

基于模型性能指标变化的因果图修剪方法 及其在黄金价格数据中的应用

张 斌, 陈雪东, 陆欣怡

(湖州师范大学 理学院, 浙江 湖州 313000)

摘要:提出了一种基于模型性能指标变化的因果图修剪方法,用于非线性时间序列的因果发现.通过使用门限自回归模型对时间序列数据进行建模,并结合因果图修剪技术,评估父节点对模型预测性能的影响,逐步剔除不重要的点,从而简化因果图.实验结果表明,剪枝方法提高了模型的准确率、精确度和 F1 分数.在均方误差、平均绝对误差以及决定系数下进行剪枝处理,剪枝方法使 F1 分数显著提升,即使在最保守的 R2 指标下也从 0.34 提升至 0.43.在 MSE 指标下则提升至 0.70,与其他因果发现方法对比,MSE 剪枝方法在准确率、精确度和 F1 分数上均表现最佳,其中 F1 分数为 0.70.在黄金价格实证分析中,研究发现商品期货持仓百分比等变量的滞后项对黄金价格的因果效应.

关键词:因果发现; 剪枝; TAR 模型; 时间序列

中图分类号: Q 212.7

文献标志码: A

doi: 10.13763/j.cnki.jhebnu.nse.202601007

Evaluation-based Causal Graph Pruning and Causal Discovery of Gold Prices

ZHANG Bin, CHEN Xuedong, LU Xinyi

(School of Science, Huzhou Normal University, Huzhou Zhejiang 313000, China)

Abstract: This paper proposes a causal graph pruning method based on changes in model performance for nonlinear time series. Using a threshold autoregressive model, the method evaluates the impact of parent nodes on predictive accuracy and removes insignificant nodes to simplify the causal graph. Experiments show that pruning improves accuracy, precision, and F1 score. Under R2, F1 increases from 0.34 to 0.43, and under MSE, it reaches 0.70, outperforming other causal discovery methods. Empirical analysis of gold prices reveals that lagged variables, such as commodity futures positions, have causal effects on gold prices.

Key words: causal discovery; pruning; TAR model; time series

因果发现是一个从数据中确定变量之间稳定的因果关系的过程^[1].与相关性分析不同,因果发现关注的是变量 A 的变化是否能够“主动”导致变量 B 的变化,从而确定 A 和 B 之间是否存在因果关系.

随着数据科学和机器学习技术的飞速发展,因果发现在人工智能、生物医学和金融等领域的应用日益增多^[2-3].例如,金融领域中,因果发现可用于因子的挖掘和识别,分析经济政策,预测市场波动,评估政策的影响^[4].传统的因果推断方法,如 Granger 因果检验用于确定一个时间序列是否能够帮助预测另一个时间序列,其通常考虑的是两两之间的因果关系,适用的场景有较大的局限^[5].随着数据规模的增大和复杂性的发

收稿日期: 2025-06-18; 修回日期: 2025-12-23

基金项目: 教育部重点实验室开放课题(KLATASDS2008); 湖州师范学院研究生科研创新项目(2025KYCX51)

作者简介: 张 斌(1998-),男,山西晋中人,硕士研究生,研究方向为统计建模与因果推断.

通信作者: 陈雪东(1967-),男,教授,硕士生导师,研究方向为机器学习与金融统计. E-mail: xdchen@zjhu.edu.cn

展,许多新的因果发现方法被提出,NOTEARS 通过将无环约束转换为一个优化目标,将因果图学习问题转化为一个优化问题^[6].DYNOTEARS 将 NOTEARS 扩展到时间序列数据,能够捕捉时间维度上的因果关系^[7].

此外,VARLiNGAM 是线性非高斯无环模型(linear non-Gaussian acyclic model,LiNGAM)的拓展,专门面对非高斯分布数据中的因果发现^[8-9].尽管这些因果发现方法仍聚焦于线性模型,但在高维时间序列建模等场景中为因果推断提供了更为灵活和有效的工具.门限自回归(threshold autoregressive,TAR)模型作为一种非线性时间序列模型,通过引入门限值将数据划分为多个不同的状态,在每个状态下进行自回归建模,从而有效捕捉数据中的非线性特征^[10].TAR 模型适用于描述具有显著变化点或不同阶段的动态系统,如经济周期、股市波动等.然而,TAR 模型的参数估计可能会受到数据高维性和冗余特征的影响,导致模型复杂性较高,过拟合问题可能会更加突出.

为了解决这些问题,本文中,笔者提出了一种基于模型性能指标变化的因果图修剪方法,结合门限自回归模型对非线性时间序列数据进行建模,并通过评估每个节点对模型性能的贡献,逐步剔除不重要的节点,从而简化因果图的结构.该方法能够有效提高因果发现的准确性,并避免因果图过度复杂化带来的问题.通过实验验证,该方法在多个因果发现指标上取得了优异的表现,特别是在金融领域的黄金价格预测中,验证了滞后效应在黄金价格波动中的重要作用.

1 相关工作

格兰杰因果分析最早由经济学家克莱夫·格兰杰提出,用于确定一个时间序列是否能够帮助预测另一个时间序列^[5].因果关系体现在时间先后顺序上,如果一个变量的历史数据 X 能显著地帮助预测另一个变量 Y ,而 Y 的历史数据不能对 X 进行预测,则可以认为 X “Granger 导致” Y .为了处理同期时刻的因果关系,NOTEARS 通过将因果结构学习问题转化为一个连续的非凸优化问题,利用实数矩阵表示有向无环图,并引入平滑的无环性约束,将图结构的学习转化为一个平滑的优化过程^[6].通过最小化目标函数并同时满足无环性约束,NOTEARS 能够高效地求解因果结构,避免了传统离散搜索方法带来的高计算成本.DYNOTEARS 在此基础上进一步引入了即时(同一时间切片内)和滞后(跨时间切片)关系,其中即时关系要求满足无环性约束,而滞后关系则遵循时间的先后顺序^[7].通过这 2 个约束的结合,DYNOTEARS 在最小化目标函数的过程中进行因果发现.VARLiNGAM 是一种结合了向量自回归(vector autoregression,VAR)和 LiNGAM 的因果发现方法^[8].将 LiNGAM 模型扩展到时间序列数据,之后信息由 VAR 模型进行刻画,瞬时因果效应由 ICALiNGAM 模型进行刻画^[9].模型假设残差服从非高斯性,采用独立成分分析进行因果方向的识别.在此基础上,Nakanishi^[11]将 NG-SVAR 模型的误差分布假设由非高斯推广至 Pearson 分布族,并给出相应 R 语言实现,进一步放宽了 VARLiNGAM 的适用条件.Zheng 等^[12]则整合 LiNGAM 全系列、NOTEARS 及 PC 等主流算法,构建了开源 Python 平台 causal-learn,成为该领域广泛采用的标准实验基准.BigVAR 专门用于高维多变量时间序列的预测,其通过引入结构化的惩罚方法,结合了 Lasso、GroupLasso 等正则化技术,用于估计和选择适合多变量时间序列的稀疏系数^[13].BigVAR 框架采用 VARX-L 结构来处理时间序列的内生性和外生性变量之间的关系,通过加入针对内生系数和外生系数的惩罚来减少过拟合并提高模型的泛化能力.

