

金融変数と実体経済変数の 因果探索と数法則発見法による 波及経路のモデル化と可視化

Modeling and Visualization of Spillover Paths
Using Causal Search and Number Law Discovery Methods
for Financial and Real Economic Variables

蒲田 涼馬 (Ryouma Gamada)
u020010@st.pu-toyama.ac.jp

富山県立大学情報システム工学科 4 年

F221, AM 9:45-10:00 Friday, February 9, 2024,

1.1 本研究の背景

2/17

近年、金融工学は計算機性能の向上やデータサイエンス手法の進化、公的機関によるオープンデータの提供に伴い、飛躍的な発展を遂げている。将来予測などの分野は成長しているにも関わらず、多様な要素の相互関係や因果関係、経済の動向をモデル化するような研究は多くない。

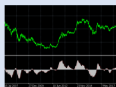


オープンデータ



計算機技術の向上

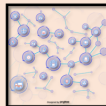
将来予測の分野



急激な成長



経済変数間の関係性明示



十分に研究されていない

図 1: 背景

1.2 本研究の目的

3/17

本研究では VAR-LiNGAM による時系列を考慮した因果探索を行い、時系列を考慮した 3D 因果グラフの作成、また数法則発見法を用いた経済変数のモデル化を行うことで、経済変数間の影響を直感的に理解できるようなシステムの実装を目指す。

