Necessary and Sufficient Conditions for Causal Feature Selection in Time Series with Latent Common Causes

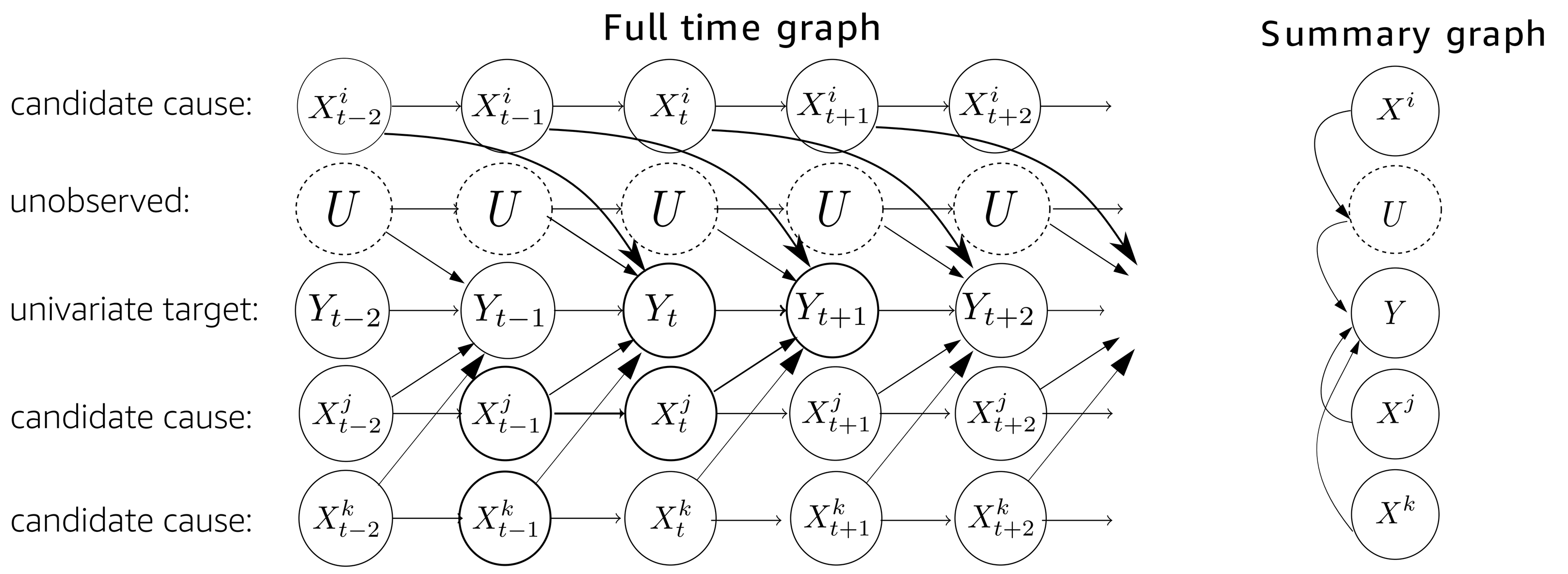
Atalanti A. Mastakouri[1], Bernhard Schölkopf[1,2], Dominik Janzing[1]

[1]Amazon Research, AWS Causality Team, Tübingen [2]Max Planck Institute for Intelligent Systems, Empirical Inference Department, Tübingen

Overview

- Given observed time series and a target time series, is it possible to identify its causes? Under which assumptions?
- **Goal:** Define **necessary and sufficient** conditions for causal feature selection in time series with latent common causes under some graph constraint.