上述因果发现方法均基于线性假设,为缓解这一局限性,TiMiNo 将每个时间序列的取值视为其直接原因与若干独立噪声共同作用的结果^[14].TiMiNo 模型首先尝试拟合时间序列数据以符合模型的假设,然后,通过检验残差的独立性来识别变量间的因果关系.PCMCI 在经典 PC 算法的基础上,结合条件独立性检验与时间延迟嵌入,有效提升了高维时间序列中的因果检测能力,并能较好地控制假阳性率^[15].神经网络格兰杰因果关系将神经网络与稀疏性惩罚方法相结合,检测非线性格兰杰因果关系^[16].通过将每个输出系列建模为单独的网络,并对输入的权重应用稀疏性惩罚,来提取因果结构.NOTEARS-MLP 解除了 NOTEARS 的线性限制,通过将每个变量的因果生成模型表示为 MLP,利用部分导数来表达变量之间的依赖关系,并通过计算梯度来判断节点之间的因果关系^[17].

2 模型

2.1 TAR 模型进行建模

为了对非线性时间序列数据进行因果发现,首先使用 TAR 模型.假设输入 p 个时间序列 $X^t = \{x_1, x_2, \dots, x_p\}$, 对于选取的目标 y^t , 每个 $x_i^{t-\tau} \in X^t \setminus \{y^t\}$ 是可能影响目标时间序列 y^t 的变量. TAR 可以抽象成为

$$y^t = f(\text{Pa}(y^t), \Gamma) + \epsilon^t.$$

其中, f 是在不同状态下用于描述时间序列的函数, ϵ^t 是误差项, Γ 表示门限值, 决定了何时进行状态切换, $\text{Pa}(\cdot)$ 表示父变量集合. 事实上, 当门限个数为 2 时, TAR 可以写为

$$y^t = \alpha_1 I(\text{Pa}(y^t) \leq \gamma) \text{Pa}(y^t) + \alpha_2 I(\text{Pa}(y^t) > \gamma) \text{Pa}(y^t) + \epsilon^t.$$

其中 I 表示示性函数.

2.2 因果图的构建

在建立 TAR 模型后, 即可构建一个因果图表示目标变量与父节点之间的因果关系. 假设目标变量 y^t 受到 k 个父节点 $x_1^{t_1}, x_2^{t_2}, \dots, x_k^{t_k}$ 的影响. 因果图 G 可以表示为

$$G = (x_1^{t_1}, x_2^{t_2}, \dots, x_k^{t_k}).$$

在建立 TAR 模型之前, 需要构建初始因果图以确定目标变量与候选父节点之间的潜在因果关系. 本研究采用全连接图策略, 具体步骤如下.

第一步 确定候选父节点集合

对于 p 个时间序列变量 $\{x_1, x_2, \dots, x_p\}$, 目标变量为 y^t , 考虑最大滞后阶数 L , 则候选父节点集合为

$$C = \{x_i^{t-\tau}; i = 1, 2, \dots, p, \tau = 1, 2, \dots, L\}.$$

候选集合共包含 $p \times L$ 个节点.

第二步 构建初始全连接因果图

基于候选父节点集合 C , 构建初始因果图 G_0 .

1) 节点集: $V_0 = C \cup \{x_i^t, i = 1, 2, \dots, p\}$;

2) 边集: $E_0 = \{x_i^{t-\tau} \rightarrow x_j^t; x_i^{t-\tau} \in C\}$.

第三步 TAR 建模

在初始因果图基础上, 使用 TAR 模型进行建模, 然后通过 2.3 的剪枝算法逐步移除对模型性能贡献不显著的节点, 得到最终精简的因果图.

2.3 剪枝

假设当前因果图为 $G = (x_1^{t_1}, x_2^{t_2}, \dots, x_k^{t_k})$, 希望通过评估每个父节点对模型性能的贡献, 来决定是否将其修剪掉. 首先计算当前因果图的 G 模型性能. 这个性能指标可以是任意的度量, 如均方误差 (mean squared error, MSE)、平均绝对误差 (mean absolute error, MAE)、对数损失等. 定义该性能指标 $L(G)$ 为当前模型的性能指标, 然后逐个删除因果图中的父节点 $x_i^{t_i}$, 并计算删除该父节点后的新因果图 $G \setminus x_i^{t_i}$ 的性能指标 $L(G \setminus x_i^{t_i})$. 这个过程通过去除一个父节点来评估其对因果图模型性能的影响. 在去除节点后, 因果图的结构发生了变化, 因此模型的性能也会发生改变. 删除父节点 $x_i^{t_i}$ 后的性能变化量可定义为删除前后性能指标的差异:

$$\Delta L = L(G) - L(G \setminus x_i^{t_i}).$$

该差异反映了父节点对模型性能的贡献. 如果删除某个父节点后, 性能变化量 ΔL 较小, 则可认为该父节点在因果图中的作用不显著. 因此, 可将其视为不重要的节点, 从而在因果图中将其修剪掉, 即 ΔL 小于某个设定的阈值, 则认为该父节点不重要, 可以被删除.

算法 1 基于模型性能指标变化的因果图修剪方法.

输入: 时间序列数据 $\{x_i, i = 1, 2, \dots, p\}$ 目标变量 y_t , 性能指标 (MSE, MAE, R2).

输出: 简化后的因果图 G .

1) 初始化 TAR 模型

使用 TAR 模型对时间序列数据进行建模构建初始因果图 G_0 ;

2) 评估初始因果图的性能

根据初始因果图 G_0 计算性能指标(MSE, MAE, R2)并使用性能指标评估模型的预测性能;

3) 迭代修剪父节点并评估性能变化

- ① for 每个父节点 x_i 在 G_0 中 do
- ② 移除父节点 x_i , 得到新的因果图 G_i
- ③ 使用新的因果图 G_i 训练模型
- ④ 计算移除该父节点前后的性能变化
- ⑤ if 性能变化小于预设阈值 δ then
- ⑥ 认为该父节点对模型贡献不大, 移除该父节点
- ⑦ end if
- ⑧ end for;

4) 得到简化后的因果图

经过修剪, 得到简化后的因果图 G_f , 计算准确率、精确度、召回率、F1 分数等指标, 评估修剪后因果图的分类能力和预测性能.

