

卒業論文

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

Curriculum Standardization and Learning Analytics for
WebBT in Teaching IR

富山県立大学 工学部 情報システム工学科

2020010 蒲田 涼馬

指導教員 奥原 浩之 教授

提出年月: 令和6年(2024年)2月

目 次

図一覧	ii
表一覧	iii
記号一覧	iv
第1章 はじめに	1
§ 1.1 本研究の背景	1
§ 1.2 本研究の目的	2
§ 1.3 本論文の概要	2
第2章 経済情報と市場データ	4
§ 2.1 経済時系列の状態変数と指標	4
§ 2.2 経済における波及メカニズム	6
§ 2.3 統計的手法における経済の分析	8
第3章 経済変動の数理モデル	13
§ 3.1 種々の計量経済モデル	13
§ 3.2 回帰を用いた数法則発見	16
§ 3.3 グラフ・ネットワークによる経済の可視化	19
第4章 提案手法	22
§ 4.1 数法則発見法による経済変数のモデル化	22
§ 4.2 VAR-LiNGAM による因果関係の導出	22
§ 4.3 経済データの 3D グラフによる可視化	25
第5章 実験結果並びに考察	27
§ 5.1 実験の概要	27
§ 5.2 実験結果と考察	27
第6章 おわりに	29
謝辞	30
参考文献	31

図一覧

2.1	EMA,BBAND,TRENDLINE	5
2.2	可視化された MACD,STOCH,ATR	5
2.3	時系列データの例: 為替時系列	5
2.4	マクロデータとミクロデータ	5
2.5	経済の波及メカニズム	6
2.6	EMA,BBAND,TRENDLINE	9
2.7	可視化された MACD,STOCH,ATR	9
2.8	パネルデータの例	10
2.9	因果関係	10
3.1	EMA,BBAND,TRENDLINE	14
3.2	可視化された MACD,STOCH,ATR	14
3.3	EMA,BBAND,TRENDLINE	19
3.4	可視化された MACD,STOCH,ATR	19
3.5	EMA,BBAND,TRENDLINE	20
3.6	可視化された MACD,STOCH,ATR	20
4.1	スクレイピングによって取得した時系列データ	23
4.2	可視化された MACD,STOCH,ATR	23
4.3	スクレイピングによって取得した時系列データ	23
4.4	可視化された MACD,STOCH,ATR	23
4.5	波及メカニズム	25
4.6	可視化システムの流れ	26
4.7	可視化に用いたデータ	26
5.1	EMA,BBAND,TRENDLINE	28
5.2	可視化された MACD,STOCH,ATR	28

表一覧

4.1	日経業種間データ	24
5.1	日経業種間データと金融変数	28
5.2	業種データ	28
5.3	金融変数	28

記号一覧

以下に本論文において用いられる用語と記号の対応表を示す。

用語	記号	用語	記号
特定のユーザー	x	g_c のユーザー数	$size(g_c)$
特定のアイテム	y	g_c の支持度数	$support(g_c)$
利用者数	n	g_c における協調度	$collaborate(g_c)$
アイテム数	m	u_{l_i} が属する頻出投稿者グループ	$G_{u_{l_i}}$
利用者集合 $\{1, \dots, n\}$	\mathcal{X}	投稿時間間隔が短いレビュー集合	g_b
アイテム集合 $\{1, \dots, m\}$	\mathcal{Y}	レビュー l_i の集中性スコア	$T_score(l_i)$
アイテム y を評価した利用者集合	\mathcal{X}_y	g_b のレビュー数	$size(g_b)$
ユーザー x が評価したアイテム集合	\mathcal{Y}_x	レビュー l_i と同じジャンルに属するレビュー数	o
対象ユーザー	a	レビュー l_i に出現する名詞集合	K_i
ユーザー x のアイテム y への評価値	r_{xy}	K_i の要素	$term_j$
ユーザー x による評価値の平均	\bar{r}_x	l_i と同じジャンルのレビュー集合において $term_j$ を含むレビューの数	$df(term_j)$
アイテム y への評価値の平均	\bar{r}_y	レビュー l_i の情報性スコア	$I_score(l_i)$
推薦されたアイテムの数	f	レビュー l_i の類似性スコア	$S_score(l_i)$
レコメンドで推薦されたアイテム集合	$ S_r $	レビュー l_i のサクラ性スコア	$F_score(l_i)$
推薦可能なアイテム集合	$ S_a $	教材 i における信頼性スコア	$K_score(i)$
教材 i につけられたレビュー文章	l_i	教材 i におけるスパムスコアの平均	$\bar{F}_score(i)$
レビュー文章 l_i を <i>bigram</i> によって区切った要素集合	X_{l_i}	教材 i のレビューの評価点の平均	\bar{l}_i
頻出投稿者グループ	g_c	レビュー l_i の協調性スコア	$C_score(l_i)$
Precision で考慮する上位ランクイングの数	N	Precision で考慮する人数	H

