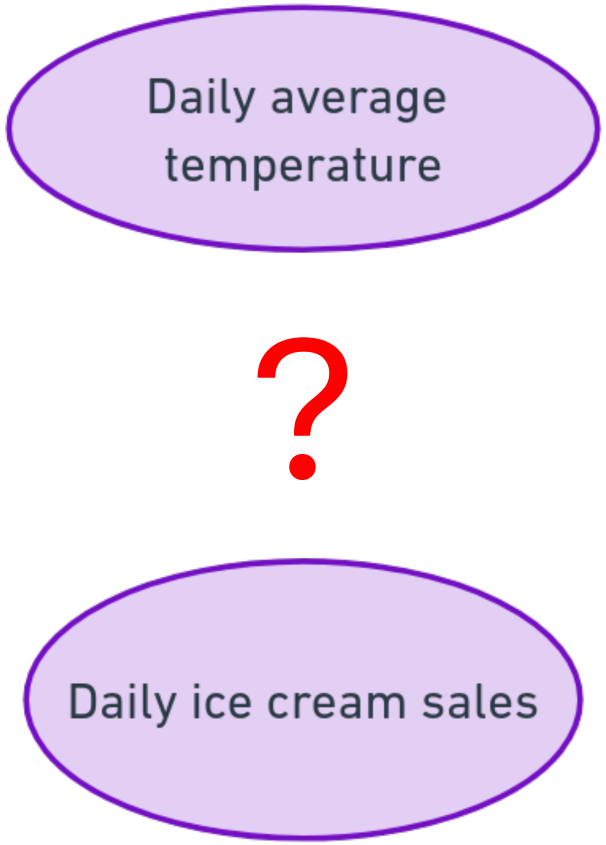
Motivation

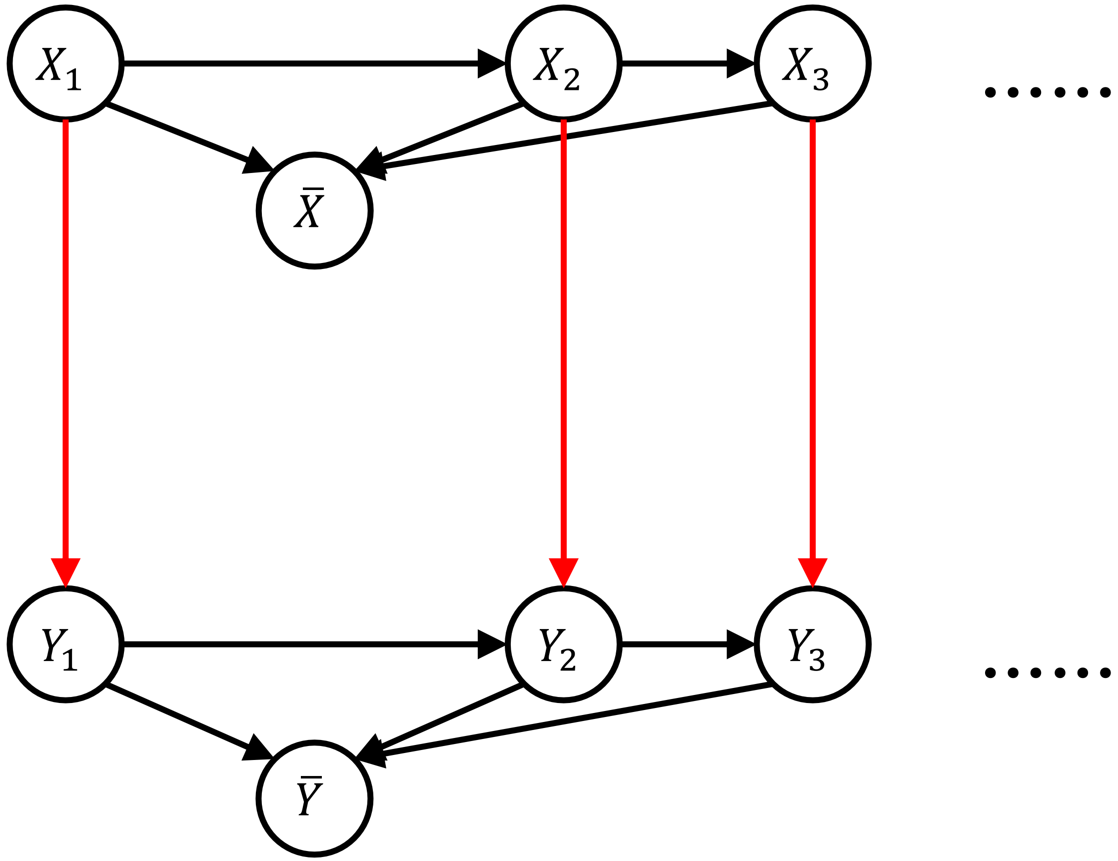
Is the causal relationship consistent across different observational levels?



Motivation

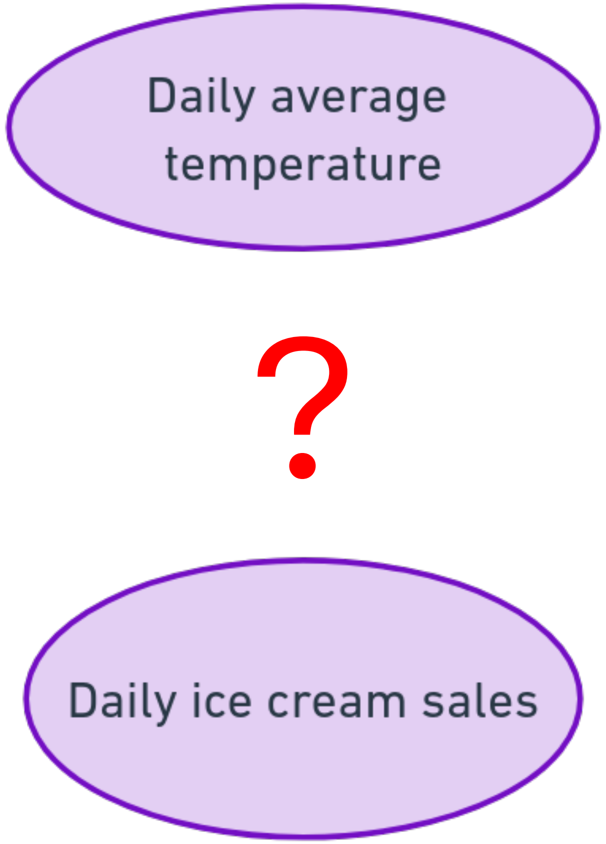
Is the causal relationship consistent across different observational levels?