2.4 理论性质

基于模型性能指标变化的剪枝方法, 为保证因果关系的正确识别, 需满足以下 3 个条件.

条件 1(门限稳定性): 门限值在时间维度上应保持恒定, 即状态切换不随时间任意变化.

条件 2(状态内可识别性): 在每个状态 k 下, 父节点对模型性能的贡献应当可以被有效测量. 设 $S_k = \{t: k\}$ 为属于状态 k 的时间点集合, $\text{perf}(\text{Pa}_y | S_k)$ 表示使用父节点集合在状态 k 中的数据上训练 TAR 模型所得的性能指标(如 MSE、MAE 或 R2).

定义删除父节点 x_i 后在状态 k 中的性能变化:

$$\Delta_i^{(k)} = \text{perf}(\text{Pa}_y | S_k) - \text{perf}(\text{Pa}_y \setminus \{x_i\} | S_k),$$

其中 $\text{Pa}_y \setminus \{x_i\}$ 表示从父节点集合中移除 x_i 后的集合.

可识别性要求: 对于真实因果父节点 $x_i \in \text{Pa}_y^*$, 至少存在一个状态 k^* , 使得

$$E[\Delta_i^{(k^*)}] = \theta_i^{(k^*)} > 0,$$

即删除该节点后, 模型在该状态下的性能会显著下降(对于 MSE 和 MAE, 性能变化为正表示误差增加; 对于 R2, 需适当调整符号).

对于非因果节点 $x_i \notin \text{Pa}_y^*$, 在所有状态下, 对任意 k , 有

$$E[\Delta_i^{(k)}] = 0.$$

条件 3(跨状态一致性): 某父节点 x_i 应从因果图中剔除, 当且仅当其在所有状态下对模型性能的贡献都不显著. 形式化地, 设模型包含 k 个状态, 剪枝阈值为 $\tau > 0$, 则

$$x_i \notin \text{Pa}_y^* \Leftrightarrow \forall k \in \{1, 2, \dots, K\}, \Delta_i^{(k)} < \tau,$$

其中, Pa_y^* 表示 y 的真实因果父节点集合.

该条件体现了 TAR 模型允许某变量仅在特定状态下产生因果效应. 因此, 若变量在任一状态下对模型的性能贡献超过预设阈值, 则应将其保留在因果图中, 避免错误删除.

引理 1(性能变化估计的集中性) 在条件 2 下, 对于任意父节点 x_i 和状态 k , 性能变化的估计满足

$$\text{Var}[\hat{\Delta}_i^k] \leq \frac{4\sigma^2}{n_k}.$$

删除节点 x_i 前后, 2 个模型的性能之差 $\hat{\Delta}_i^{(k)}$ 是 2 个样本均值之差. 根据中心极限定理和独立性假设, 每个 MSE 估计的方差为 $O(\sigma^2/n_k)$, 因此差值的方差为 $O(\sigma^2/n_k)$, 具体常数为 4.

定理 1 设 Pa_y^* 为目标变量 y 的真实因果父节点集合, $\hat{\text{Pa}}_y$ 为算法 1 输出的父节点集合. 在满足条件 1~3 的前提下, 当每个状态 k 的样本量满足时, 基于性能指标变化的剪枝算法以概率至少 $1 - \delta$ 正确识别因果父节点集合:

$$n_k \geq C \cdot p \cdot \ln\left(\frac{2p}{\delta}\right),$$

$$P(\hat{Pa}_y = Pa_y^*) \geq 1 - \delta.$$

其中, $p = |Pa_y|$ 为候选父节点总数, $\delta \in (0, 1)$ 为容许的错误概率, C 为依赖于噪音方差 σ^2 和最小因果效应 θ_{\min} 的常数.

进一步, 错误概率关于样本量呈指数衰减:

$$\delta = O(\exp(-\frac{c \cdot n_{\min}}{p})),$$

其中 $n_{\min} = \min_k n_k$, $c > 0$ 为常数.

证 记 $\text{perf}(\cdot)$ 为性能指标函数(如 MSE、MAE), $\Delta_i^{(k)} = \text{perf}(Pa_y | S_k) - \text{perf}(Pa_y \setminus \{x_i\} | S_k)$ 为删除 x_i 在状态 k 的真实性能变化, $\hat{\Delta}_i^{(k)}$ 为基于有限样本的性能变化估计值, $\tau > 0$ 为剪枝阈值, σ^2 为模型预测误差的方差, $\theta_{\min} = \min\{E[\Delta_i^{(k)}]: x_i \in Pa_y^*, E[\Delta_i^{(k)}] > 0\}$ 为最小因果效应.

证明分为 2 部分: 假阳性控制和假阴性控制.

I) 控制假阳性率(避免错误保留非因果节点)

对于非因果节点 $x_i \notin Pa_y^*$, 根据条件 2, 对任意 k , $E[\Delta_i^{(k)}] = 0$, 即删除该节点不影响真实预测性能. 但由于有限样本的估计误差, 观测到的 $\hat{\Delta}_i^{(k)}$ 可能偏离 0.

应用 Chebyshev 不等式, 得

$$P(|\hat{\Delta}_i^{(k)} - 0| > \tau) \leq \frac{\text{Var}[\hat{\Delta}_i^{(k)}]}{\tau^2}.$$

由引理 1, $P(|\hat{\Delta}_i^{(k)}| > \tau) \leq \frac{4\sigma^2}{n_k \tau^2}$. 为使该概率足够小, 取 $\tau = 2\sigma \sqrt{\frac{\ln(2p/\delta)}{n_k}}$, 则

$$P(|\hat{\Delta}_i^{(k)}| > \tau) \leq \frac{4\sigma^2}{n_k \cdot 4\sigma^2 \cdot \frac{\ln(2p/\delta)}{n_k}} \leq \frac{1}{\ln(2p/\delta)} \leq \frac{\delta}{2p} \text{ (当 } \delta \text{ 足够小时成立)}.$$

节点 x_i 被错误保留当且仅当至少存在一个状态 k 使得 $|\hat{\Delta}_i^{(k)}| > \tau$. 对于 K 个状态, $P(\exists k: |\hat{\Delta}_i^{(k)}| > \tau) \leq \sum_{k=1}^K P(|\hat{\Delta}_i^{(k)}| > \tau) \leq K \cdot \frac{\delta}{2p}$. 跨所有非因果节点, 设有 $p - p^*$ 个非因果节点, 总假阳性错误的概率

$$P(\text{错误保留某个非因果节点}) \leq (p - p^*) \cdot K \cdot \frac{\delta}{2p} \leq \frac{K\delta}{2}.$$

对于 2 个状态 TAR 模型, $P(\text{假阳性}) \leq \delta$.