はじめに

§ 1.1 本研究の背景

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に伴い、飛躍的な発展を遂げている。その中でも特に顕著なトレンドは、1990年代半ばの外国為替証拠金取引自由化以降本格化した外国為替市場の拡大と高度情報化である。コンピューターと情報通信技術に支えられた市場への新規参入障壁の大幅引き下げに伴い、個人投資家を含む参加者がこの巨大市場に急増した。その結果、2022年時点での主要通貨間の平均日次売買高は驚異的な7.5兆ドルに達している[1]。これは外国為替市場の流動性が比類なき水準に高まったことを示していると言えるだろう。

この超流動的市場を支える背景には、銀行をはじめとした公的機関が市場関連データ、四半期ごとの企業業績データ、消費者信頼感指数や失業率等の各種マクロ経済指標、政治的イベントに関するニュース記事やSNSデータなど、関連する膨大かつ多様なデータソースを公開したことがある。それにより、機関や企業を問わずに個人がリアルタイムに入手・分析できる環境が実現した。こうした大量の構造化・非構造化データを深層学習等の最新アルゴリズムで処理することで、過去を大きく上回る精度の市場予測が可能になりつつある上、そこから自動売買戦略を抽出する試みも活発化している。例として、2022年にはグレンジャー因果性分析を用いて市場間分析を行い、MetaTrader5を用いて自動売買を行うシステムの研究がされている[2]。

しかしながら、市場変動や関連指標変化の原因解明と体系的理解は必ずしも十分とは言い難い状況にある。要因分析は個別事例に止まり、要素間相互作用や全体動態のモデル化には至っていない。経済システムを構成する多様な要素(変数)の相互作用や動的挙動をモデル化することは、複雑系としての経済の本質的理解や予測、リスク管理支援において有用であると考えられる。そのため、超流動的外国為替市場と膨大な関連データを最大限に活用しつつ、市場変動や他指標変化の本質的メカニズムを解明すること、またそれを可視化することで直感的理解を支援することはが金融工学分野において非常に重要だと言えるだろう。

§ 1.2 本研究の目的

1.1 節でも述べたように、近年、金融工学についての研究は国内外問わず盛んにおこなわれているものの、多様な要素の相互関係や因果関係、経済の動的挙動をモデル化するような研究は十分に行われていない。金融工学分野の研究は近年大きく進展しているものの、経済を構成する多様な要素間の相互作用や因果関係、ひいては経済全体の動的挙動を包括的にモデル化し解明する試みは必ずしも十分ではない。例えば外国為替市場の場合、マクロ経済動向や金利変動、政治リスクといった複数の外部要因が複雑に影響し合いながら為替レートを変動させているが、これら多要因間の定量的相関関係や変動メカニズムの体系的解明は立ち遅れがちである。

その大きな理由として、巨大で複雑な経済システムにおいて、様々なマクロ要因やミクロ要因の相互依存性を包括的かつ動的にモデル化することの困難さが挙げられる。しかし、この多要因間交互作用と経済全体の挙動原理を体系的に理解することこそが、変動予測の精度向上やリスク管理の高度化には不可欠な要素である。

したがって今後、高度な計算機シミュレーションやデータサイエンス的手法を駆使して、経済を構成する多様要素間の相関連鎖を包括的に解明し、マクロあるいはミクロレベルでの経済変動の本質的メカニズムのモデル化を強力に推進していくことが金融工学分野における最重要課題であると言える。これにより変動予測とリスク管理の精度が格段に向上し、実学面での更なる貢献の道が拓けることが期待される。そのため、本研究では時系列を考慮することが可能である、因果探索手法の VAR-LiNGAM を用いて変数間の因果関係を分析する。また VAR-LiNGAM を用いて得られた因果関係を人々が直感的に理解しやすいとされている 3D グラフ・ネットワークを用いて描画し、それについて考察を行う。

また、3 層パーセプトロンを用いた多変量多項式回帰法を使って経済変数におけるモデル式を導出し、それについても考察を行う。

§ 1.3 本論文の概要

本論文は次のように構成される。

第1章 本研究の背景と目的について説明した。背景では従来の金融工学研究の歴史と事例について述べ、目的では VAR-LiNGAM による因果探索、数法則発見法による 3D グラフによる可視化、またモデル化を行い、課題に挑むことについて述べる

第2章 従来の金融工学について、また経済における基本知識について解説する。

第3章 従来研究の金融工学に対するアプローチと、機械学習による数法則発見手法について述べる。また、可視化システムの実用事例について述べる。

第4章 データセットを作成し、金融時系列について因果探索、数法則発見を行う。また、得られた結果からグラフ・ネットワークによる可視化、同定式を表示するシステムの作成、またその流れを説明する。

第5章 実際の金融時系列データを用いて、第4章で述べた手法で、因果探索による3Dグラフの作成、数法則発見法による数法則の表示を行い、その精度、またシステムの有用性について解説する。