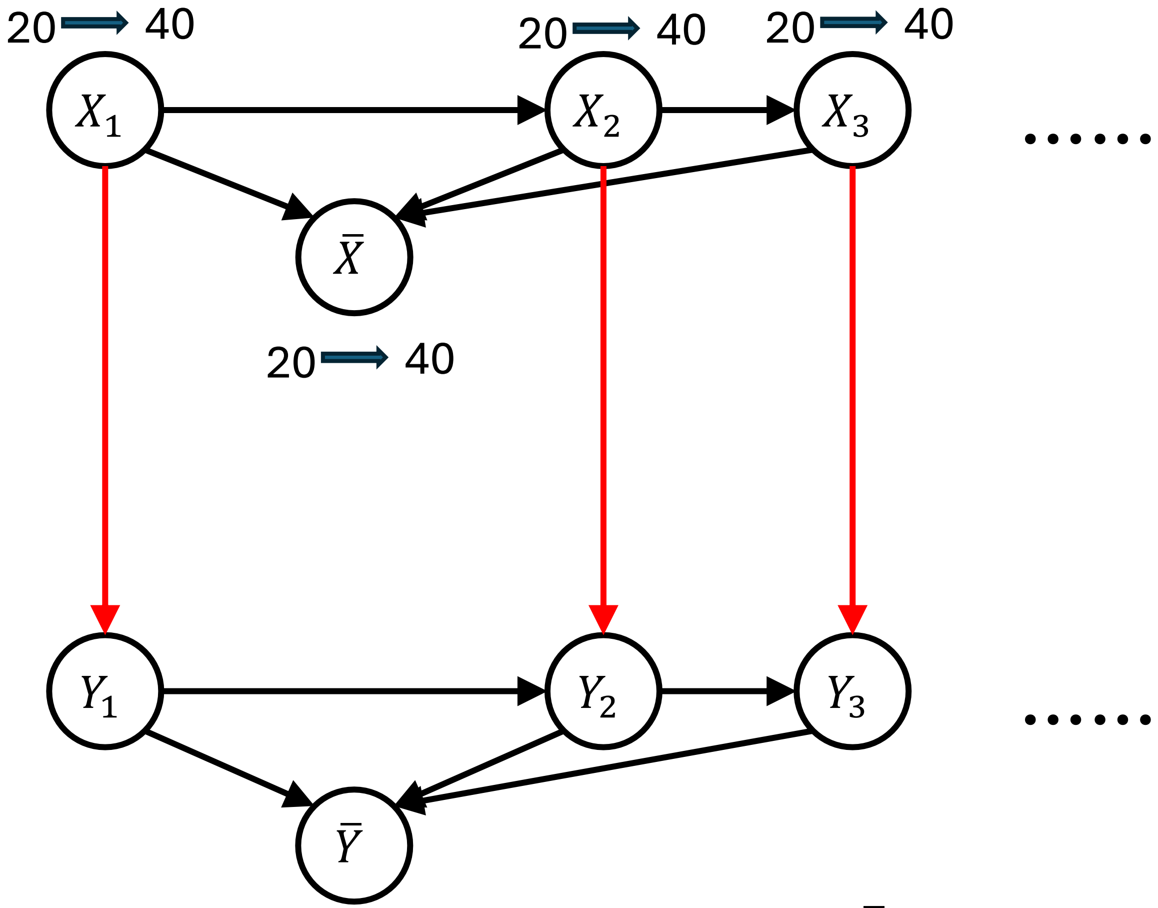




Motivation

Is the causal relationship consistent across different observational levels?

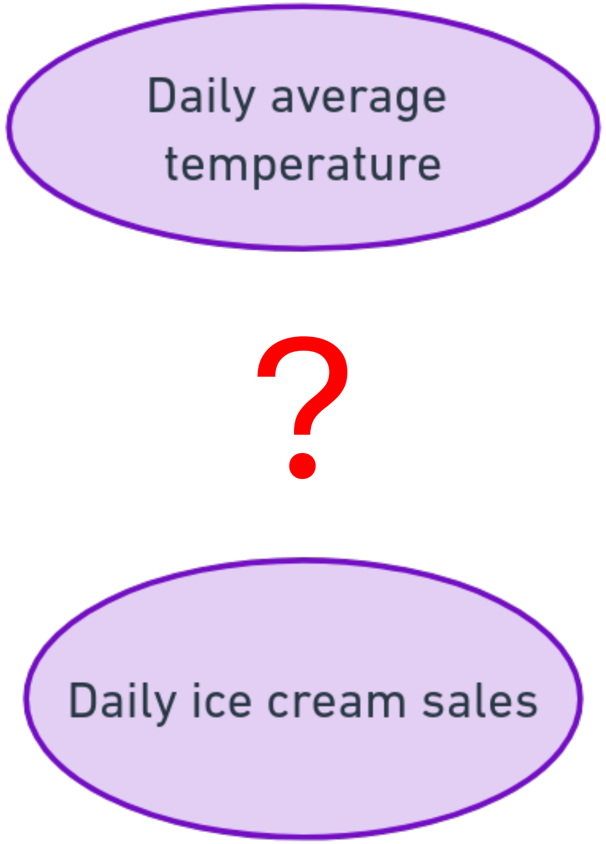


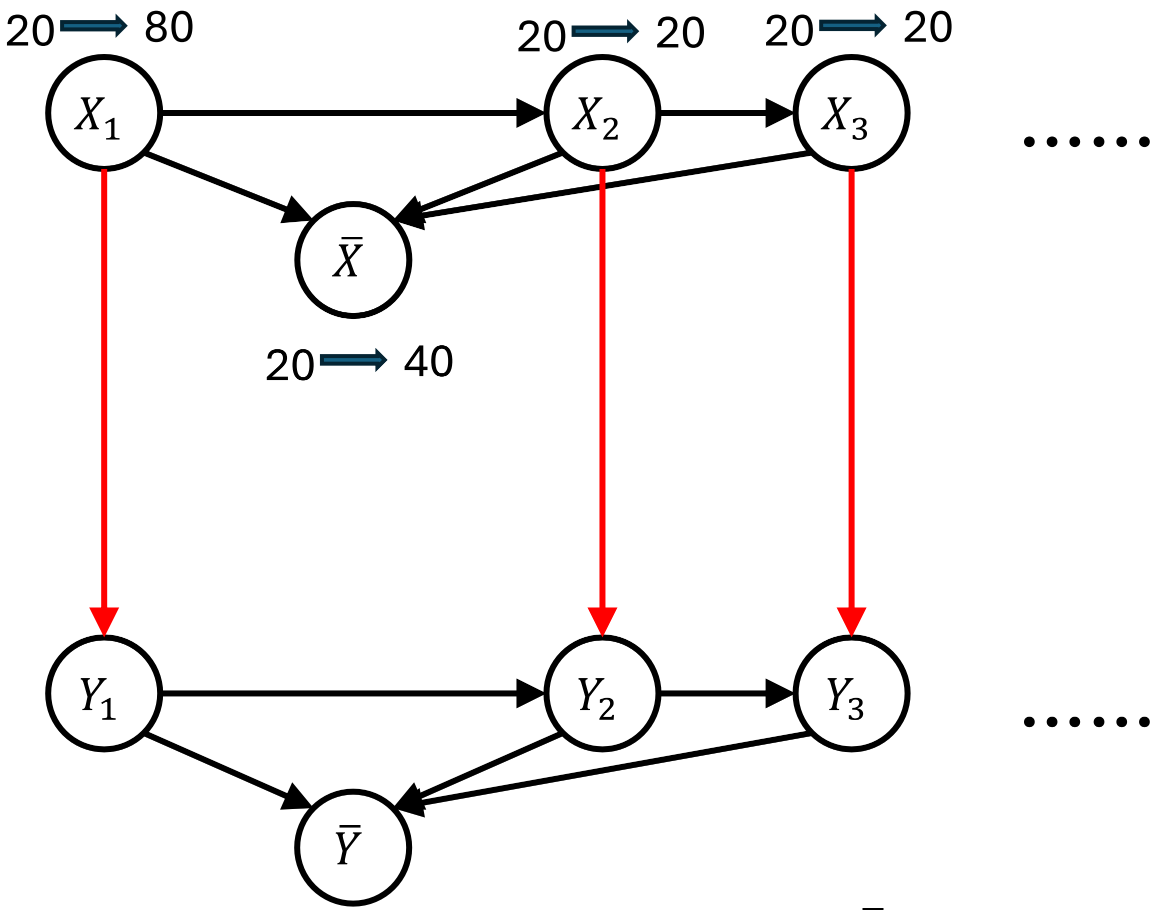


Intervention on $\bar{X}$ is not well-defined

Motivation

Is the causal relationship consistent across different observational levels?