II) 控制假阴性率(避免错误删除真因果节点)

对于真因果节点 $x_i \in Pa_y^*$, 根据条件 2, 至少存在一个状态 k^* 使得 $E[\Delta_i^{(k^*)}] = \theta_i^{(k^*)} > 0$, 即删除该节点会降低该状态下的模型性能.

设所有真因果节点在其有效状态下的最小性能变化为

$$\theta_{\min} = \min_{x_i \in Pa_y^*} \{E[\Delta_i^{(k)}]: E[\Delta_i^{(k)}] > 0\}.$$

为保证真因果节点不被删除, 阈值应满足 $\tau < \theta_{\min}$. 实际中取 $\tau = \frac{\theta_{\min}}{2}$.

在状态 k^* 中, 节点 x_i 被错误判定为不重要的概率为

$$P(\hat{\Delta}_i^{(k^*)} < \tau) = P(\hat{\Delta}_i^{(k^*)} < \frac{\theta_{\min}}{2}).$$

由于 $E[\Delta_i^{(k^*)}] = \theta_i^{(k^*)} \geq \theta_{\min}$, 有

$$P(\hat{\Delta}_i^{(k^*)} < \frac{\theta_{\min}}{2}) = P(\hat{\Delta}_i^{(k^*)} - \theta_i^{(k^*)} < \frac{\theta_{\min}}{2} - \theta_i^{(k^*)}) \leq$$

$$P(\hat{\Delta}_i^{(k^*)} - \theta_i^{(k^*)} < -\frac{\theta_i^{(k^*)}}{2}) \leq P(|\hat{\Delta}_i^{(k^*)} - \theta_i^{(k^*)}| > \frac{\theta_i^{(k^*)}}{2}) \leq$$

$$\frac{4\text{Var}[\hat{\Delta}_i^{(k^*)}]}{\theta_i^{(k^*)}} \text{(切比雪夫不等式)} \leq \frac{16\sigma^2}{n_k \cdot (\theta_i^{(k^*)})^2} \text{(引理 1)}.$$

由于 $\theta_i^k \geq \theta_{\min}$, $P(\hat{\Delta}_i^{(k^*)} < \tau) \leq \frac{16\sigma^2}{n_k \cdot \theta_{\min}^2}$. 节点 x_i 被删除当且仅当在所有状态下都有 $\hat{\Delta}_i^{(k)} < \tau$. 但由于 x_i 在状态 k^* 中有显著效应, 只要在该状态中正确检测到, 就不会被删除.

$$P(x_i \text{ 被错误删除}) = P(\text{在所有状态下都} < \tau) \leq P(\hat{\Delta}_i^{(k^*)} < \tau) \leq \frac{16\sigma^2}{n_k \cdot \theta_{\min}^2}.$$

当样本量满足 $n_{k^*} \geq \frac{32\sigma^2}{\theta_{\min}^2} \cdot \frac{p}{\delta}$ 时,

$$P(x_i \text{ 被错误删除}) \leq \frac{\delta}{2p}.$$

跨所有真因果节点, 设有 p^* 个真因果节点, 总假阴性错误的概率为 $P \leq p^* \cdot \frac{\delta}{2p} \leq \frac{\delta}{2}$.

III) 总错误率

定义事件 E_1 : 至少一个非因果节点被错误保留(假阳性). E_2 : 至少一个真因果节点被错误删除(假阴性). E : 算法输出错误, 即 $\hat{P}_{a_y} \neq P_{a_y}$.

显然, $E \subseteq E_1 \cup E_2$, 因此, $P(E) \leq P(E_1) + P(E_2)$. 由(1)和(2), 得

$$P(\hat{P}_{a_y} \neq P_{a_y}) \leq \delta + \frac{\delta}{2} = \frac{3\delta}{2}.$$

为使总错误率不超过 δ , 可适当调整各部分的容许错误(例如将每部分的 δ 替换为 $\delta/3$). 综合假阳性和假阴性的要求, 当每个状态的样本量满足 $n_k \geq C \cdot p \cdot \ln(\frac{2p}{\delta})$, 其中, $C = \max\{4, \frac{32\delta^2}{\theta_{\min}^2}\}$ 时, 定理成立.

IV) 指数衰减率

使用 Hoeffding 不等式进行更精细的分析. 对于均值为 μ 、方差有界的随机变量 Z , 其样本均值 \bar{Z} 满足

$$P(|\bar{Z} - \mu| > \epsilon) \leq 2\exp(-\frac{2n\epsilon^2}{(b-a)^2}).$$

其中, a, b 为随机变量的上下界.

对于性能变化 $\hat{\Delta}_i^{(k)}$, 应用 Hoeffding 不等式, $P(|\hat{\Delta}_i^{(k)} - E[\hat{\Delta}_i^{(k)}]| > \epsilon) \leq 2\exp(-\frac{c \cdot n \cdot \epsilon^2}{\delta^2})$, 其中 c 为常数.

对于非因果节点 $E[\hat{\Delta}_i^{(k)}] = 0$, 取 $\epsilon = \tau$, $P(|\hat{\Delta}_i^{(k)}| > \tau) \leq 2\exp(-\frac{cn_k\tau^2}{\sigma^2})$.

对于真因果节点, 取

$$\epsilon = \theta_i^{(k^*)} - \tau = \theta_{\min}/2,$$

$$P(|\hat{\Delta}_i^{(k)}| > \tau) \leq 2\exp(-\frac{cn_k\theta_{\min}^2}{4\sigma^2}).$$

总错误率可界定为 $\delta \leq 4pK \cdot \exp(-\frac{c'n_{\min}}{p})$, 其中 $c' = \min\{\frac{c\tau^2}{\sigma^2}, \frac{c\theta_{\min}^2}{4\sigma^2}\}$, $n_{\min} = \min_k n_k$. 这表明错误概率

δ 关于最小状态样本量 n_{\min} 呈指数衰减, $\delta = O(\exp(-\frac{c \cdot n_{\min}}{p}))$.