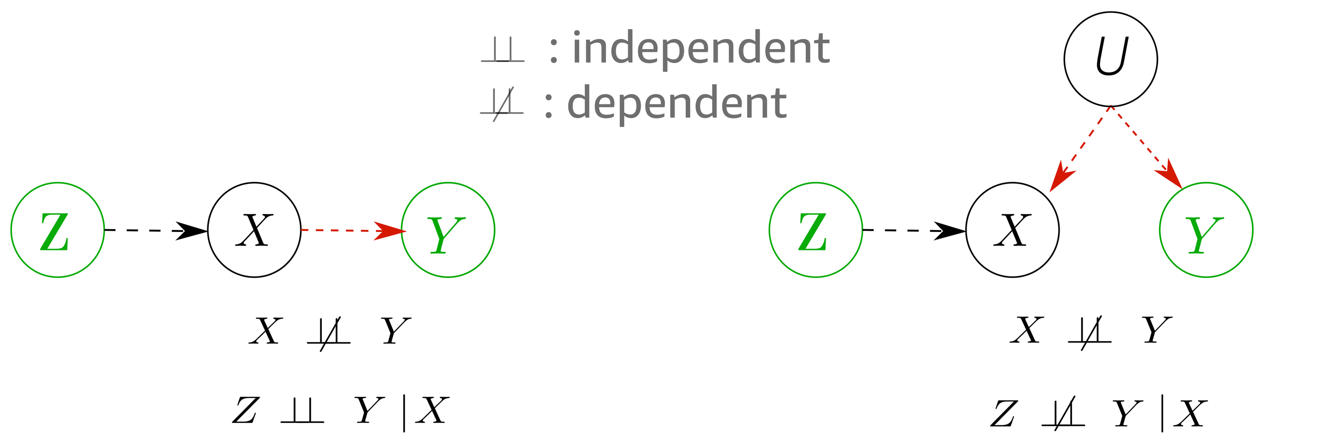
Formal problem description



Intuition

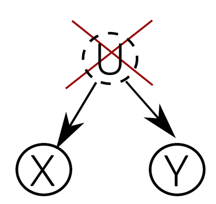
X causes Y

X and Y are confounded by U

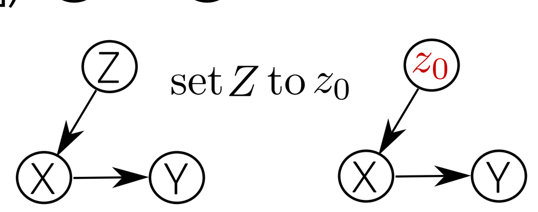


- The two graphs entail different patterns of conditional independences. [5]
- Finding different patterns of conditional independences in time series is more complex and requires conditioning on a larger **conditioning set**.

Issues of existing methods which we want to avoid



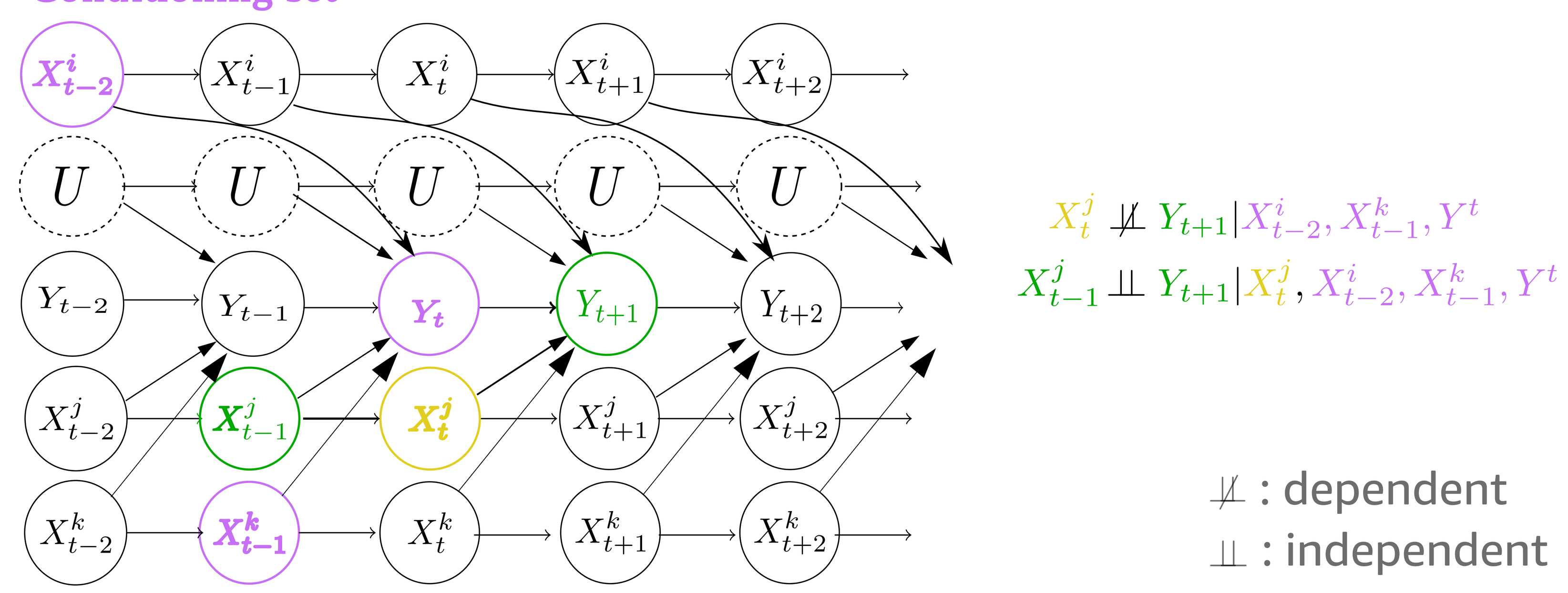
- **Exclude hidden common causes** (PCMCI [7], Granger [2])