第6章 本論文における前章までの内容をまとめつつ、本研究で実現したことと今後の課題について述べる。

経済情報と市場データ

§ 2.1 経済時系列の状態変数と指標

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。これは経済活動や経済指標が時間とともにどのように変化しているかを示すためのデータである。経済指標は国の経済状況を数値化したものであり、各国の公的機関などによって公開されている。この背景には政府、地方自治体、公的機関などによる情報公開の取り組みである「オープンデータ」が注目を集めていることがある[4]。株価や為替レート、消費者物価指数などがその典型例であり、これらの時系列データは日本銀行を始めとする様々な公的機関によって公開されている。経済変数の時系列データには、景気循環に関する成分や、長期的なトレンド分析、季節変動や短周期の不規則変動に関する成分などが含まれている。
[1]

マクロデータとミクロデータ 経済データには大きく分けるとマクロデータとミクロデータの2種類がある。経済マクロデータは、特定の国や地域全体の経済を把握するための統計データを指し、これらのデータは経済の健全性や成長の指標として使用され、政府、企業、投資家、研究者などが経済の状態や動向を理解するために利用される。例として、国内総生産(Gross Domestic Product, GDP)、失業率、国際収支などがあり、これらのデータは日本銀行など各機関で公開されている。

マクロデータを分析する研究は長年行われており、財務省によってマクロ経済政策の影響等を資産するための経済モデルによる経済分析が行われている[5]。ミクロデータは、マクロデータよりももっと狭く詳細なデータを指す。例えば、個々の企業の売上高や利益などのデータ、各家計の支出などのデータ、労働市場などのデータがある。ミクロデータは政策の策定や意思決定などに用いられることが多い。

どちらのデータも機械学習において重要な役割を果たすが、一般的な経済変動予測などではマクロデータが用いられることが多い。

マクロデータとミクロデータの両方を効率的に使うことが統計的経済分析には重要である。

金融データ 金融データとは金融市场や金融活動に関連する様々なデータのことである。金融データには、マクロ的な経済・金融指標から個別企業の決算情報や株価データなど様々な種類がある。代表的な例を挙げると銀行が公表する金利や通貨供給量日本における金融データは、日本銀行の提供するマクロ金融統計を始め、金融商品取引所の株価や先物取引データ、REIT市場の運用資産や配当利回りデータ、企業の決算短信や有価証券報告書に含



図 2.1: EMA, BBAND, TRENDLINE



図 2.2: 可視化された MACD,STOCH,ATR



図 2.3: 時系列データの例: 為替時系列



図 2.4: マクロデータとミクロデータ

まれるミクロ金融情報、政府統計の金融関連項目といった公的データに加え、最近では API 連携によって容易に利用可能な実務データが充実しており、金融工学の研究や金融リスク管理、個人資産運用といった多目的に利用できる反面、個人金融資産に関するデータの限界も留意する必要があるなど、全体として金融分析活用のための有用なデータが整備されつつある状況にある。

データ収集

経済に関するデータを扱ううえで、データ収集の手法は非常に重要である。

経済データは政府機関や研究機関など、様々な組織によって公表されている。こうしたデータを効率よく収集し、分析を行えるようにすることは、経済分析システムにとって不可欠な要素だ。

データ収集手法の中でも、最近注目を集めているのがウェブスクレイピングである。Python や Ruby, R など、多くのプログラミング言語を使ってウェブページからデータを自動収集することができる。中でも Python によるスクレイピングでは様々なモジュールが提供されている。ここではそのモジュールについていくつか説明する。Python で頻繁に用いられるモジュールは 2 つある。まず 1 つ目は Selenium であり、Selenium ではクリックや入力といったユーザーアクションを自動的に模倣することで JavaScript によって生成されたコンテンツを取得することができる。しかし JavaScript が動作するサイトの場合は動的なコン