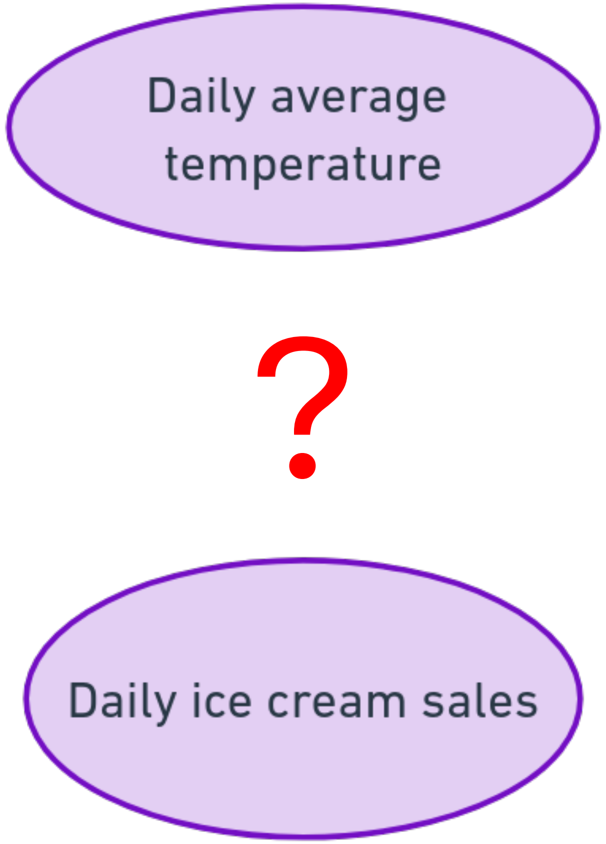


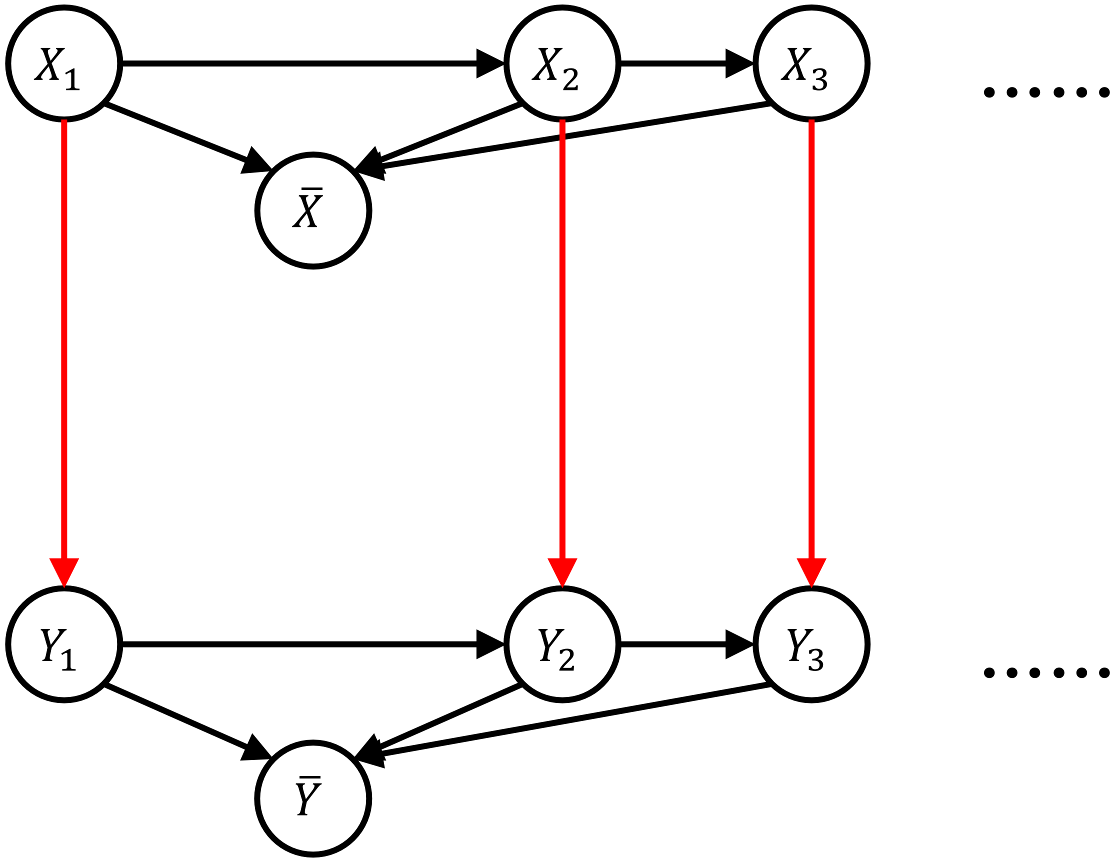


Intervention on $\bar{X}$ is not well-defined

Motivation

Is the causal relationship consistent across different observational levels?

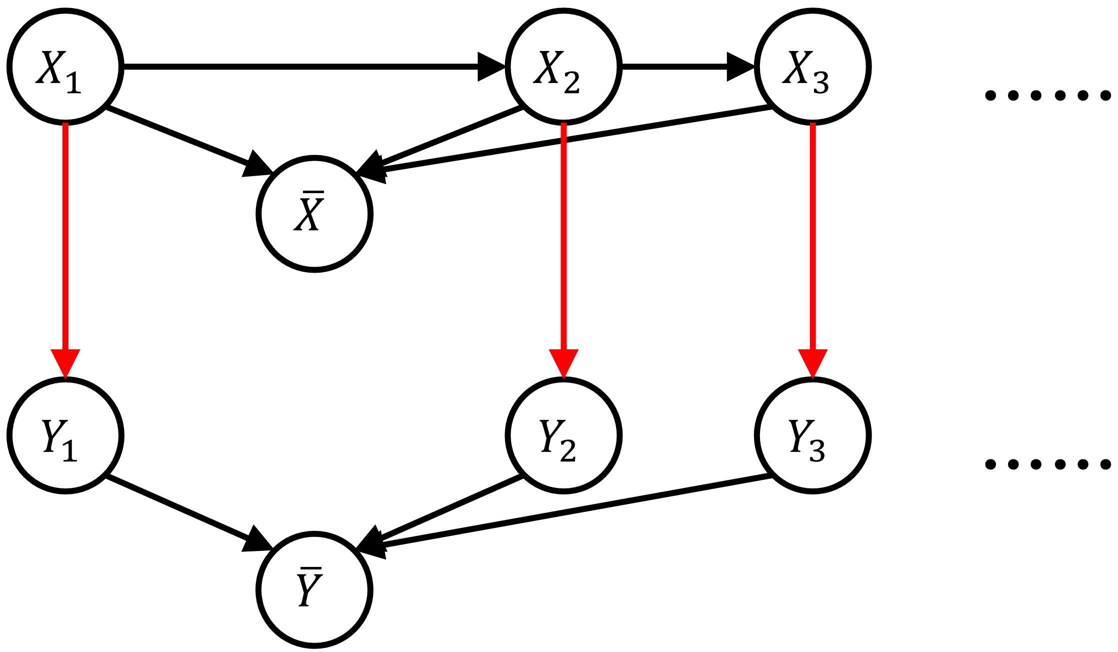




X_t is the cause of Y_t, but $\bar{X}$ may not be the cause of $\bar{Y}$.