- Need to **have interventions on the system** (seqICP [6])
- Require large conditioning sets (SVAR-FCI [4], tsFCI [1]) → **low statistical strength**
- Require exhaustive conditional independence tests → **low statistical strength**

Finding conditional independence patterns in time series

By conditioning on the purple nodes we find similar characteristic patterns of conditional independences which hold true only when X → Y



Theorem (Sufficient conditions for a direct or indirect sg-unconfounded cause of Y in single-lag dependency graphs)

Assuming A1-A9 (see paper) and single-lag dependency graphs, let w_i be the minimum lag between X^i and Y. Further, let $w_{ij} := w_i - w_j$. Then, for every time series $X^i \in \boldsymbol{X}$ we define a conditioning set $\boldsymbol{S^i} = \{X^1_{t+w_{i1}-1}, X^2_{t+w_{i2}-1}, ..., X^{i-1}_{t+w_{ij}-1}, X^{i+1}_{t+w_{ij}-1}, ..., X^n_{t+w_{in}-1}\}$.
If

$$X^i_t \not\perp\!\!\!\perp Y_{t+w_i} \mid \{\boldsymbol{S^i}, Y_{t+w_i-1}\} \tag{1}$$

and

$$X^i_{t-1} \perp\!\!\!\perp Y_{t+w_i} \mid \{\boldsymbol{S^i}, X^i_t, Y_{t+w_i-1}\} \tag{2}$$

are true, then

$$X^i_t \dashrightarrow Y_{t+w_i}$$

and the path between the two nodes is sg-unconfounded.

Theorem (Necessary conditions for a direct sg-unconfounded cause of Y in single-lag graphs)

Let the assumptions and the definitions of the previous Theorem hold.
If X^i_t is a direct, "sg-unconfounded" cause of Y_{t+w_i} ($X^i_t \rightarrow Y_{t+w_i}$), then cond. 1 and 2 of the previous Theorem hold.