3 数值模拟

3.1 实验场景设计

本实验包含 5 个变量 x_1, x_2, \dots, x_5 的时间序列模型, 按照如下模型生成 2 000 条数据.

$$\begin{cases} x_1^t = x_1^{t-1} + \epsilon_1^t, \\ x_2^t = 0.5(x_1^{t-1})^2 + 0.3\sin(0.1t) + 0.2x_2^{t-1} + \epsilon_2^t, \\ x_3^t = 0.7\ln(1 + x_2^{t-1}) + 0.2x_1^{t-2} + \epsilon_3^t, \\ x_4^t = -0.5(x_3^{t-1})^2 + 0.4x_1^{t-2} + 0.3\cos(x_2^{t-1}) + \epsilon_4^t, \\ x_5^t = \begin{cases} 0.6x_4^{t-1} + 0.2(x_2^{t-2})^2 + 0.1x_1^{t-3} + \epsilon_5^t, & x_3^{t-1} > 0, \\ -0.3(x_4^{t-1})^2 + 0.5x_2^{t-1} + \epsilon_5^t, & x_3^{t-1} \leq 0. \end{cases} \end{cases}$$

实验使用了 3 种度量指标:均方误差(MSE)、平均绝对误差(MAE)和 R2(决定系数)。其中,MSE 衡量的是预测值与实际值之间差异的平方的平均值,其对较大误差赋予更高的权重,因此特别适用于对大误差敏感的场景。MAE 则通过计算预测误差的绝对值的平均值来衡量模型的预测精度,其不如 MSE 对大误差敏感,因此在处理异常值时更为稳健。R2 则衡量模型对数据变异性的解释能力,值越接近 1 表示模型对数据的拟合越好,能够更有效地捕捉数据中的变化趋势。通过对上述 3 种指标的综合评估,可从多个维度全面分析模型性能,从而更准确地判断模型优劣,并为后续模型优化提供依据。

3.2 实验结果

实验结果基于不同的剪枝方法和因果发现方法进行分析。使用准确率(accuracy)、精确度(precision)、召回率(recall)及 F1 分数(F1 score)来验证模型的好坏。其中,准确率表示模型预测正确的样本占总样本的比例。

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

精确度衡量的是模型预测为正类的样本中有多少实际上是正类。即在所有被预测为正类的样本中,真正正类的比例。

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

召回率衡量的是所有真正正类样本中,有多少被模型正确预测为正类。召回率高表示模型能够识别更多的正类样本,减少假阴性,但可能会增加假阳性的风险。

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1 分数是精确度和召回率的调和平均值,综合了模型的准确性和召回能力。其特别适用于数据不平衡的情况,可以平衡模型对正类样本的识别能力和精度。F1 分数越高,表示模型在精确度和召回率之间的平衡越好。

$$\text{F1 score} = 2 \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

根据不同剪枝方法下的结果分析(表 1),剪枝显著改善了模型的准确率和精确度,尤其是在 MSE 和 MAE 指标下。未进行剪枝时,模型的准确率仅为 0.53,而在剪枝后,准确率分别提高至 0.92、0.91 和 0.89,这表明剪枝方法能够显著提升模型的分类准确性。此外,精确度也得到了显著提高,未剪枝时的精确度为 0.21,而剪枝后的精确度在 MSE 和 MAE 指标下分别为 0.70 和 0.67,在 R2 指标下为 0.75,表现最好。这表明剪枝能有效减少假阳性的发生,从而提高模型对正类样本的识别能力。

表 1 不同指标下剪枝效果对比

Tab.1 Comparison of Pruning Effects Under Different Metrics

指标	未剪枝	MSE	MAE	R2
准确率	0.53	0.92	0.91	0.89
精确度	0.21	0.70	0.67	0.75
召回率	0.90	0.70	0.60	0.30
F1 分数	0.34	0.70	0.63	0.43

然而,对于召回率而言,未剪枝时为 0.90.这意味着未剪枝时模型可能会对每个变量都做出较为宽松的预测,导致较高的召回率.在这种情况下,模型更多地倾向于预测正类样本,即使这些预测可能并不完全准确,从而提高了召回率.

此外,剪枝后,F1 分数也有所提升.未剪枝时的 F1 分数为 0.34,而在剪枝后,F1 分数在 MSE 和 MAE 指标下分别提升至 0.70 和 0.63,但在 R2 指标下仅为 0.43,尽管有所提高,但与其他剪枝方法相比,表现仍显得较弱.这表明剪枝可以在一定程度上改善模型的平衡性,但不同剪枝方法的效果存在差异,尤其是在 R2 指标下,剪枝后的平衡性较差.

根据不同因果发现方法的对比结果(表 2),MSE 剪枝方法在准确率、精确度和 F1 分数上均表现最佳,准确率达到 0.92,精确度为 0.70,F1 分数为 0.70,显示出其在分类任务中的卓越性能,能够有效地平衡精确度和召回率.相比之下,PCMCi 方法在召回率方面表现最佳,达到 1.00,但其精确度较低,仅为 0.40,导致 F1 分数为 0.57,表明该方法可能存在较高的假阳性率.VARLiNGAM 的各项指标均较低,尤其是在准确率和召回率上,表现相对较差.TCDF 方法的表现介于 MSE 剪枝和 PCMCi 之间,准确率和精确度较好,但召回率较低,F1 分数为 0.50.综上所述,MSE 剪枝在所有方法中表现最为均衡,尤其适合需要高准确率和较好平衡性的任务.

表 2 本文方法与其他因果发现方法的性能对比

Tab.2 Performance Comparison of Different Causal Discovery Methods

指标	MSE 剪枝	PCMCi	VARLiNGAM	TCDF
准确率	0.92	0.40	0.12	0.68
精确度	0.70	0.40	0.17	0.67
召回率	0.70	1.00	0.30	0.40
F1 分数	0.70	0.57	0.21	0.50

3.3 稳健性与敏感性分析

为全面评估本方法的稳健性,依次做了如下 3 组实验,分别考察阈值选择、样本量和模型准确率在不同随机数种子下的结果,所有实验基于 2.1 的合成数据模型.

3.3.1 阈值参数的敏感性分析

图 1 展示了在不同阈值下模型性能的变化.具体来说,图 1a 展示了随着阈值的调整,模型的性能指标(如 F1 分数、精确度等)随阈值变化的趋势.可以观察到,绿色部分对应的阈值范围内,模型性能保持较为稳定,表明在这一范围内,模型对剪枝阈值的敏感性较低.因此,选择该范围内的一个值作为固定值能确保模型性能的稳定性,从而有效提升因果图修剪方法的可靠性.图 1b 进一步展示了不同阈值下因果发现过程中保留的边数.可以看到,在选定的稳定阈值区间内,保留的因果边数变化较小,表明在该区间内,剪枝阈值对因果图的结构影响较小.这表明,选择该阈值可以有效简化因果图,同时保证图中的因果关系得到合理保留,从而提高模型的可解释性和预测性能.