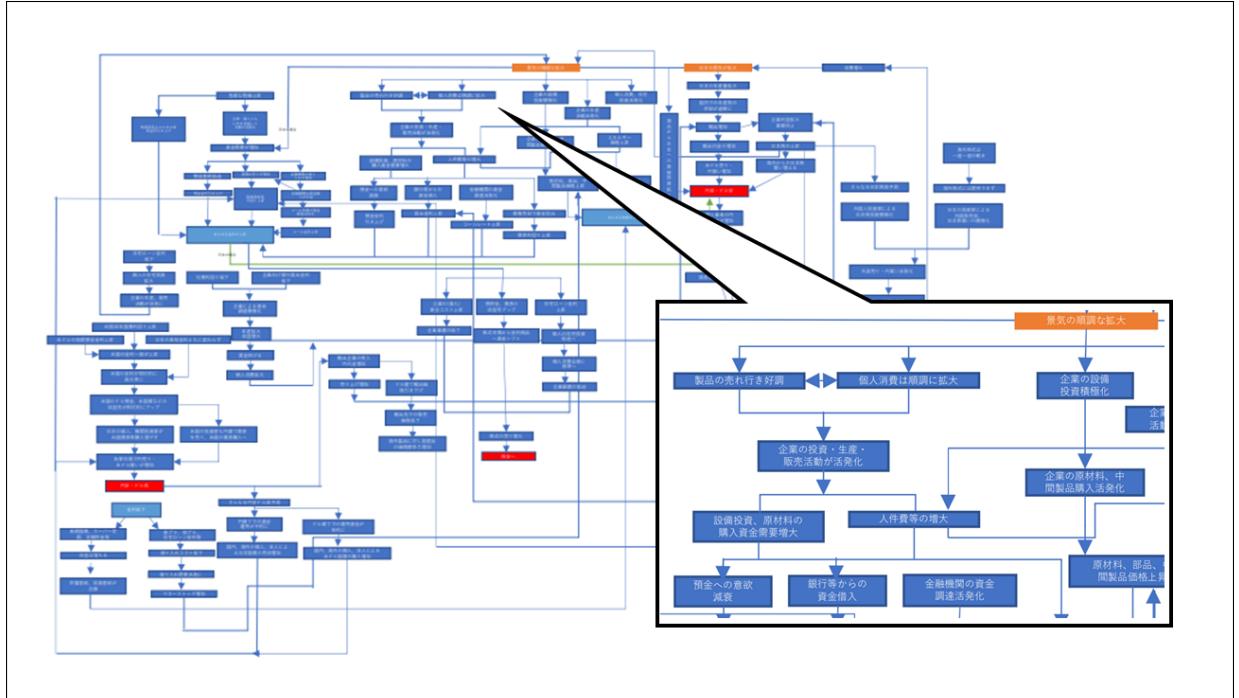


図 2.5: 経済の波及メカニズム

テンツを取得することが難しいと言ったような欠点が存在する。もう一つは requests である。requests では HTTP リクエストを発行し、サーバーからデータを取得することができる。これらのモジュールはそれぞれ長所と短所が存在し、効率的に使い分けてウェブスクレイピングを行う必要がある。政府統計や金融データ、企業業績など、経済分析に必要なデータの多くがウェブ上に存在しているため、ウェブスクレイピングは強力なツールとなりうる。しかしながら、ウェブスクレイピングには技術的課題や法的課題もあることに注意が必要である。たとえば、大量のデータ収集を行うとサーバに負荷がかかる可能性がある。また、著作権法やサイトの利用規約に抵触する形でのスクレイピングは違法となる。慎重な検討が不可欠であり、データ提供元との協力関係を構築することが望ましい。

このように、経済データ収集におけるウェブスクレイピングは可能性と課題が併存している状況だが、適正な利用が推進されていけば、経済分析の精度向上に大きな影響を与えることが期待される。

§ 2.2 経済における波及メカニズム

経済における波及メカニズムとは、ある経済変数の変動が、他の関連する経済変数にどのように影響を及ぼしていくかを表すメカニズムである。これはマクロ経済学における中心的な研究テーマであり、経済システムの複雑な動力学を理解する上で欠かせないものとなっている [4]。近年ではグローバル化と IT 化の進展により、世界的な経済変動の波及スピードが加速しているほか、波及経路自体の複雑化も指摘されている。したがって波及メカニズムの実証的解明と影響度評価は益々重要性を増しており、変動予測と適切な政策対応の精度向上に不可欠な要素として、経済学の更なる研究蓄積が望まれる領域である。以下では

いくつか経済波及の例について述べる。

i. 金融政策の波及メカニズム

金融政策の波及メカニズムとは、中央銀行が金利操作や金融調整といった金融政策手段を用いた場合、それが実体経済や金融システムにどのような影響を及ぼしていくかを表すメカニズムである。具体的には中央銀行が公開市場操作によってマネーサプライ量を拡大・縮小した場合、この政策によって市場金利が低下あるいは上昇し、銀行の貸出姿勢が緩和あるいは引き締まることになる。それにより、企業の設備投資意欲や個人の住宅ローン需要が刺激あるいは抑制され、マクロな景気循環にも影響を与えることになる。この一連の因果関係の流れが金融政策の波及メカニズムである。しかしその影響効果や時間プロファイルは、個人や企業の特徴によって異なってくる。例えば消費性向の高い個人ほど支出への影響が大きく、投資機会に恵まれた企業ほど設備投資への反応が敏感になる。このため波及効果を正確に把握するには、経済主体の異質性も合わせて考慮する必要がある。金融政策の波及メカニズムを分析するにはマクロとミクロの多面的アプローチが重要といえる。

ii. 原油高による経済波及

炭化水素資源である原油は、現代社会における重要なエネルギー・資源である。そのため、原油を通じたエネルギー価格の変動は、幅広い産業や経済主体に多大な影響力を持つことになる。特に原油高は企業の生産コスト上昇を通じて消費者物価に直接的に影響を与える。エネルギー集約型企業ほど影響は大きく、運輸・航空を始めとする各種移動サービスの料金上昇に波及する。また原油高は輸入原価の上昇を通じて交易条件の悪化を招きやすく、資源国通貨高・非資源国通貨安に加えてインフレ圧力を高める効果がある。こうしたコストプッシュ型インフレは企業収益の悪化を通じ、実体経済にマイナスに作用する。日本の2021年における過去10年間のデータを見ると、輸入量と輸入価格の積である総支払額は、最大の2013年で約14.5兆円、原油価格が下落した2016年では約6.2兆円になっている。これより原油価格の変動により、8兆円程度の対外収支が増減する効果があった。
[原油価格] 原油価格変動のこのような「多段階的波及効果」を正確に把握することは適切な政策対応には欠かせないとされている。さらに波及効果の国・地域差も考慮する必要がある。こうした点から原油価格変動に伴う波及メカニズムの精緻なモデル化と実証的検証が不可欠である。