Linear

$$Y_t = aX_t + N_t$$

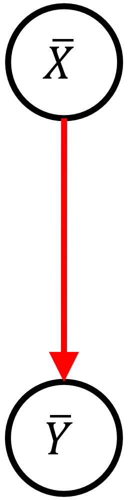


Linear

$$Y_t = aX_t + N_t$$

Thanks to additivity and homogeneity of linear transformation:

$$\bar{Y} = a\bar{X} + \bar{N}$$

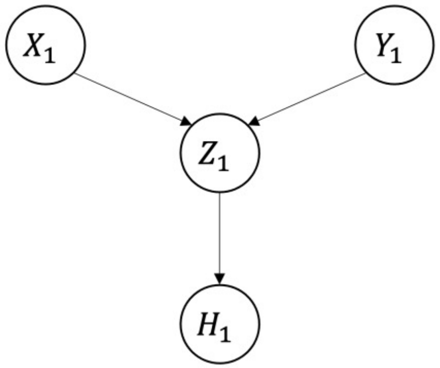


Gong, M., Zhang, K., Schölkopf, B., Glymour, C., & Tao, D. (2017, August). Causal discovery from temporally aggregated time series. UAI 2017

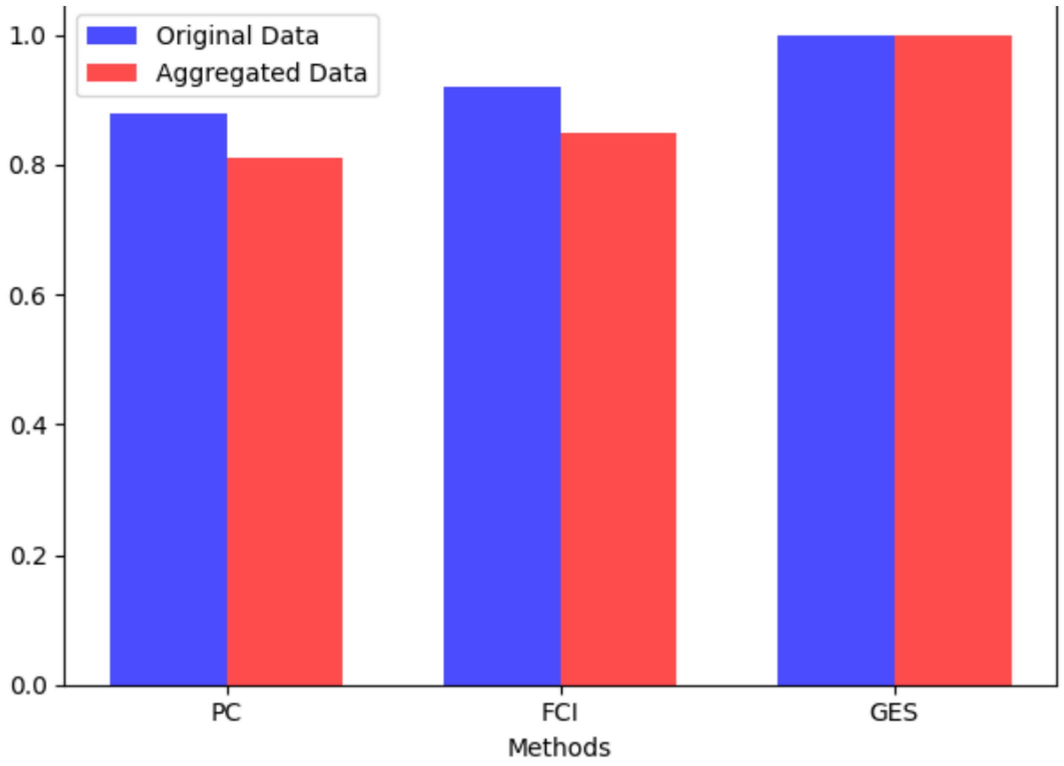
Causal model:

$$
\begin{aligned}
Z &= X + Y + N_Z, \\
H &= Z + N_H \quad \text{(for linear);} \\
Z &= X^2 + Y^2 + N_Z, \\
H &= Z^2 + N_H \quad \text{(for nonlinear).}
\end{aligned}
$$

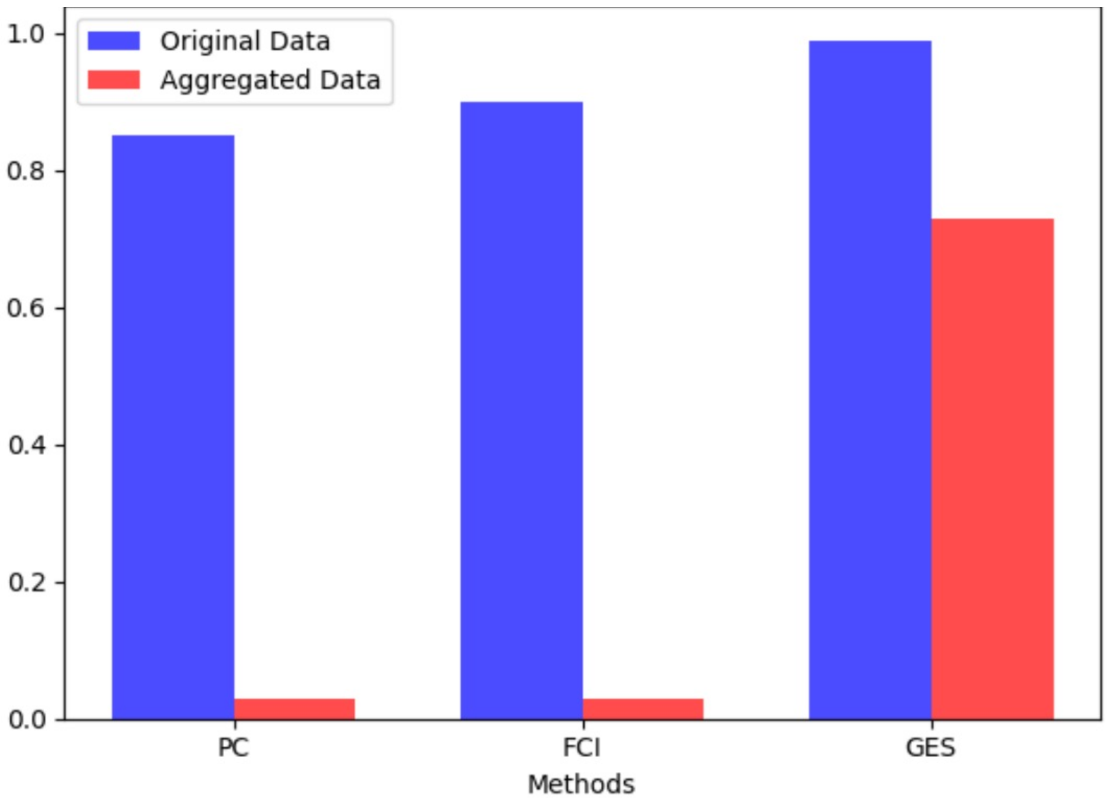
Ground truth:



Accuracy in linear case



Accuracy in non-linear case



Motivation

Many real-world observational data can be considered as the result of aggregation from fine-grained, micro-level, non-linear causal processes. How can we trust the real-world results from non-linear causal discovery methods given that the causal relationship may be inconsistent across different levels?