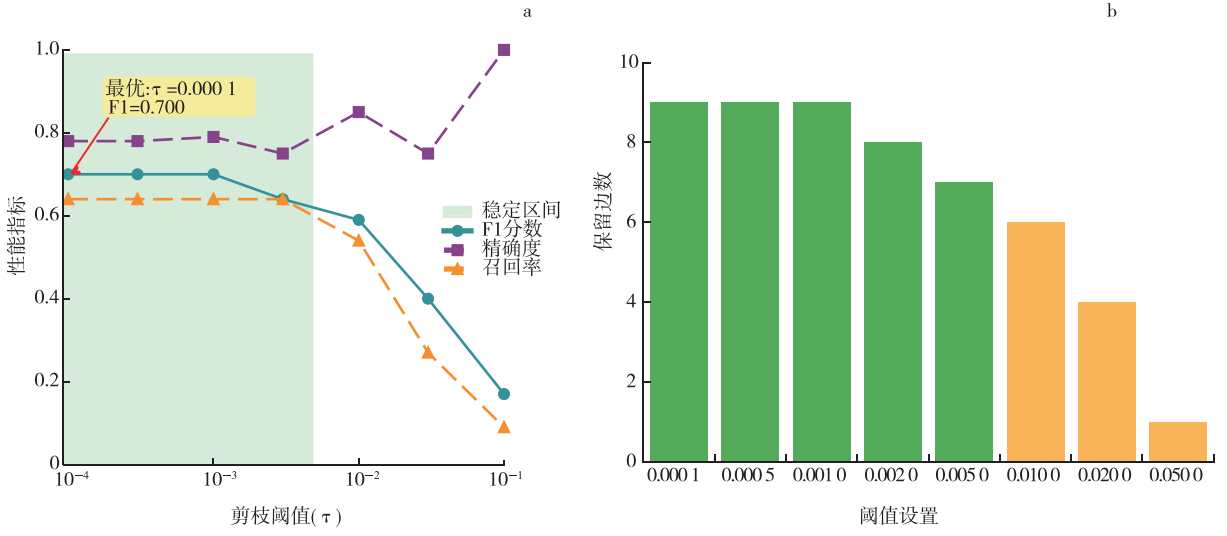
3.3.2 样本量影响分析

图 2 展示了样本量对模型性能的综合影响.可以观察到,随着样本量的增加,模型的各项性能指标(如 F1 分数、精确度、召回率和准确率等)呈现出明显的变化.具体来说,适当增加样本量有助于提升模型的分类能力和预测精度,尤其在样本量达到一定程度时,性能逐渐稳定.然而,过小的样本量可能导致模型效果不稳定,无法有效捕捉数据中的规律;而样本量过大时,模型的性能提升趋于平稳,过多的样本量可能引入计算负担和冗余信息.因此,选择合适的样本量范围对于提高模型的稳定性和可靠性至关重要.

3.3.3 稳定性检验

为了验证方法的稳定性,使用不同的随机种子进行了 20 次独立运行.图 3 展示了准确率在 20 次运行中的表现.结果显示,准确率的均值为 0.802,标准差为 0.040,变异系数仅为 5.0%,表明方法具有良好的稳定性.20 次运行中,准确率均保持在 0.696~0.856 之间,且大部分结果集中在均值加减一倍标准差范围内(黄

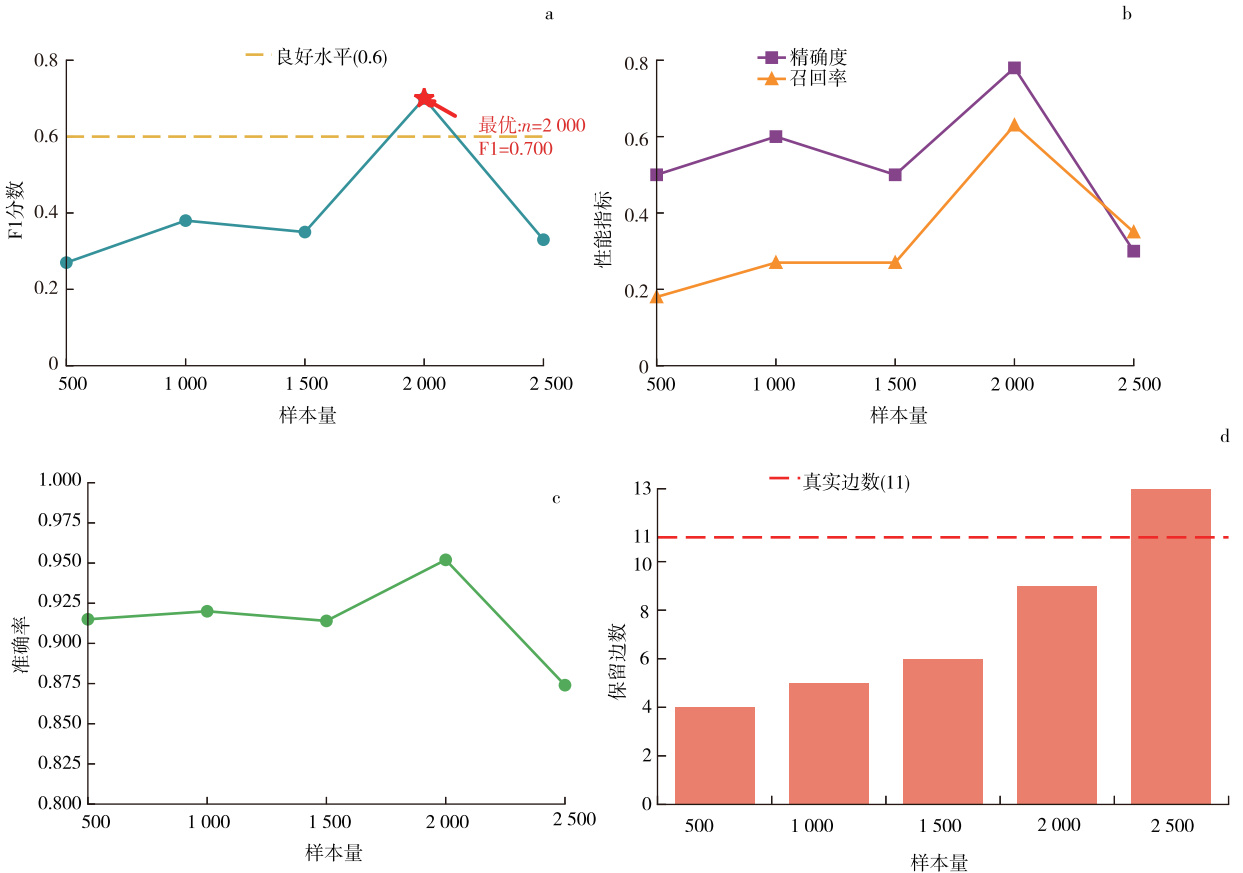
色阴影区域), 未出现极端异常值. 准确率的高稳定性主要得益于方法对非因果边的可靠判断, 这些判断不易受随机波动影响. 这一结果证明了本研究在不同数据实现下都能保持较高的整体正确性, 具有良好的可靠性和实用性.