iii. 景気と雇用、所得の関係

景気、雇用、所得は経済の重要な要素であり、経済成長という複雑なメカニズムを形作るものである。これらの変動がお互いに影響しあう。経済は通常、景気循環と呼ばれる一定の周期で変動する。景気が好調で成長している時期には企業の生産が拡大し、雇用と所得が増える。雇用が増えると需要と供給の法則に基づいて、賃金が上昇することがある。実際に2013年には金融緩和政策による景気回復と雇用への影響を単回帰にて分析した研究があり、この結果によると景気が雇用に影響を与えるという結果が得られている
[千葉経済論叢]。

iv. 経済の急成長とエネルギー価格の関係

急成長する経済は通常、産業の発展や都市化などに伴い、エネルギーの需要が急増し

エネルギー価格に対するエネルギー価格に対する需要圧力が高まる。急成長する経済は、エネルギー供給インフラの需要も高める。発電所やパイプラインなどの建設が必要になり、これがエネルギー価格に影響を与える。急成長する経済がエネルギー価格に与える影響は多岐にわたり、需要と供給、エネルギー政策などが絡み合って影響する。

v. 大幅為替変動による波及

大幅な為替変動は経済に広く影響をもたらすことがある。為替レートの変動は輸出入業界に大きな影響を与える。円安円高の関係(図参照)により為替に左右されることになる。また、為替変動は国内企業の業績へ影響を与える。特に多国籍企業にとっては重要で、海外での売上高が大きい企業は、為替変動により報告される売上高や利益に影響を受けることがある。為替変動は外国からの投資にも影響を与える。通貨が安定している国は外国からの投資をひきつけやすくなる。他にも金融市場、中央銀行による政策、国際的な取引にも影響を与える。

上で挙げた経済波及はほんの一部であり、実際はもっと複雑な形で経済波及メカニズムは構築される。図に「」によって得られた情報から作成した経済波及メカニズムの図を示す。経済波及の研究は長年行われており、特に金融政策の波及メカニズムについての研究は日本銀行などの金融機関を中心に行われている。

2020年には Heterogeneous Agent New Keynesian(HANK) モデルを用いた金融政策の波及メカニズムについての研究が日本銀行研究所で行われており、この研究では波及メカニズムの要因分解を行うことで金融政策の波及効果に重要な影響を与えると考えられている要因などについて議論がされている。また、波及メカニズムにおいて異時点間の代替効果以外の経路が重要な役割を果たす可能性があること、経済環境に応じて、波及効果自体が変わることの可能性があるという結論に至った。また、結果から政策効果の分析をどのように活用していくべきかを比較検証することが今後重要になるということを述べられていた。

§ 2.3 統計的手法における経済の分析

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。計量経済学とも呼ばれ、経済学の主要な手法の一つとして発展してきた。

代表的な統計的手法としては、回帰分析、時系列分析、相関分析など大別できる。これらの手法を用いることで、GDP成長率の決定要因分析や、為替変動の周期性分析、設備投資と景気循環の相関分析といった実証研究が可能になる。

近年ではこうした実証分析手法に加え、シミュレーションや推定の精度向上を目的とした統計的手法も高度化している。マルコフ連鎖モンテカルロ法やブートストラップ法などを用いることで、より複雑なマクロ経済モデルの推定や経済予測の信頼区間の算出などに活用されています。この結果として、有限の過去データから得られる確度の低い情報を、コンピュータシミュレーションによって補完し予測精度の向上を図る試みも増えている。

一方、近年の人工知能やビッグデータの発展も経済分析に新たな可能性をもたらしている。膨大な非構造化データから有用な情報を抽出するテキストマイニングや、サポートベ



図 2.6: EMA, BBAND, TRENDLINE



図 2.7: 可視化された MACD, STOCH, ATR

クターマシンに代表される機械学習など、新しい統計的・計算科学的手法によって分析効率の向上や新展開も期待されている。

いずれにせよ統計的手法は、過去の経験則やデータに依拠した認識原理を重視する経済学の実証分析において、主要な役割を果たし続けている。その変遷の過程で計量経済学という1つの体系的分野を形成し、今もって経済現象の解明に大きく貢献しているといえます。今後もコンピュータの計算能力向上や統計理論の発展、更なる大量の経済データの蓄積を受けて、経済分析への統計的アプローチの有用性は益々高まっていくものと考えられる。

i. 回帰による分析

回帰分析とは目的変数と複数個の説明変数との関係をモデル化する手法である。回帰分析は幅広い分野において適用されているが、経済分野でも長年注目を集めている。回帰分析には単回帰分析や重回帰分析など様々な種類があり、それぞれの分野において適した形で使われる。これらは事象の予測、検証、要因分析を行うときに用いられ、一般的に以下の回帰式にて表される。

