

修士論文について (仮)(仮) 未観測変数を考慮した時系列因果探 索とその検証について

蒲田 涼馬 (Ryoma Gamada)
u455007@st.pu-toyama.ac.jp

富山県立大学 電子・情報工学専攻 情報基盤工学部門

October 8, 2024

背景

経済の発展や衰退の裏には様々な要素が絡み合っており、それらの相互関係や因果関係を推定することは経済の動向を知ったり、新たな金融政策を考える上で非常に重要なものである。しかし、経済における因果探索についての研究はそこまで充実しておらず、また未観測変数を考慮したリアルデータに直接適用できるシステムを開発している研究はさらに少ない。

目的

未観測変数を考慮した因果探索を行い経済変数の因果関係を求める。
また、求めた因果性を用いて自動売買を行い、システムの有効性を
検証する。

未観測変数を考慮した因果探索では時系列を考慮したい。

因果探索の手法 1

4/10

LiNGAM

LiNGAM は統計的因果探索手法の 1 つである。

LiNGAM

定式化

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i \quad (i = 1, \dots, p)$$

【パラメータ】

観測値: x_i 誤差: e_i
観測値間の関係値: b_{ij}



$$\begin{cases} x_1 = x_2 + e_1 \\ x_2 = e_2 \\ x_3 = -x_1 + 2x_2 + e_3 \\ x_4 = x_2 + e_4 \end{cases}$$

LiNGAMにおける仮定

1. 因果関係を線形モデルで表現する
2. 変数が非ガウス性を持つ
3. グラフ構造は非循環である。
4. 変数が完全に観測されている。

VARモデル

定式化

$$\mathbf{x}(t) = \sum_{\tau=1}^k \mathbf{B}_{\tau} \mathbf{x}(t-\tau) + \mathbf{e}(t)$$

VARモデルにおける仮定

1. 各変数はその過去の値に依存する。
2. 誤差項が多変量正規分布に従う。
3. 各変数の誤差項は独立

VAR-LiNGAM

定式化

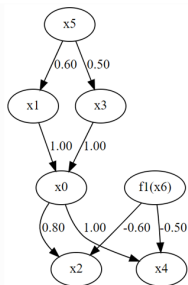
$$\mathbf{x}(t) = \sum_{\tau=0}^k \mathbf{B}_{\tau} \mathbf{x}(t-\tau) + \mathbf{e}(t)$$

VAR-LiNGAMにおける過程

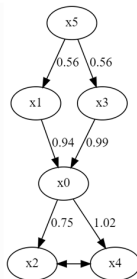


1. $e_i(t)$ は変数間で独立で時間的にも独立
2. $e_i(t)$ が従う分布は非ガウス分布
3. B_0 が表す因果グラフは日巡回グラフ
4. 変数が完全に観測されている
5. 各変数は過去の時間点の変数に影響を受ける

RCD 1



正解の因果グラフ



RCDによって得られる因果グラフ

RCDのモデル式

$$x_i = \sum_j b_{ij} x_j + \sum_k \lambda_{ik} f_k + e_i$$

x_i : 観測変数

b_{ij} : 観測変数間の因果強度

f_k : 潜在交絡因子(未観測変数)

λ_{ik} : 潜在交絡因子から観測変数
への因果強度

e_i : 外部効果

RCD

RCD では線形性, 非ガウス性, 有向非循環グラフ, 潜在交絡因子の独立性が仮定されている. また, RCD は LiNGAM と同じく回帰分析, 独立性テスト, 因果方向の推定のステップで因果性を求めるが, 潜在交絡因子を考慮するという点で異なる. そのため, RCD は観測変数間の残差に相関が残っている場合, それらの変数が潜在交絡因子によって影響を受けていると判断する. この場合, 二つの変数間に双方向の矢印を引いて潜在交絡因子の存在を示す.

CAM-UV

CAM-UV も同じく因果探索の手法であり、LiNGAM の拡張版である。CAM-UV と LiNGAM との違いはその仮定にある。CAM-UV では未観測変数を考慮しているのに加えて非線形のものも扱えるというメリットがある。

RCD と同じく未観測変数については双方向矢印 (\leftrightarrow) で表現する。これについてまだ勉強中。

進捗

RCD のプログラムを実装し、卒論で用いた金融データの一部について実行した。

結果としては双方向矢印が多く現れたが、具体的な因果性は導出されなかった。

この理由としてはデータが不足していること、RCD では線形性の因果関係のみしか扱えないため非線形のもの进行分析できていないということが考えられる。

ただ、全変数間に双方向因果が表れているわけではなかったもので、それが出力されていない変数間の関係についても調査する必要がある。

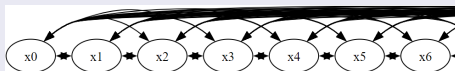


図 2: RCD の実行結果 (1 部抜粋)

今後の方針

今まで通り，未観測変数を考慮した因果性分析についての研究を進めていく．

現状の RCD では非線形性を考慮することができていないので，非線形性を考慮した CAM-UV について実装する．

CAM-UV を実装したあとには正解があるデモデータについて数値実験を行い手法そのものの有効性を確かめる．

CAM-UV で問題がないようであれば CAM-UV と VAR-LiNGAM についてより詳しく勉強し，CAM-UV を時系列データに利用できるようにする．

そこまでできれば自動売買システムに時系列を考慮した CAM-UV を適用しデモ取引を行うことで有効性を示す．

まとめ

LiNGAM の拡張型である RCD モデルを卒論で用いた金融データについて適用してみた.

結果としては具体的な因果性が導出されることはなくあまり良い結果とは言えなかった.

今後は CAM-UV をはじめとした別の手法について勉強し, 実装する必要がある.

新規性についてもまだ決まっていないので早めに因果探索に適用可能な手法を見つけ, その改善を考える必要がある.