a. 不同阈值下的性能表现; b. 不同阈值下的图稀疏性.

图 1 阈值参数的敏感性分析图

Fig.1 Sensitivity Analysis of the Threshold Parameter



a. 样本量对 F1 分数的影响; b. 样本量对精确度和召回率的影响; c. 样本量对准确率的影响; d. 样本量对识别边数的影响.

图 2 样本量对方法性能的影响分析图

Fig.2 Effect of Sample Size on Method Performance

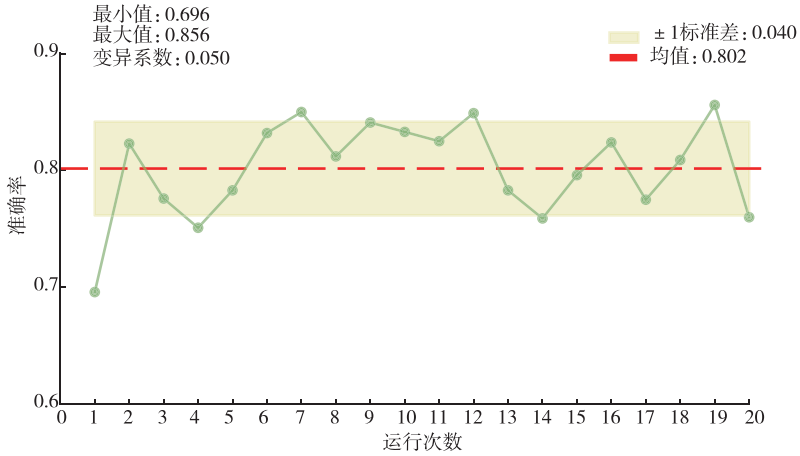


图3 20次独立运行中的准确率稳定性

Fig.3 Accuracy Stability over 20 Independent Runs

4 黄金期货交易价格的归因分析

4.1 问题与数据

以黄金价格影响因素的因果分析为案例展开.黄金作为一种重要的贵金属资产,其价格波动受多方面因素的影响,涉及到宏观经济、市场需求、金融政策等多个层面.为了深入理解这些因素,经过对文献和实际需求的调研,选择7个影响黄金价格的因子.表3显示了用于分析黄金价格波动的7个特征,表格中的特征包括商品期货市场的净仓位、债券收益率与黄金价格的关系、经济政策不确定性指数与黄金价格的关联等.本文将最大滞后阶数设定为 $L=2$,即每个变量考虑1阶、和2阶滞后项,共 $7 \times 2=14$ 个候选父节点.该设定基于以下考虑:1) 文献依据,黄金价格建模研究普遍采用1~4阶滞后^[18-19],短期滞后能够有效捕捉价格动态特征;2) 数据匹配,本文使用周频数据,2阶滞后对应约2周时间,符合金融市场短期信息传导的时间尺度;3) 复杂度控制, $L=2$ 的设定在捕捉动态关系与避免过拟合之间取得平衡.数据选择的起止时间为2000年02月27日至2024年02月18日,在数据清洗和整理阶段,首先进行了数据的预处理工作,包括去除异常值、填补缺失值以及统一数据的频率.为了确保各个变量能够在同一时间点对齐,所有指标的时间频率均进行了统一,确保了不同数据源的可比性和分析的一致性.为方便处理,在后续的分析中以 y 表示黄金价格. $x_i, i=1,2,\dots,7$, 表示7个指标.

表3 符号说明

Tab.3 Description of Notations

特征	说明	代号
$\text{diff1}_{\{\text{Comm_Net_Pct-088691-FutureOptions-Report}\}}$	商品期货市场净仓位百分比的一阶差分	x_1
$\text{log_diff1}_{\{\text{3month-bond-real-yield-gold-price-Gold Futures Price (R)}\}}$	3个月期债券实际收益率与黄金价格及黄金期货价格之间的对数差分	x_2
$\{\text{Comm_Position_Pct-098662-FutureOptions-Report}\}_{\text{lag2}}$	商品期货持仓百分比的对数差分滞后2期	x_3
$\text{log_diff1}_{\{\text{Comm_Position_Pct-088691-FutureOptions-Report}\}_{\text{lag2}}}$	商品期货持仓百分比的对数差分滞后2期	x_4
$\text{log_diff1}_{\{\text{world-economic-policy-uncertainty-index-vs-gold-price-US - Economic Policy Uncertainty Index (MA30)}\}_{\text{lag2}}}$	全球经济政策不确定性指数与黄金价格之间对数差分滞后2期	x_5
$\text{log_diff1}_{\{\text{3month-bond-real-yield-gold-price-Gold Futures Price (R)}\}_{\text{lag4}}}$	3个月期债券实际收益率与黄金价格及黄金期货价格的对数差分滞后4期	x_6
$\text{log_diff1}_{\{\text{low-GC-price}\}_{\text{lag6}}}$	低价黄金期货的一阶对数差分滞后6期	x_7

4.2 结果与讨论

图 4 为因果发现结果,线条粗细代表因果强度,从图中可以看出,目标变量 y (黄金价格)与多个变量之间存在显著的因果关系.首先, x_1 、 x_3 、 x_4 、 x_5 、 x_6 和 x_7 的一阶滞后项与黄金价格 y 存在因果关系,这些变量对黄金价格的变化产生了较为直接的影响.除此之外, x_3 的二阶滞后项也与黄金价格 y 存在因果关系.因此,时间序列数据中的滞后效应在分析黄金价格的动态变化时不可忽视.