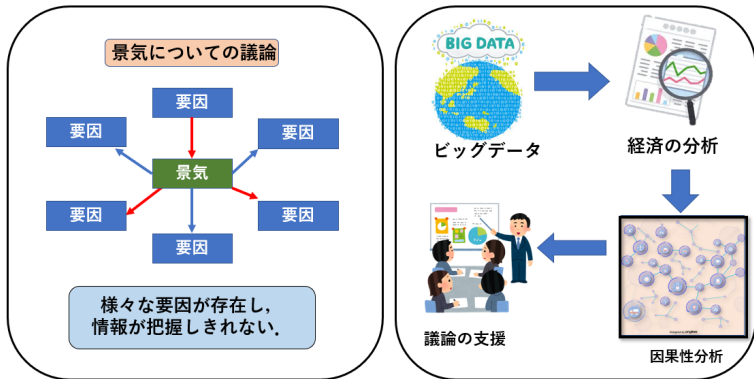


図 2: 本研究の目的

2.1 経済時系列の状態変数と指標

4/17

経済時系列は時間の経過に伴って変動する経済データの系列や指標を指す。これらは日本銀行など様々な機関によって公開されており、経済の時系列分析などで利用されている。

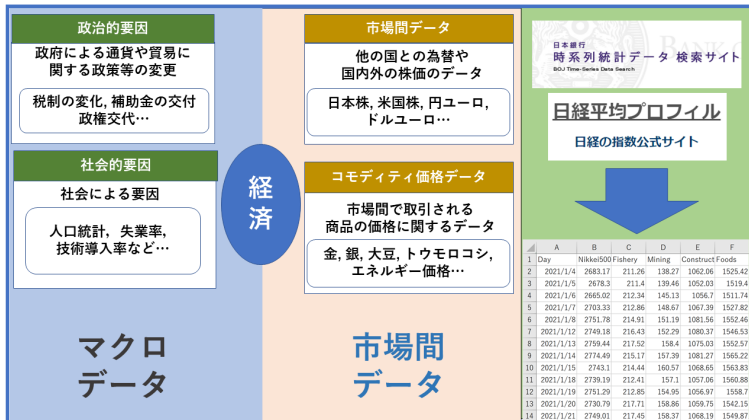


図 3: 経済情報の例

2.2 経済における波及メカニズム

5/17

経済状況进行分析するために用いられる概念として、経済の波及メカニズムというものがある。

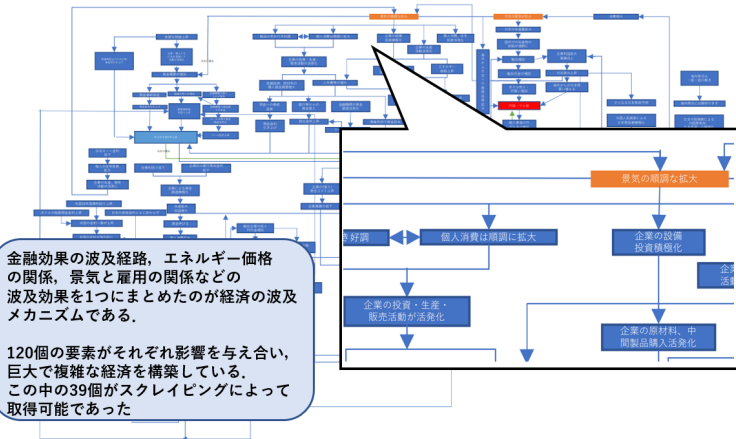


図 4: 経済波及メカニズムの例

2.3 統計的手法における経済の分析

6/17

統計的手法による経済分析は実際の経済データを用いて経済現象の定量化や関連要因の分析を行う手法である。時系列分析や回帰は典型的な例であり、将来予測などに役立つため特に注目を集めている。

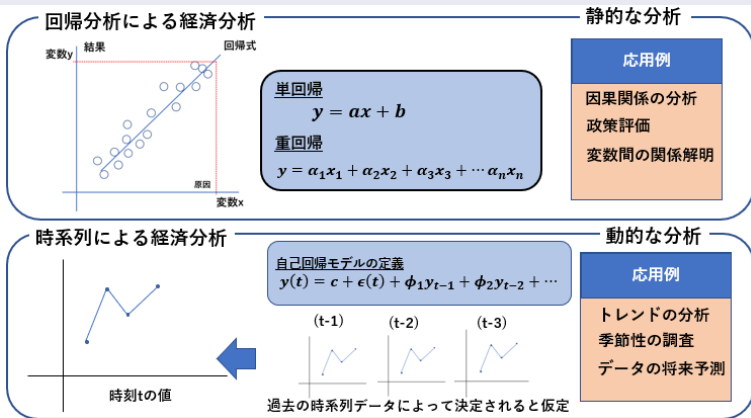


図 5: 基本的な統計的手法

3.1 種々の計量経済モデル

7/17

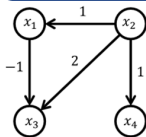
経済波及の分析を行うために、計量経済モデルを用いて因果性を導出する研究は様々ある。LiNGAM と VAR モデルでは変数間の因果関係を読み取ることができる。

LiNGAM

定式化

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

観測値: x_i
誤差: e_i
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

多変量時系列データ x_t に対し、自己回帰の次数が τ の VAR モデル

定式化

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

$x_{t-\tau}$: 目的変数ベクトルのラグ τ
 $e(t)$: K 次元の定数ベクトル
 B_{τ} : ラグ τ の $K \times K$ 係数行列

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

図 6: 統計的手法による因果探索

はじめに

経済情報と市場
データ

経済変動の数理モ
デル

提案手法

実験結果並びに
考察

おわりに

3.2 回帰を用いた数法則発見

8/17

経済においてモデル化は経済の動向を把握するために有用な手法である。経済のモデル化を行う手法の一つとして数法則発見法がある。

RF5法による数法則の発見

RF5法では多変量多項式回帰法の1つであり、以下の式でモデルを考える。

$$y = f(x; \theta) + \epsilon, \quad f(x; \theta) = w_0 + \sum_{j=1}^K w_j \prod_{k=1}^K x_k^{w_{jk}}$$

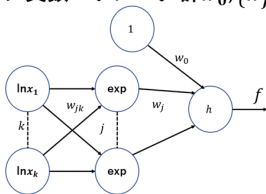
θ に実数パラメータ群 $w_0, \{w_j\}, \{w_{jk}\}$ が入る

K 個の説明変数

$x = (x_1, x_2, x_3, \dots, x_K)$

1個の目的変数

y



3層パーセプトロンの学習により、組み合わせ爆発を回避することができる。指数部分が整数でない多項式を求めることもでき、複雑な多項式を事前知識なしで容易に発見できる。