Efficient conditioning set

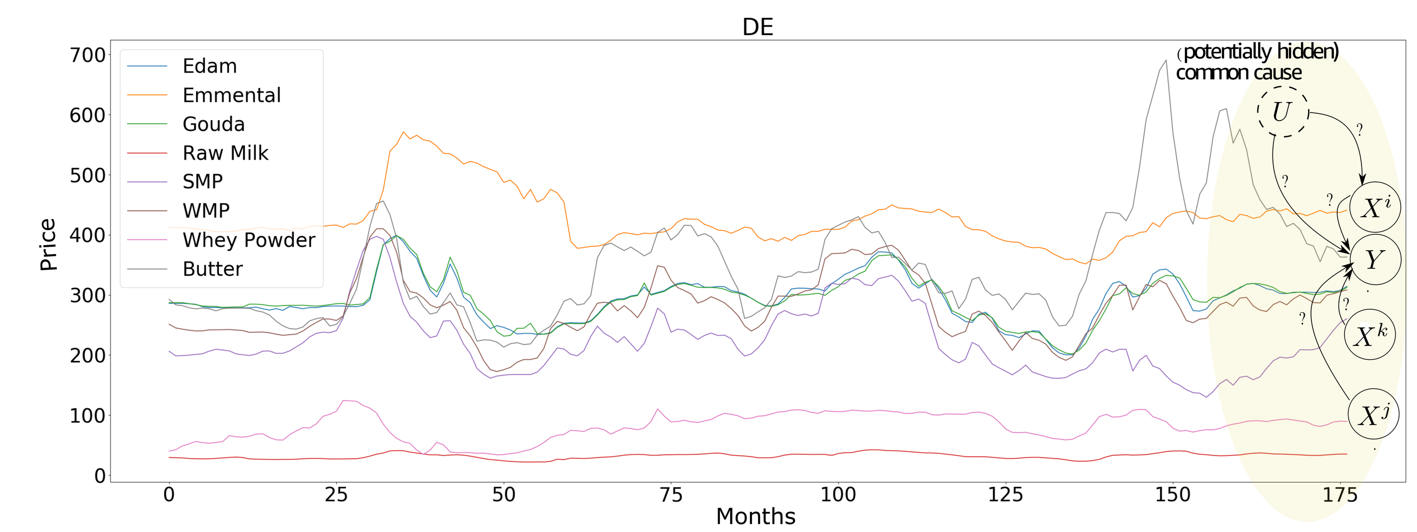
- The resulting conditioning set $\{\mathbf{S^i}, Y_{t+w_i-1}, X^i_t\}$ contains covariates that enter the outcome node Y_{t+w_i-1}, and not the potential cause X^i_{t-1}.
- Adjustment sets that include **parents of the potential cause** node are considered **inefficient**, as they can reduce the variance of the cause if they are strongly correlated with it, and thus reduce the signal [3].
- On the other hand, adding **nodes that explain variance in the outcome node** (precision variables) can contribute to a **better signal to noise ratio** for the dependences under consideration.
- Therefore, our choice of conditioning set could also **strengthen the statistical outcome.**

Algorithm

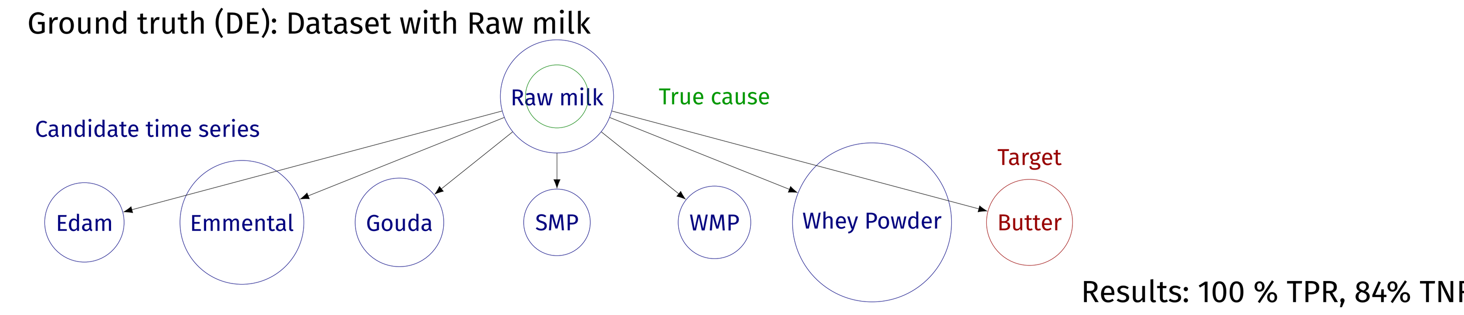
Input: $\mathbf{X}, Y$.
Output: causes_of_Y
n_{vars} = shape$(\mathbf{X}, 1)$; causes_of_Y$=[]$
$w = min\_lags(\mathbf{X}, Y)$
for $i = 1$ ***to*** n_{vars} **do**
  $\mathbf{S_i} = \bigcup_{j=1, j\neq i}^{n_{\text{vars}}} \{X^j_{t+w[i]-w[j]-1}\}$
  pvalue1 $= cond\_ind\_test(X^i_t, Y_{t+w[i]}, [\mathbf{S_i}, Y_{t+w[i]-1}])$
  if *pvalue1 < threshold1* **then**
    pvalue2 $= cond\_ind\_test(X^i_{t-1}, Y_{t+w[i]}, [\mathbf{S_i}, X^i_t, Y_{t+w[i]-1}])$
    if *pvalue2 > threshold2* **then**
      causes_of_Y = [causes_of_Y, X^i]
    end
  end
end