Our work focuses on:

- When will the causal discovery fail or succeed on aggregated data?
- How the causal discovery results go wrong?

Functional Consistency (LiNGAM, ANM…)

Definition 3.2 (Functional Consistency Regarding Additive Noise). Consider the bivariate aligned model defined in 3.1 incorporates additive noise: $Y_t = f(X_t) + N_{Y,t}$. This process exhibits functional consistency regarding additive noise if there exists a function $\hat{f}$ such that the aggregated variables can be represented as $\overline{Y} = \hat{f}(\overline{X}) + N$, where N is independent of $\overline{X}$, and such $\hat{f}$ exists only in the correct causal direction.

Functional Consistency

Theorem 3.3 (Construction of $\hat{f}$). *If such $\hat{f}$, as defined in Definition 3.2, exists, then $\hat{f}$ must take the form:*

$$\hat{f}(T) = \mathbb{E}\left(\sum_{i=1}^{k} f(X_i) \mid \overline{X} = T\right) + c, \tag{2}$$

where c is any constant (which can be incorporated into the noise term) and the expression $\mathbb{E}(\cdot \mid \overline{X} = T)$ denotes the conditional expectation. For simplicity, we set $c = 0$. Consequently, this implies:

$$\mathbb{E}(\hat{f}(\overline{X})) = \mathbb{E}\left(\sum_{i=1}^{k} f(X_i)\right). \tag{3}$$

Theorem 3.4 (Necessary and Sufficient Condition). *The necessary and sufficient condition for the existence of the additive noise causal model defined in Definition 3.2 is that $N = \sum_{i=1}^{k} N_{Y,i} + \left(\sum_{i=1}^{k} f(X_i) - \hat{f}(\overline{X})\right)$ is independent of $\overline{X}$, where $\hat{f}$ is defined by Eq. 2.*

Functional Consistency

Direct LiNGAM

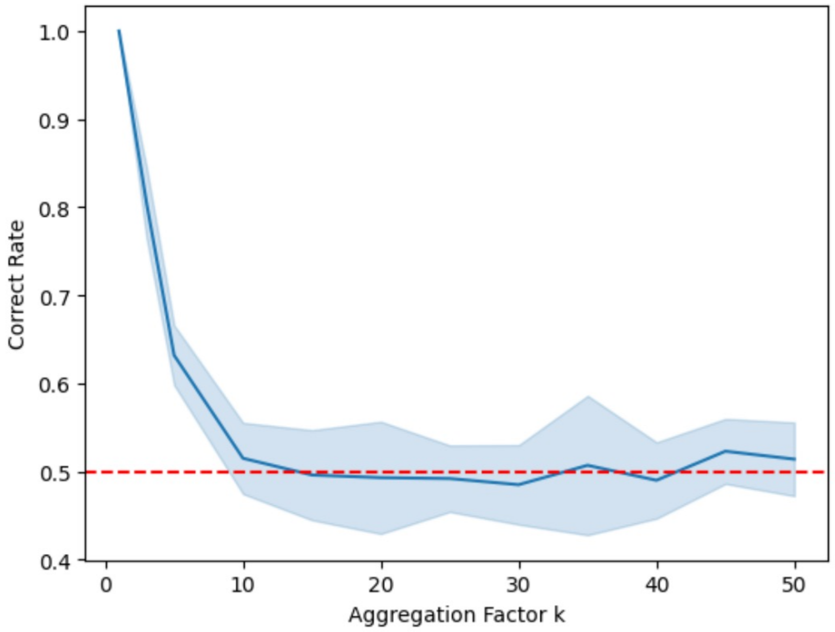


Figure 7. Linear Case: Direct LiNGAM Correction Rate with Different Aggregation Factors k. The blue area represents the standard deviation. The red line represents the random guess baseline.

Additive noise model:

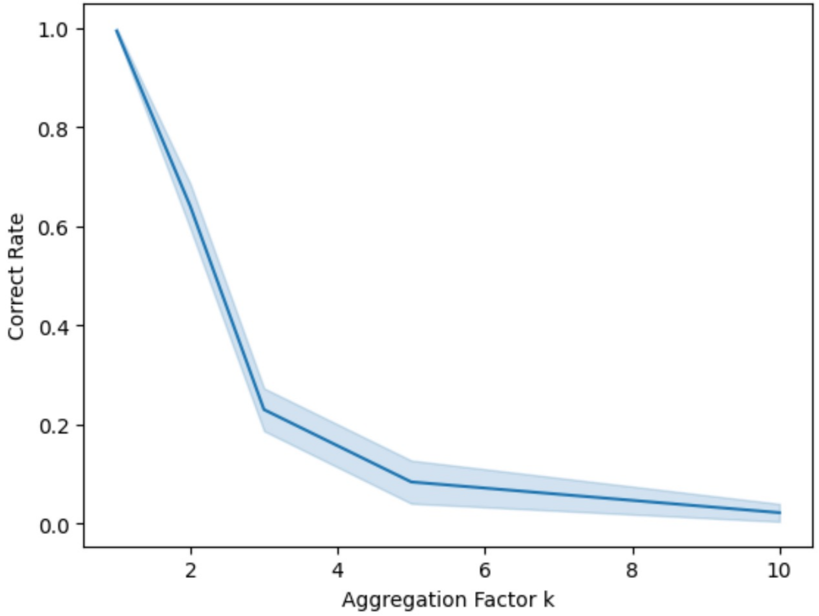
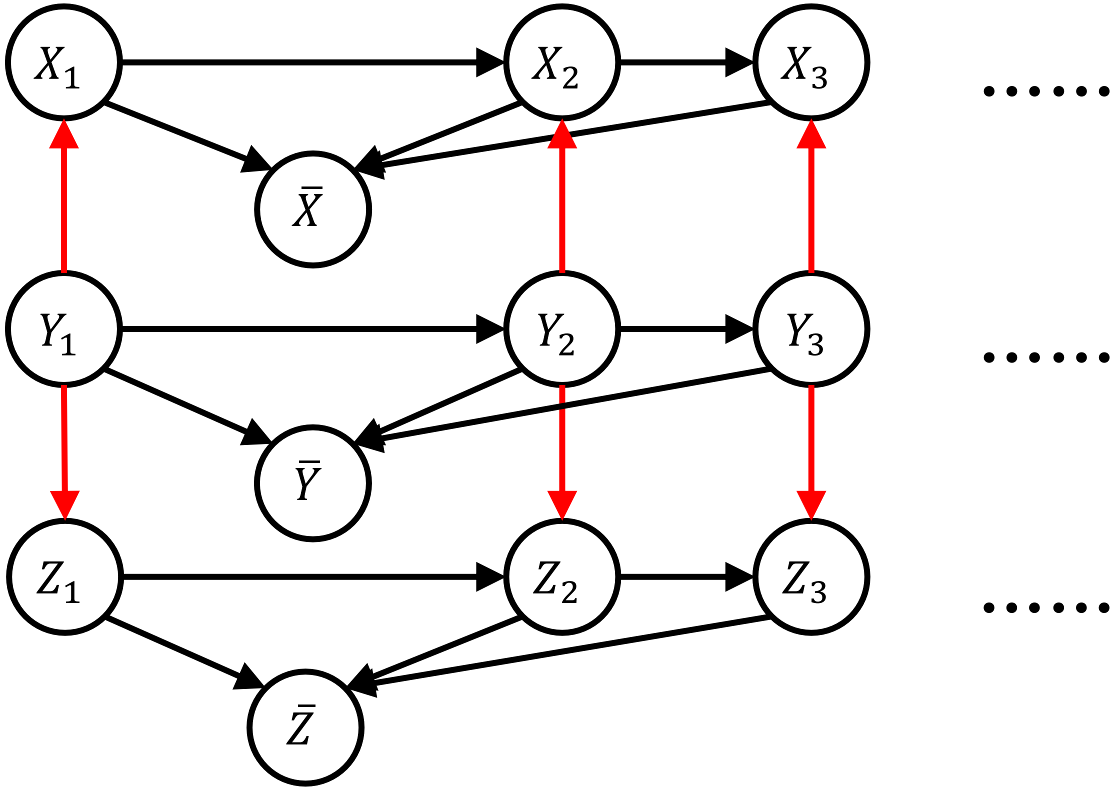
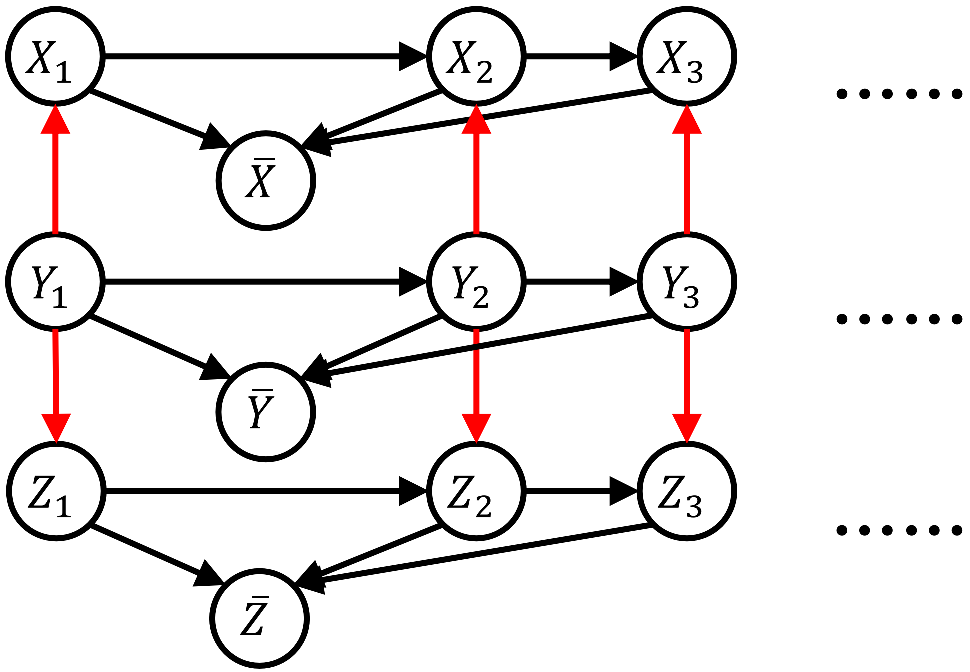


Figure 8. Nonlinear Case: ANM Correction Rate with Different Aggregation Factors k. The blue area represents the standard deviation.

Conditional Independence Consistency (PC, FCI...)