図 7: RF5 について

3.3 グラフネットワークによる経済の可視化

9/17

グラフネットワークは、複雑なデータ構造を複数のノードとそれをつなぐエッジで表現し要素の相互作用を可視化するための手法である。経済においてもグラフネットワークは利用されている。

消費者側
↓

	農業	工業	商業	最終需要	総生産量
農業	80	120	170	70	440
工業	250	230	80	40	600
商業	70	80	190	150	490
付加価値	40	170	50		
総生産量	440	600	490		

生産者側 ⇒

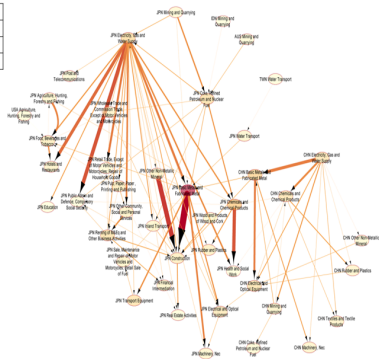
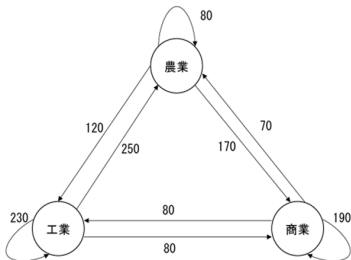


図 8: 経済におけるネットワーク活用の例

4.1 数法則発見法による経済変数のモデル化

10/17

本研究では経済変数のモデル化を行い，同定式を導出することで直感的に経済の動向を把握できるようにする．そのために，RF5 を用いてモデル化を行う．

スクレイピングによるデータ収集

日本銀行
時系列統計データ検索サイト
BOJ Time Series Data Search

日経平均プロフィール

日経の指数公式サイト

csvファイルに格納

	A	B	C	D	E	F
1	Day	Nikkei500	Fishery	Mining	Construct	Foods
2	2021/1/4	2683.17	211.26	138.27	1062.06	1525.42
3	2021/1/5	2678.3	211.4	139.46	1052.03	1519.4
4	2021/1/6	2665.02	212.34	145.13	1056.7	1511.74
5	2021/1/7	2703.33	212.86	148.67	1067.39	1527.82
6	2021/1/8	2751.78	214.91	151.19	1081.56	1552.46
7	2021/1/12	2749.18	216.43	152.29	1080.37	1546.53
8	2021/1/13	2759.44	217.52	158.4	1075.03	1552.57
9	2021/1/14	2774.49	215.17	157.39	1081.27	1565.22
10	2021/1/15	2743.1	214.44	160.57	1068.65	1563.83
11	2021/1/18	2739.19	212.41	157.1	1057.06	1560.88
12	2021/1/19	2751.29	212.85	154.95	1056.97	1558.7
13	2021/1/20	2730.79	217.71	158.86	1059.75	1542.15
14	2021/1/21	2749.01	217.45	158.37	1068.19	1549.87

モデルの作成

$$y = f(x; \theta) + \epsilon, \quad f(x; \theta) = w_0 + \sum_{j=1}^K w_j \prod_{k=1}^K x_{kj}^{\theta_{jk}}$$

	A	B	C	D	E	F	G
1	w_01	w_1	v_2	w_1_1	w_1_2	w_1_3	w_1_4
2	3.65333	-0.93447	-1.80242	0.63265	-0.48938	0.62482	0.34483