$$y = a + bx \quad (2.1)$$

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n \quad (2.2)$$

その中でも特に重回帰分析は統計的データ手法の中で最も多く応用される有用な手法であるとされている[3]。重回帰分析は単回帰分析とは異なり、複数の説明変数を同時に扱うことができる。重回帰を用いた経済研究の例として、企業倒産についての重回帰モデルの構築をした研究がある。この研究では企業倒産数や企業倒産率、負債総額を目的変数として設定し、それを求めるための説明変数として為替相場、輸入額、輸出額などのマクロ経済情報を用いて分析を行った。結果としては倒産件数が、株価、有効求人倍率の2つのマクロ経済指標で説明されることを明らかにしており

ii. 時系列分析

時系列分析は、時間の経過に伴うデータの変動や形態を分析し、主にトレンドの特定や未来の値の予測などに用いられる統計的手法である。これは気象学や医学などを含める多くの分野においても用いられているが、経済の分野でも長年注目されている。2.1節にて述べた経済データにこれを適用し、経済の分野での予測、主に為替の変動



図 2.8: パネルデータの例



図 2.9: 因果関係

予測などに用いられることが多い。これによって経済変数の推移の実態把握、変動の周期性分析、変数間の動学的関係の解明などが可能になる。経済における時系列分析の研究としては、2022年

iii. パネルデータ分析

パネルデータ分析について述べるにはまず、パネルデータそのものについて説明する必要がある。パネルデータというものは異なる対象に対して同じ変数を複数の時間店で観測したデータのことである。このデータの形式はクロスセクションデータと時系列データを同時に扱うことができる。経済における対象は複数の国や地域などを指し、例えばGDP、株価、雇用などを複数の年度にわたって収集したデータがパネルデータになる。パネルデータ分析ではこれを分析することでそれぞれの国の異質性や時間の効果などを詳細に分析し、比較することができる。この手法を用いた研究には、長寿社会と経済格差の関係に焦点をあてたものがある。この研究では長寿社会に起因する経済格差の発生メカニズム、さらには経済格差がちゅおじゅ社会での人々の暮らしや働き方などに与える影響について、家計を追跡するパネルデータの構築、解析を通じている[パネルデータ論文]。

iv. 構造方程式モデリング

構造方程式モデリングは、観測変数と潜在変数のあいだの関係をモデル化する統計的手法である。これは、複雑なデータ構造や多様な変数間の関係性を理解するのに役立つ。潜在変数というのは直接観測することはできないが、複数の観測変数の背景に存在する概念や構造のことを指す。構造方程式モデリングではこれを考慮してモデリングを行うことができる。経済の分野においても、観測できない潜在的な概念や構造が存在することがある。構造方程式モデリングではこれらの潜在変数をモデル化し、観測変数との関係性を解明することが可能である。経済学に構造方程式モデリングを適用した研究もいくつかある。構造方程式を用いた経済分析では

v. 因果推論

因果推論は特定の出来事や介入が何らかの減少や結果に対してどのように影響を与えるかを推定するための手法である。経済における因果推論は、経済政策やビジネスの意思決定などに対して因果関係を明確にし、効果を評価するために重要である。基本

的に因果推論は時系列を考慮しないが、近年では時系列を考慮して因果関係を推論する手法が登場し、注目を集めている。

経済変動の数理モデル

§ 3.1 種々の計量経済モデル

計量経済学は、経済現象を数理・統計的手法を用いて分析・モデル化する学問分野です。計量経済モデルとは、この計量経済学の考え方に基づき、経済のメカニズムや変数間の関係性を数学的な式や方程式系によって定式化したものを指します。

計量経済モデルには大きく分けて、経済構造そのものを表現する構造形モデルと、特定の経済変数間の総合的な相関関係を表現する縮小形モデルがあります。前者は経済主体の意思決定や市場の働きなど、経済の根本メカニズムを体系的にモデル化する点に特徴があります。後者は実証分析に基づいて、複数経済変数間の相関関係を簡潔に定式化することに主眼が置かれます。

これらのモデルは、消費者や企業の最適化行動の仮定、市場の均衡条件の設定、複数の内生変数や外生変数の導入などの方式で構築されます。その際、個々の変数の挙動はパラメータによって定量的に規定されます。これらパラメータは実際の時系列データや断面データを用いた計量的手法(回帰分析等)に基づいて推定されることが多く、可能な限り実態を反映するよう設定されます。

構築されたモデルは、政策変更がマクロ経済・ミクロ経済に与える影響の定量的分析や、経済変数の将来予測に利用されます。特に政策当局にとって、異なる政策オプションを比較検討する上での有用性が高いとされている。

VAR モデル VAR モデルは多变量時系列データを分析する統計的手法である。複数の変数が時間と共にどのように捉えることができ、将来予測の分野などで特に注目を集めている。為替変動や売上高成長率の予測 [3] などの研究があり、経済の分野でもよく利用されていることがわかる。VAR モデルの中でも一般的な VAR モデル (p) の構造は以下のようになる。

