

証拠に基づく政策立案のための 時系列を考慮したオープンデータ間の 因果関係分析とその検証

英語タイトル

長瀬 永遠 (Towa Nagase)
u255013@st.pu-toyama.ac.jp

富山県立大学 工学研究科
電子・情報工学専攻 情報基盤工学講座

December 6, 2022

はじめに
先行研究
時系列を考慮した
因果探索の既存
手法
数値実験並びに
考察
おわりに

はじめに

先行研究

時系列を考慮した
因果探索の既存
手法

数値実験並びに
考察

おわりに

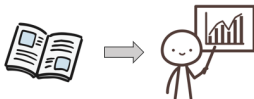
近年、世界各国で証拠に基づく政策立案（Evidence-Based Policy Making: EBPM）に注目が集まっている。日本においても例外ではなく、研究機関でも取り上げられている。政府を中心に推進されている EBPM であるが、地方自治体ではいまだにエピソードベースの意思決定が根強い。

EBPMの利点

- データに基づくため予測を行うことができ、問題が顕在化する前に対策を打つことが可能
- その場限りのエピソードによるものではなく、明確な根拠があるため、住民の理解が得やすい
- 政策による効果が事前に逆算できるため、状況に応じた微調整が可能

エビデンスベース

政策によって改善したい対象を明確化したうえでデータを収集し意思決定。



エピソードベース

住民によって役場に持ち込まれた問題に対して対面処理的に意思決定。



政策決定における対象の複雑性という課題に対して、多種多様なデータを収集し、データ間の因果関係を分析することで対象と因果関係のあるものを導出する。その結果を用いることで政策決定の支援を行う手法を提案する。

はじめに

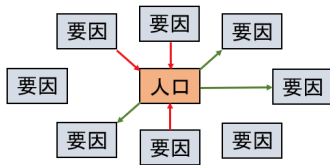
先行研究

時系列を考慮した
因果探索の既存
手法

数値実験並びに
考察

おわりに

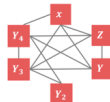
人口を増やしたい。



問題が複雑すぎる。
全貌が把握できない。



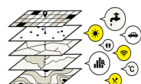
データ収集



因果探索

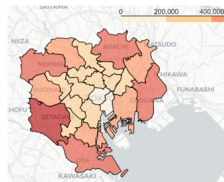


データ分析



結果の可視化

政策決定支援



はじめに

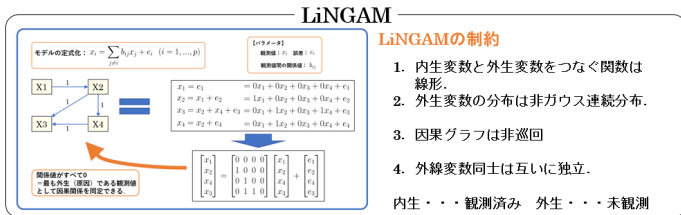
先行研究

時系列を考慮した
因果探索の既存
手法

数値実験並びに
考察

おわりに

因果探索とは、観測データを用いて、データ群の因果グラフ（複数の観測データにおいて、それぞれの値がお互いに及ぼしあっている影響の度合いを構造的に示したもの）を導出するための教師なし学習である。手法の一つとして、線形非ガウス非巡回モデル（Linear Non-Gaussian Acyclic Model: LiNGAM）がある。



現状の課題

- 通常の LiNGAM では同一時点のデータしか扱えない
- 因果グラフが非巡回とは限らない
- 未観測な変数が存在する

はじめに

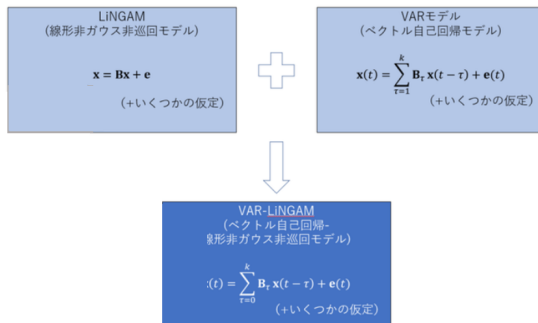
先行研究

時系列を考慮した
因果探索の既存
手法

数値実験並びに
考察

おわりに

時系列を考慮した LiNGAM の拡張モデルに VAR-LiNGAM¹ がある。VAR-LiNGAM とは、通常の LiNGAM にベクトル自己回帰モデル (Vector Auto Regressive model: VAR model) を加えることで異なる内生変数の過去データを考慮できるように拡張したモデルである。



¹A. Hyvarinen, K. Zhang, S. Shimizu, P. O. Hoyer, Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity, *Journal of Machine Learning Research*, vol. 11, pp. 1709-1731, 2010

VAR モデル² の定義

$\mathbf{y}_t$ を定数と自身の p 期の過去の値に回帰したモデルとすると式は以下のようになる。

$$\mathbf{y}_t = \mathbf{c} + \phi_1 \mathbf{y}_{t-1} + \cdots + \phi_p \mathbf{y}_{t-p} + \epsilon_t, \epsilon_t \sim W.N.(\Sigma) \quad (1)$$