データの事前処理

$$\text{目的変数: } y = \frac{y_i - \mu}{\sigma}$$

$$\text{説明変数: } x'_i = \frac{x_i}{\text{MAX}x}$$

図 9: RF5 の流れ

4.2 VAR-LiNGAM による因果関係の導出

11/17

スクレイピングによって経済情報や金融情報を収集し、正規化などデータの前処理を行い、VAR-LiNGAM による因果性探索を行う。

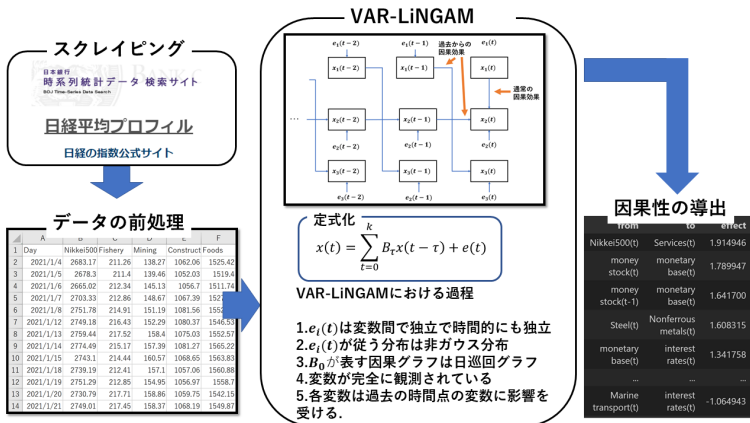
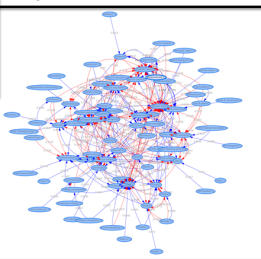
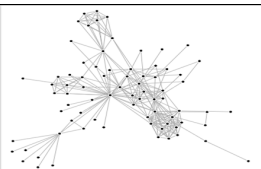
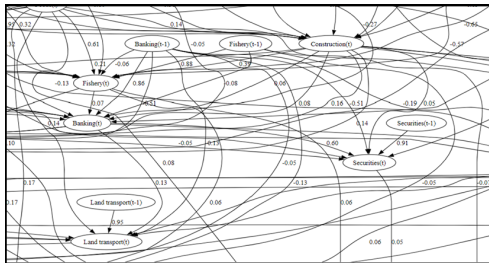


図 10: 時系列を考慮した因果探索システムの流れ

4.3 経済データの3D グラフによる可視化

12/17

直感的に因果関係を理解しやすくするために、3D グラフによる因果性の可視化を行う。



2Dグラフによる可視化ではエッジが増えるとスパゲッティ状になってしまい、情報が読み取りづらくなってしまふ。
3Dグラフによる可視化はエッジが増えた際に2Dグラフよりも可読性に優れている。

図 11: VAR-LiNGAM の実行結果

5.1 実験の概要

13/17

はじめに

経済情報と市場
データ

経済変動の数理モ
デル

提案手法

実験結果並びに
考察

おわりに

本研究では日経 33 業種の時系列データに加え、いくつかの金融データについて分析を行い 3D グラフによる可視化を行う。また、RF5 を用いた数法則発見により経済変数のモデル化を行う。今回は日足のデータに対して分析を行う。

データ項目	時間足	データ項目	時間足	データ項目	時間足
水産	日	窯業	日	金利	月
鉱業	日	鉄鋼	日	マネーストック	月
建設業	日	非鉄・金属	日	マネタリーベース	月
食品業	日	機械	日	エネルギー価格	月
繊維業	日	電気機器	日	日本物価指数 (総平均)	月
パルプ・紙業	日	ガス	日		
科学業	日	自動車	日		
医薬品	日	輸送用機器	日		
石油	日	精密機器	日		
商社	日	海運	日		
小売業	日	空運	日		
銀行	日	倉庫	日		
その他金融	日	通信	日		
証券	日	電力	日		
造船	日	サービス	日		
保険	日	ゴム	日		
不動産	日	陸運	日		
鉄道・バス	日	その他製造	日		

今回は経済の波及メカニズムを参考に40個のデータを収集し、実験を行う。市場間データはすべて株価であり、単位は全て円で表記される。物価指数、金利の単位のみ%で表記される。今回は日足と月足のデータを取り扱う。月足のデータは一か月間同じ値をとり続けると仮定し、日足の形に直してデータセットに格納した。