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \varepsilon_t \quad (3.1)$$



図 3.1: EMA, BBAND, TRENDLINE



図 3.2: 可視化された MACD, STOCH, ATR

$$\begin{aligned}
 \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ \vdots \\ Y_{k,t} \end{bmatrix} &= \begin{bmatrix} a_{11,1} & a_{12,1} & \dots & a_{1k,1} \\ a_{21,1} & a_{22,1} & \dots & a_{2k,1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1,1} & a_{k2,1} & \dots & a_{kk,1} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ \vdots \\ Y_{k,t-1} \end{bmatrix} \\
 &+ \begin{bmatrix} a_{11,2} & a_{12,2} & \dots & a_{1k,2} \\ a_{21,2} & a_{22,2} & \dots & a_{2k,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1,2} & a_{k2,2} & \dots & a_{kk,2} \end{bmatrix} \begin{bmatrix} Y_{1,t-2} \\ Y_{2,t-2} \\ \vdots \\ Y_{k,t-2} \end{bmatrix} + \dots \\
 &+ \begin{bmatrix} a_{11,p} & a_{12,p} & \dots & a_{1k,p} \\ a_{21,p} & a_{22,p} & \dots & a_{2k,p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1,p} & a_{k2,p} & \dots & a_{kk,p} \end{bmatrix} \begin{bmatrix} Y_{1,t-p} \\ Y_{2,t-p} \\ \vdots \\ Y_{k,t-p} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{k,t} \end{bmatrix}
 \end{aligned}$$

近年では VAR モデルは単体として使われることは少なく、構造 VAR モデルなど拡張されて使われることが多いが、未だに最も基本的な多変量時系列データを分析するための手法として注目されている。近年では VAR モデルは単体として使われることは少なく、構造 VAR モデルなど拡張されて使われることが多いが、未だに最も基本的な多変量時系列データを分析するための手法として注目されている。経済分析に特に頻繁に用いられるのは構造 VAR である。構造 VAR モデルの構造については以下の式にして示す。

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + B X_t + \varepsilon_t \quad (3.2)$$

構造 VAR モデルでは VAR モデルに外部説明変数を取り入れることで経済ショックが内政変数にもたらす動学的な影響をインパルス応答などで数量的に表すことができる。また変数間の因果関係の方向性や即時効果の程度を知ることができる。これにより、金融政策の効果検証や政策分析への研究に適用されることがあり、構造 VAR モデルを用いた日本経済の資産蓄積、所得分配、負債の動態分析などの研究が行われている。

計量経済モデルの一種として LiNGAM によるものがある。

VAR-LiNGAM

VAR-LiNGAM は因果探索手法の一つであり、時系列データに対する因果関係の推定を行うためのモデル化手法である。 VAR-LiNGAM は 2010 年に提唱され [VAR-LiNGAM 元論文]，様々な分野で活用されている。 VAR-LiNGAM は VAR モデルと Linear Non-Gaussian Acyclic Model(LiNGAM) を組み合わせたものであり，図のような形になる。 LiNGAM は因果探索のセミパラメトリックなアプローチの手法であり定式化は以下のようになる。

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i \quad (i = 1, \dots, p) \quad (3.3)$$

そしてこれは行列表現で以下のように表すことができる。

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{bmatrix}, \quad E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}$$

VAR-LiNGAM ではこれに VAR モデルを加える。 VAR-LiNGAM のイメージは図に示す。 VARLiNGAM における仮定として VAR-LiNGAM の一般的な定式化としては以下のようになる

$$x(t) = \sum_{\tau=0}^k b_{ij} B_\tau x(t-\tau) + e(t) \quad (3.4)$$

これは VAR モデルと似た形になっているが，時間差を表す τ が 0 から始まっている。 B_0 は LiNGAM における B のように現時点での多変数からの因果を表す係数であり， $B_\tau (\tau = 1, \dots, k)$ は過去時点の $x(t-\tau)$ からの直接的な影響を表す係数であり，これにより現在のベクトル変数の値 $x(t)$ を現時点の他の変数だけでなく過去時点のベクトル変数も用いて推定することができる。

モデルの推定方法

VAR-LiNGAM の推定方法はいくつかあるがここでは VAR モデルとして推定を行ったあとに通常の LiNGAM モデルの推定を行う 2 段階法について紹介する。まず，時系列ベクトル $x(t)$ に対する通常の VAR モデルを以下のように構築する。

$$x(t) = \sum_{\tau=1}^k M_\tau x(t-\tau) + n(t) \quad (3.5)$$

これから最小二乗推定量 M^τ を計算する。ここで求めた M^τ を代入することで搅乱項の推定値 n^t を計算する。

$$\hat{n}(t) = x(t) - \sum_{\tau=1}^k \hat{M}_\tau x(t-\tau) \quad (3.6)$$

$\hat{n}(t)$ に対する LiNGAM 推定を行う。これにより因果係数行列の推定値 \hat{B}_0 が求まる。

$$\hat{B}_\tau = (I - \hat{B}_0)\hat{M}_\tau \quad (3.7)$$

上で求めた \hat{M}_τ と \hat{B}_0 から \hat{B}_τ を求める。サンプル数を大きくすると正しい値に収束することが VAR-LiNGAM の元論文では示されている。