(ただし $\mathbf{c}$ は $n \times 1$ 定数ベクトル、 ϕ_i は $n \times n$ 係数行列)

VAR モデルの推定

VAR モデルは各方程式が誤差項の相関を通じて関係する見かけ上無関係な回帰モデルに分類される。見かけ上無関係な回帰モデルは通常では誤差項同士の相関を考慮するために全ての回帰式を同時に同定しなければならないが、VAR モデルは全ての回帰式が同一の説明変数を持つという特殊性があるため、各回帰式を個別に最小二乗法で解けばよい。

VAR モデルでできること

ある時点 t におけるデータ y が過去の $t-1$ 時点のデータから受ける影響を分析できる。

²北坂真一, ”日本経済における構造変化と景気変動”, 日本経済学会誌, vol. 44, no. 2, pp. 142-158, 1993

はじめに

先行研究

時系列を考慮した
因果探索の既存
手法

数値実験並びに
考察

おわりに

VAR-LiNGAM の定義

$$\mathbf{y}_t = \mathbf{c} + \phi_1 \mathbf{y}_t + \phi_1 \mathbf{y}_{t-1} + \cdots + \phi_p \mathbf{y}_{t-p} + \epsilon_t, \epsilon_t \sim W.N.(\Sigma) \quad (2)$$

(ただし $\mathbf{c}$ は $n \times 1$ 定数ベクトル、 ϕ_i は $n \times n$ 係数行列)

VAR-LiNGAM の推定

- 1 VAR モデルの推定
- 2 VAR モデルの残差の推定
- 3 残差に対する LiNGAM モデル推定
- 4 VAR-LiNGAM の推定

VAR-LiNGAM でできること

ある時点 t におけるデータ y が同一時点のデータおよび過去の $t-1$ 時点のデータとどのような因果関係を構築するかを分析できる。

はじめに

先行研究

時系列を考慮した
因果探索の既存
手法

数値実験並びに
考察

おわりに

LiNGAM における 4 つの仮定

- 観測された変数と未観測の変数をつなぐ関数は線形
- 未観測の変数の分布は非ガウス連続分布
- 因果グラフは非巡回
- 未観測の変数同士は互いに独立

エラーの判定

LiNGAM の仮定が壊れていないかの判定において、現状いくつかの方法が考えられているが、その中に未観測の変数同士の独立性を判定する方法がある。LiNGAM における誤差項同士の独立性をもとめることによってこれを実現する手法が提案されている。

政策決定と VAR-LiNGAM



データ収集



時点 t
のデータ



時点 $t-1$
のデータ

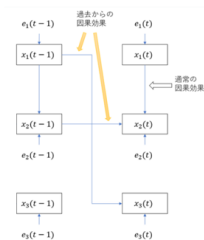


ターゲット

VAR-LiNGAM
適用



その後の工程へ



LiNGAM, VAR-LiNGAM のそれぞれを用いて年少人口に対する因果関係を持つデータ項目をもとめた。用いたデータ項目を以下の表 1 に示す。データはすべて内閣府が提供する RESAS³ というサイトから収集した。

表 1 用いたデータ

データ項目	データ項目	データ項目
年少人口	衛生費	経営耕地面積
生産年齢人口	教育費	耕作放棄地面積
老年人口	警察・消防費	農業経営者平均年齢
企業数	商工費	農業就労者平均年齢
事業所数	民生費	
就業者数	土木費	
一人当たりの住民税	総務費	
一人当たりの地方税	公債費	
一人当たりの固定資産税	労務費	
年間商品販売額	農林水産業費	
海面漁獲物等販売費	雑費	
製造品出荷額	住宅用地平均取引価格	
農業産出額	商業用地平均取引価格	
林産物販売額	農地平均取引価格	
林作業請負収入	林地平均取引価格	

³内閣府, ” RESAS 地域経済分析システム ”, <https://resas.go.jp/#/13/13101>

結果

年少人口に対して影響を与えるデータ項目は以下ようになった。

表 2 LiNGAMの結果

データ項目	パス係数
生産年齢人口	0.249
教育費	0.017
老年人口	-0.559
住宅用地平均取引価格	-0.043
警察・消防費	-0.038
商工費	-0.024
衛生費	-0.019
企業数	-0.006

表 3 VAR-LiNGAMの結果

データ項目	パス係数
民生費 (2015)	0.940
老年人口 (2015)	0.220
生産年齢人口 (2020)	0.180

考察

- LiNGAM と比較して VAR-LiNGAM では項目数が減ったが、因果関係が示された項目は全て一見妥当そうに見える
- 外生変数間の独立性の検証が必要

はじめに

先行研究

時系列を考慮した
因果探索の既存
手法

数値実験並びに
考察

おわりに

現状の課題

- 因果探索の精度に疑問がある
- 現状のデータベースと制約式との間にズレがある
- 時系列に対する増減を考慮できない

今後の方針

- 1 外生変数間の独立性を算出し，因果探索の精度を求める
- 2 データの拡充，数式の見直しによって因果探索の精度を向上させる
- 3 時系列に対する増減を考慮する手法を提案する