図 12: 本実験で用いるデータ

5.2 実験結果と考察

14/17

VAR-LiNGAM によって 3320 もの因果性を得ることができた。高い因果性を得られたものとブートストラップ法による確率が高かったものについて考察する。

from	to	effect
Nikkei500(t)	Services(t)	1.914946
money stock(t)	monetary base(t)	1.789947
money stock(t-1)	monetary base(t)	1.641700
Steel(t)	Nonferrous metals(t)	1.608315
monetary base(t)	interest rates(t)	1.341758
...
Marine transport(t)	interest rates(t)	-1.064943

	from	to	effect	probability
0	Transportation instruments(t-1)	Transportation instruments(t)	0.859349	0.987755
1	monetary base(t-1)	monetary base(t)	0.973287	0.985034
2	Petroleum(t-1)	Petroleum(t)	0.909080	0.983673
3	money stock(t-1)	money stock(t)	0.984165	0.982313
4	Communications(t-1)	Communications(t)	0.920379	0.982313
...
3316	Pharmaceuticals(t-1)	gross average(t)	-0.016641	0.300680
3317	Shipbuilding(t-1)	Retail(t)	-0.080427	0.297959
3318	Air transport(t-1)	Retail(t)	-0.050161	0.295238
3319	Air transport(t-1)	Nikkei500(t)	0.047179	0.295238
3320	Electric power(t-1)	Glass & Ceramics(t)	-0.121124	0.281633

海運業界の株価と金利には負の因果関係があると導出されている。これは一般的に知られている関係であり、正しい因果関係が導出されたと言えるだろう。

ブートストラップ法による検定結果で得られた確率を見ると、そのデータの過去のデータから得られる因果性の確率が高く、それ以外のデータから得られる因果性の確率は低くなっている。

これには変数間の関係が時系列によって異なっている可能性が考えられる。

はじめに

経済情報と市場
データ

経済変動の数理モ
デル

提案手法

実験結果並びに
考察

おわりに

5.2 実験結果と考察

15/17

2021 年 1 月から 2022 年 12 月までのデータを学習データ, 2023 年 1 月から 2023 年 12 月までのデータをテストデータとし, RF5 で実験を行った。



RMSE=0.444325



RMSE=0.648943



RMSE=0.142644



RMSE=0.29158

RF5によって得られた出力式

$$y = 4.112976 - 2.06447x_1 - 0.00311x_2 - 1.20865x_3 - 0.74313x_4 - 0.32073x_5 - 0.10715x_6 - 0.42619x_7 - 0.04326x_8 - 0.26576x_9 - 0.081282x_{10} - 0.572935x_{11} - 0.24974x_{12} - 0.58252x_{13} - 0.51609x_{14} - 0.67534x_{15} - 0.08419x_{16} - 0.20341x_{17} - 0.53962x_{18} - 0.692409x_{19} - 0.08506x_{20} - 0.35025x_{21} - 0.12891x_{22} - 0.055805x_{23} - 0.199628x_{24} - 0.051147x_{25} - 1.39007x_{26} - 0.19629x_{27} - 1.839659x_{28} - 0.68007x_{29} - 0.12772x_{30} - 0.22785x_{31} - 0.35563x_{32} - 0.63077x_{33} - 1.05056x_{34} - 0.07256x_{35} - 0.30954x_{36} - 0.92422x_{37} - 0.239614x_{38} - 1.44142x_{39}$$

学習データに対する検証では, RF5は局所的な解を捉えられず, RMSEは重回帰の方が低くなった。

部分的に別のアルゴリズムを適用することで対処する必要がある。

テストデータに対するモデルのRMSEはRF5の方が低かった。

RF5では重回帰分析で発見することができなかった汎化性のある数法則を発見したと考えられる。

今後はルールや状況を判断し, そのときに合った手法で数法則の発見を行う必要があると考えられる。

図 14: RF の実行結果

6 おわりに

16/17

動画でシステムの流れをお見せします。

はじめに

経済情報と市場
データ

経済変動の数理モ
デル

提案手法

実験結果並びに
考察

おわりに

まとめ

本研究では時系列を考慮した因果グラフネットワークの作成および RF5 法による数法則の発見を行い，利用者が経済の状況を直感的に理解できるシステムを作成した．

今後の課題

今後は数法則を発見する際に，季節や状況を考慮しモデルを作成できるようなシステムを作成する必要がある．また，金融データや市場データ以外の様々な要素を考慮して分析を行う必要があるだろう．また，今回作成した 3D グラフネットワークをさらにインタラクティブな操作をできるようにする必要がある．