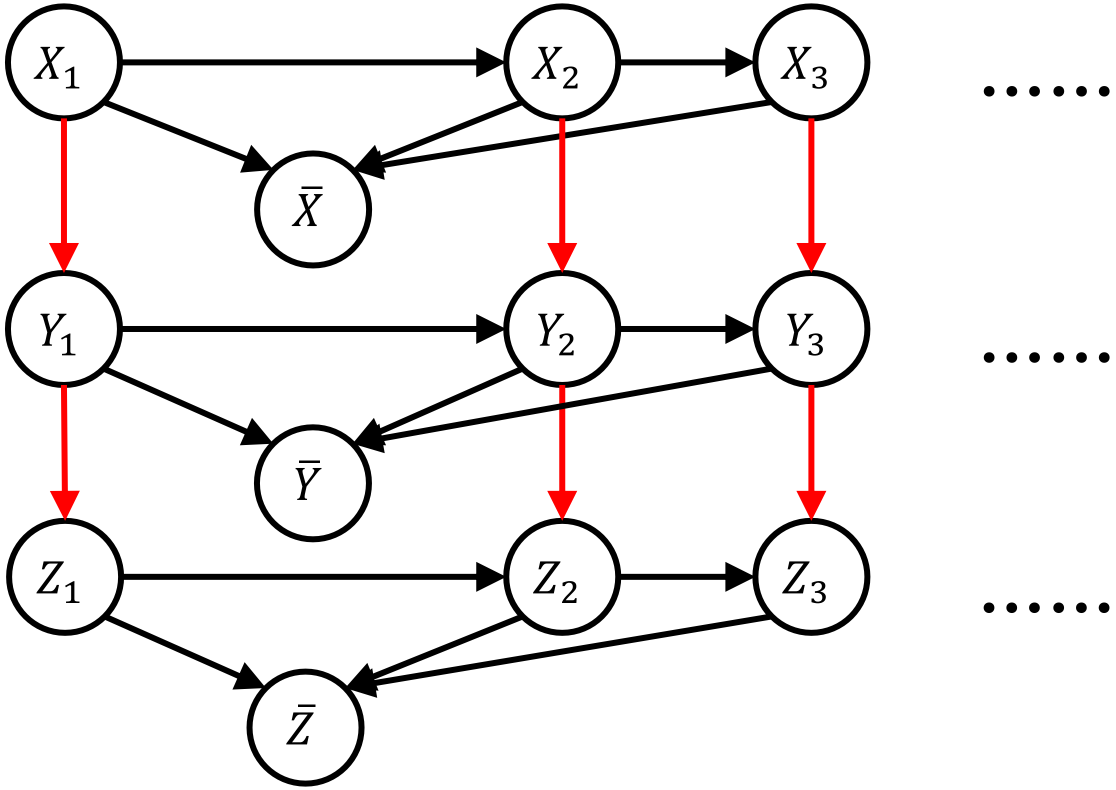


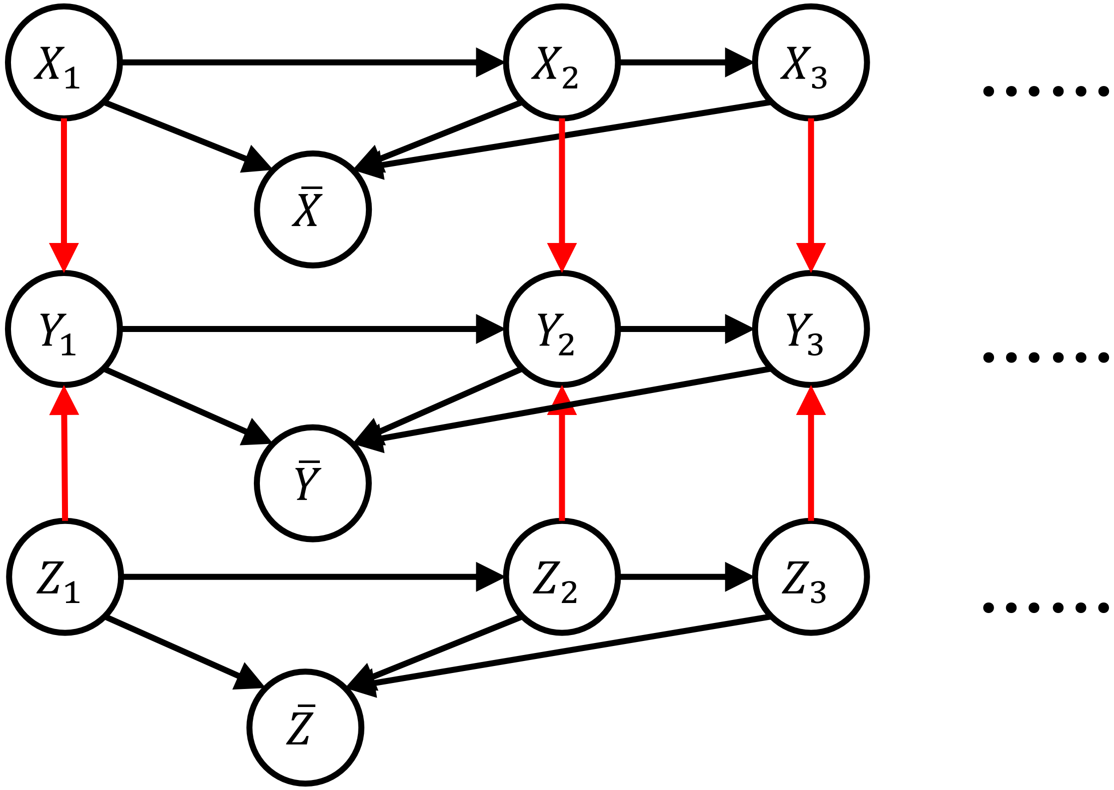
(I) $\overline{X} \perp\!\!\!\perp \overline{Y}$, (II) $\overline{Y} \perp\!\!\!\perp \overline{Z}$, (III) $\overline{X} \perp\!\!\!\perp \overline{Z}$, (IV) $\overline{X} \perp\!\!\!\perp \overline{Y} \mid \overline{Z}$, (V) $\overline{Y} \perp\!\!\!\perp \overline{Z} \mid \overline{X}$, (VI) $\overline{X} \perp\!\!\!\perp \overline{Z} \mid \overline{Y}$.

Rejection Rate for CIT:

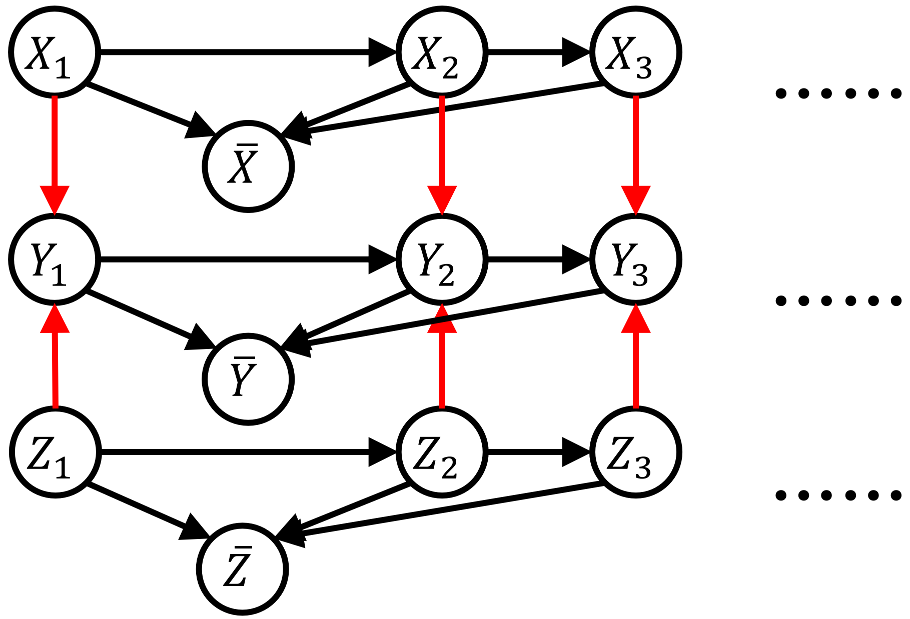
(b) Fork Structure

$X_t \to Y_t$	$Y_t \to Z_t$	I	II	III	IV	V	VI
Linear	Linear	100%	100%	100%	100%	100%	5%
Nonlinear	Linear	92%	100%	84%	92%	100%	5%
Linear	Nonlinear	100%	93%	85%	100%	93%	5%
Nonlinear	Nonlinear	92%	93%	72%	86%	87%	58%


X_1
X_2
X_3
$\bar{X}$
Y_1
Y_2
Y_3
$\bar{Y}$
Z_1
Z_2
Z_3
$\bar{Z}$
……
……
……



X_1
X_2
X_3
$\bar{X}$
......
Y_1
Y_2
Y_3
$\bar{Y}$
......
Z_1
Z_2
Z_3
$\bar{Z}$
......



(I) $\overline{X} \perp\!\!\!\perp \overline{Y}$, (II) $\overline{Y} \perp\!\!\!\perp \overline{Z}$, (III) $\overline{X} \perp\!\!\!\perp \overline{Z}$, (IV) $\overline{X} \perp\!\!\!\perp \overline{Y} \mid \overline{Z}$, (V) $\overline{Y} \perp\!\!\!\perp \overline{Z} \mid \overline{X}$, (VI) $\overline{X} \perp\!\!\!\perp \overline{Z} \mid \overline{Y}$.