Algorithm 1: SyPI

SyPI on real world data

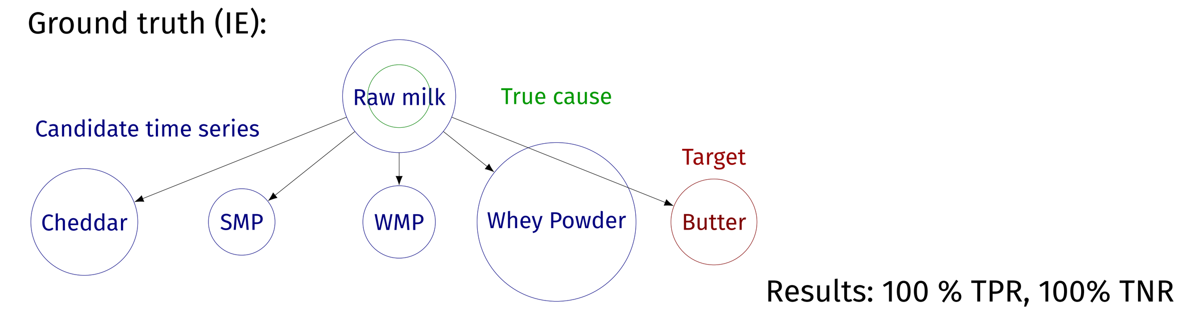


Ground truth (DE): Dataset with Raw milk



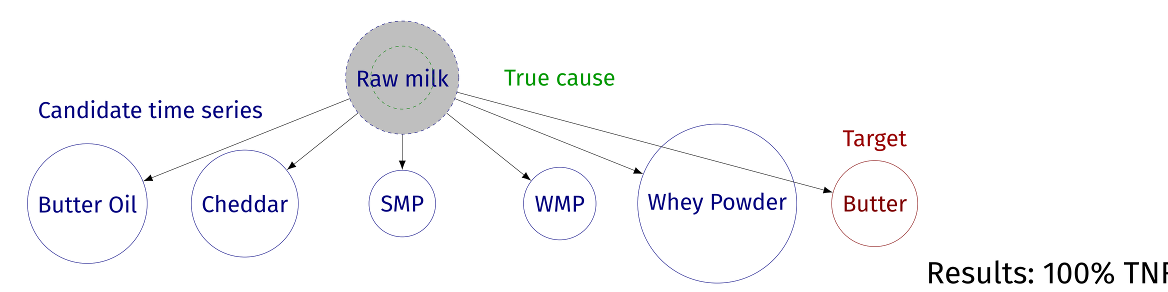
Results: 100 % TPR, 84% TNR

Ground truth (IE):



Results: 100 % TPR, 100% TNR

Ground truth (UK): Dataset without Raw milk



Results: 100% TNR

Conclusion

Findings

- Our method addresses one of the key problems of causal time series analysis: distinguish causal influence from confounding.
- Algorithm scales linearly with the number of time series
- Both necessary and sufficient conditions
- Successful application on simulated and real data
- Conditioning set efficient in terms of SNR

Future work

- Simplify some of the graph constraints
- Extend for multiple lags (not necessary conditions for multiple lags so far)

References

[1] D. Entner and P. O. Hoyer. On causal discovery from time series data using FCI. *Probabilistic graphical models*, pages 121–128, 2010.
[2] C. W. J. Granger. Investigating causal relations by econometric models and crossspectral methods. *Econometrica*, 37:424–438, 1969.
[3] L. Henckel, E. Perković, and M. H. Maathuis. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *arXiv*, 2019.
[4] D. Malinsky and P. Spirtes. Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of 2018 ACM SIGKDD Workshop on Causal Disocvery*, volume 92 of *Proceedings of Machine Learning Research*, pages 23–47, 2018.
[5] A. Mastakouri, B. Schölkopf, and D. Janzing. Selecting causal brain features with a single conditional independence test per feature. In *Advances in Neural Information Processing Systems 32*, 2019.
[6] N. Pfister, P. Bühlmann, and J. Peters. Invariant causal prediction for sequential data. *Journal of the American Statistical Association*, 114(527):1264–1276, 2019.
[7] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11):eaau4996, 2019.

atalanti@amazon.de