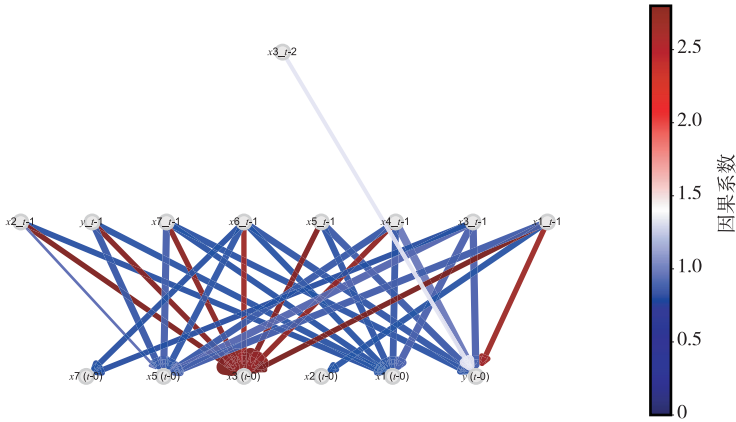


图 4 因果图

Fig.4 Estimated Causal Graph

5 结论

本文中,笔者提出了一种基于模型性能指标变化的因果图修剪方法,结合门限自回归模型,用于非线性时间序列数据的因果发现.通过合成数据与黄金价格实际数据的实验验证,该方法在简化因果图结构的同时显著提升了模型的预测性能.剪枝后模型的准确率从 0.53 提升至 0.92, F1 分数从 0.34 提升至 0.70 (基于 MSE 指标),表明其能有效剔除冗余节点并增强因果推断的鲁棒性.相较于 PCMCI、VARLiNGAM 和 TCDF 等方法, MSE 剪枝在准确率(0.92)、精确度(0.70)和 F1 分数(0.70)上均表现最优,展现了其在平衡模型复杂度与解释性方面的优势.在黄金价格影响因子的实证分析中,研究揭示了商品期货持仓百分比、经济政策不确定性指数及债券收益率等滞后变量对黄金价格的显著因果效应,验证了金融时间序列中滞后动态的复杂性.未来工作将进一步探索多变量交互场景下的因果推断,优化门限机制以适应更高维非线性模式,并结合深度学习框架,拓展该方法在复杂金融系统中的应用.

参考文献:

- [1] PEARL J. Causality: Models, Reasoning, and Inference[M]. 2nd ed. Cambridge: Cambridge University Press, 2009: 1-450.
- [2] SCHÖLKOPF B. Causality for Machine Learning[C]// GEFNER H, DECHTER R, HALPERN J Y. Probabilistic and Causal Inference: The works of Judea Pearl. New York: ACM, 2022: 765-804. doi: 10.1145/3501714.35017
- [3] SCHUCH H S, NASCIMENTO G G, DEMARCO F F, et al. Causal Inference in Dentistry: Time to Move Forward[J]. Community Dentistry and Oral Epidemiology, 2023, 51(1): 62-66. doi: 10.1111/cdoe.12802
- [4] LOPEZ-PAZ D, NISHIHARA R, CHINTALA S, et al. Discovering Causal Signals in Images[C]// IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 58-66. doi: 10.1109/CVPR.2017.14
- [5] GRANGER C W J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods[J]. Econometrica: Journal of the Econometric Society, 1969, 37(3): 424-438. doi: 10.2307/1912791
- [6] ZHENG X, ARAGAM B, RAVIKUMAR P K, et al. DAGs with NO TEARS: Continuous Optimization for Structure Learning[C]// BENGIO S, WALLACH H M. Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2018: 9492-9503.
- [7] PAMFIL R, SRIWATTANAWORACHAI N, DESAI S A, et al. DYNOTEARS: Structure Learning from Time-series Data[C]// SILVIA C, ROBERTO C. Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics. Cambridge: PMLR, 2020: 1595-1605.

- [8] HYVÄRINEN A, ZHANG K, SHIMIZU S, et al. Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity[J]. *Journal of Machine Learning Research*, 2010, 11(56):1709-1731.
- [9] SHIMIZU S, HOYER P O, HYVÄRINEN A, et al. A Linear Non-Gaussian Acyclic Model for Causal Discovery[J]. *Journal of Machine Learning Research*, 2006, 7(4):2003-2030.
- [10] SUN Y Y, HAN A, HONG Y M, et al. Threshold Autoregressive Models for Interval-valued Time Series Data[J]. *Journal of Econometrics*, 2018, 206(2):414-446. doi: 10.1016/j.jeconom.2018.06.009
- [11] NAKANISHI T. The NG-SVAR Model Under the Pearson Family of Distributions: Implementation with R Packages[J]. *Journal of International Economic Studies*, 2024(38):35-46.
- [12] ZHENG Y J, HUANG B W, CHEN W, et al. Causal-learn: Causal Discovery in Python[J]. *Journal of Machine Learning Research*, 2024, 25(60):1-8.
- [13] NICHOLSON W, MATTESON D, BIEN J. BigVAR: Tools for Modeling Sparse High-dimensional Multivariate Time Series[PP/OL]. V2. (2017-02-23)[2025-06-18]. <https://arxiv.org/abs/1702.07094>.
- [14] PETERS J, JANZING D, SCHÖLKOPF B. Causal Inference on Time Series Using Restricted Structural Equation Models [C]//BURGES C J Curges, BOTTOU L, WELLING M. *Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 1*. New York: Curran Associates Inc, 2013:154-162.
- [15] RUNGE J, NOWACK P, KRETSCHMER M, et al. Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets[J]. *Science Advances*, 2019, 5(11):eaau4996. doi:10.1126/sciadv.aau4996
- [16] TANK A, COVERT I, FOTI N, et al. Neural Granger Causality[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(8):4267-4279. doi:10.1109/TPAMI.2021.3065601
- [17] YANG D Z, HE X T, WANG J, et al. Federated Causality Learning with Explainable Adaptive Optimization[C]//AAAI. *Proceedings of the AAAI Conference on Artificial Intelligence*. Washington: AAAI Press, 2024: 16308-16315. doi: 10.1609/aaai.v38i15.29566
- [18] BECKMANN J, CZUDAJ R. Gold as an Inflation Hedge in a Time-varying Coefficient Framework[J]. *The North American Journal of Economics and Finance*, 2013, 24:208-222. doi:10.1016/j.najef.2012.10.007
- [19] MADZIWA L, PILLALAMARRY M, CHATTERJEE S. Gold Price Forecasting Using Multivariate Stochastic Model [J]. *Resources Policy*, 2022, 76:102544. doi:10.1016/j.resourpol.2021.102544

(责任编辑 赵静慧)