Rejection Rate for CIT:

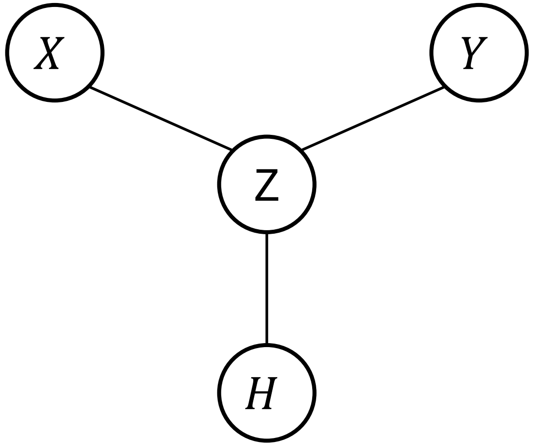
(c) Collider Structure

$X_t \to Y_t$	$Y_t \to Z_t$	I	II	III	IV	V	VI
Linear	Linear	100%	100%	5%	100%	100%	99%
Nonlinear	Linear	95%	89%	5%	96%	91%	51%
Linear	Nonlinear	90%	95%	5%	91%	96%	48%
Nonlinear	Nonlinear	81%	81%	6%	83%	81%	29%

PC algorithm:

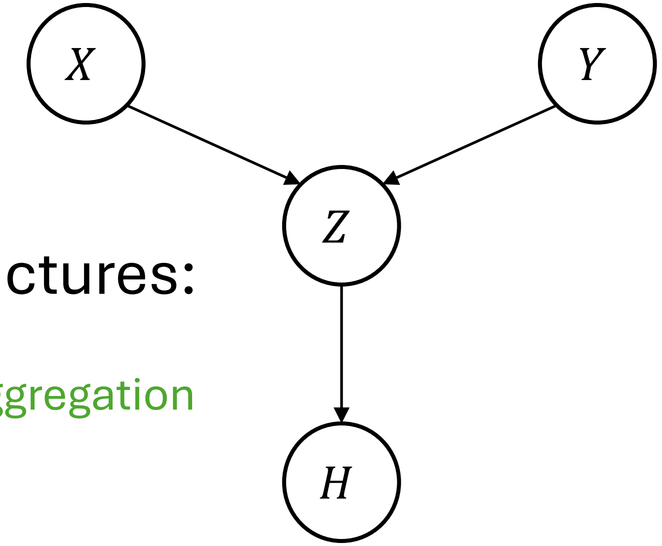
1.Skeleton Discovery:

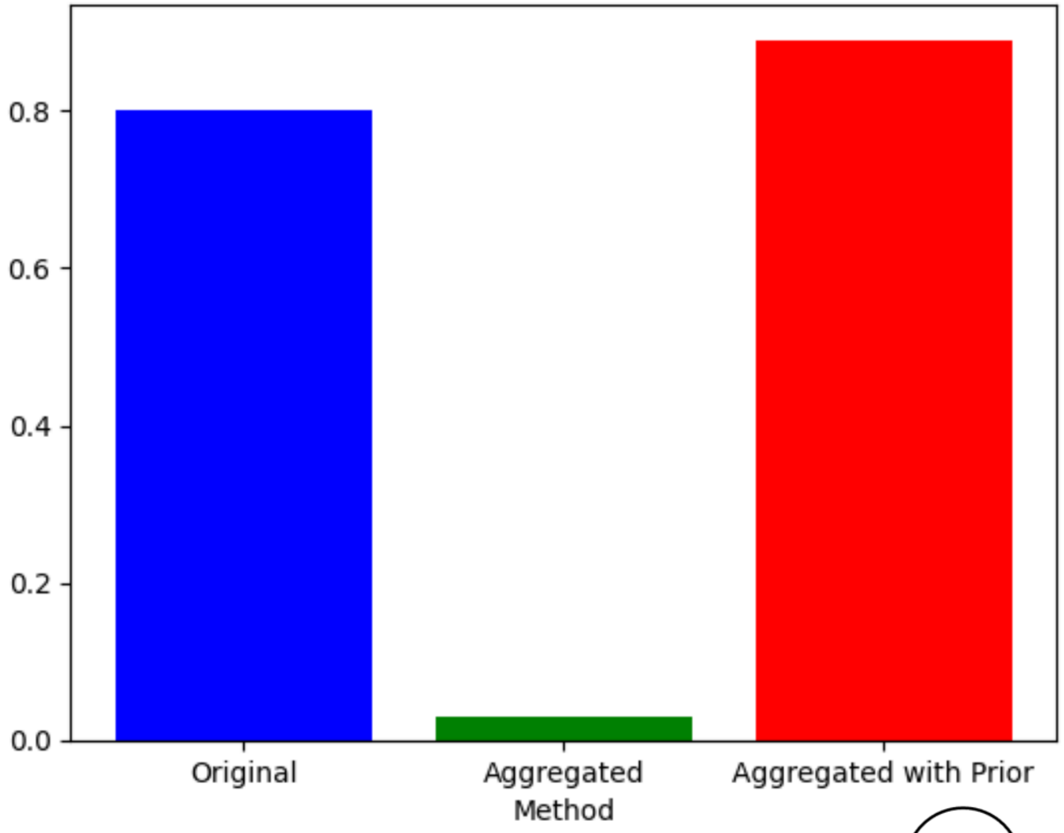
Suffer from aggregation

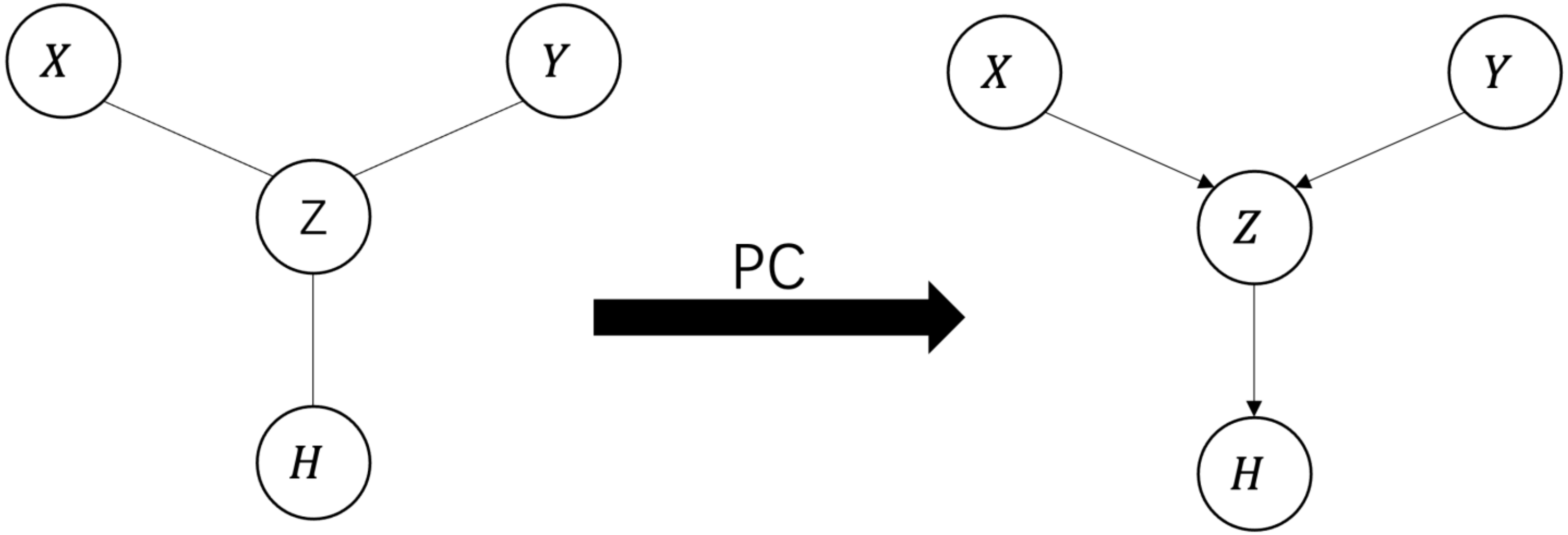


2.Finding V-structures:

Not affected by aggregation




0.8
0.6
0.4
0.2
0.0
Original
Aggregated Method
Aggregated with Prior


X
Y
Z
H
PC
X
Y
Z
H

Thanks for listening!