ブートストラップ法による係数の有意性検定

VAR-LiNGAM 含め LiNGAM の手法において推定された因果係数が 0 と有意に異なるかどうかは非常に重要である。そこでブートストラップサンプリングを用いて係数の経験分布を求ることで有意性を確認する必要がある。分散を示すシンプルな統計量を提案し、それを用いて仮説検定を行う方法を紹介する。

§ 3.2 回帰を用いた数法則発見

機械学習は、経済分析において非常に有用なツールとして広く利用されています。機械学習は、大量のデータからパターンや関係性を学習し、未知のデータに対する予測や意思決定を行うための手法を指す。現在、機械学習はあらゆる分野で注目されており、経済においてもその例外ではない。中でも機械学習による経済のモデル化については長年多くの研究が行われている。1979 年には機械学習である GMDH(Group Methods of Data Handling) を用いた日本経済のマクロ計量経済モデルについての研究が行われており、マクロ計量モデルの構成がされた。

RF5 法による数法則の発見

数法則発見法では質的変数を考慮した方法と質的変数を考慮せずに量的変数のみを考慮する方法がある。3 層パーセプトロンを用いた多変量多項式回帰法である RF5 を定式化する。K 個の説明変数を $x = (x_1, x_2, \dots, x_K)$ 、目的変数を y として、以下の多変量多項式回帰モデルを考える。

$$y = f(x; \theta) + \epsilon, \quad f(x; \theta) = w_0 + \sum_{j=1}^J w_j \prod_{k=1}^K x_k^{w_{jk}} \quad (3.8)$$

このとき、定数項 w_0 、係数 w_j や指数 w_{jk} は未知の実数パラメータ、工数 J は未知の整数パラメータ、 ϵ は誤差項である。要するにデータから求めるべき数法則は項数が未知で指数が実数の多項式である。対象とする法則が周期関数や不連続関数からなる場合、上記の式では対応できない。しかし、このような関数であっても入力ベクトルのレンジが限られている時には、有限項数の多項式を用いて、ある程度の精度で近似することが可能である。さらに実数指数の多項式ならば、勝っく耕空の表現能力が向上する。そのため少ない項数での近似が期待できる。ここでは実数パラメータ群 w_0, w_j, w_{jk} をまとめて θ と書く。パラメータ θ の次元を M とする。 x_k を仮定すると、式の $f(x; \theta)$ は以下のように変形することができる。 $\prod_{k=1}^K x_k^{w_{jk}}$ に対して対数をとり、その上に指数をとる形にする。

$$f(x; \theta) = w_0 + \sum_{j=1}^J w_j \exp(\ln(\prod_{k=1}^K s_k^{w_{jk}})) \quad (3.9)$$

さらにこれを下のように変形する.

$$f(x; \theta) = w_0 + \sum_{j=1}^J w_j \exp\left(\sum_{k=1}^K w_{jk} \ln x_k\right) \quad (3.10)$$

これは、3層パーセプトロンの前向き伝播である. ただし、隠れユニットにバイアスはなく、活性化関数は指数関数であるので、隠れユニットの出力は以下となる.

$$m_j = \exp\left(\sum_{k=1}^K w_{jk} \ln x_k\right) \quad (3.11)$$

出力ユニットは線形で、バイアスを持つので、その出力は以下となる.

$$f(x; \theta) = w_0 + \sum_{j=1}^J w_j m_j \quad (3.12)$$

このような3層パーセプトロンを多項式ネットという. なお、中間ユニット数は多項式の定数項を除いた項数 J に対応することがわかる. J の最適な値は未知のため、何らかの方法を用いて決定する必要がある. いま、データとして N 個のサンプル $\{(x^\mu, y^\mu) | \mu = 1, \dots, N\}$ が与えられていると仮定する. 最適な J を何らかのモデル選択法によって決定すれば、データを最も良く説明するパラメータベクトル θ は以下の2乗和誤差を最小するものとして決まる.

$$E(x; \theta) = \frac{1}{2} \sum_{\mu=1}^N (f(x^\mu; \theta) - y^\mu)^2 \quad (3.13)$$

式(3.8)によって表される多項式は表現能力が高く、非線形なデータにも柔軟に対応することができる. しかし、これは多くの局所会を含むため、そのパラメータの学習は難しい. したがって、標準的な再急降下法では望ましい結果を得ることはできず、1次収束であるため、学習が収束するのも遅い. そこで、常に効率よく望ましい結果を得るために、RF5法では、2次の学習アルゴリズム Back Propagation based on Quasi-Newton(BPQ法)を用いて学習を行う.

BPQ法を用いた多層パーセプトロンの学習

BPQ法は準ニュートン法の基本枠組みとし、最適ステップ幅を2次近似のサイ焦点として計算する. 準ニュートン法には特徴があり、探索の仮定で反復により2次微分の逆行列 $(\nabla^2 E(x; \theta))^{-1}$ の近似値を各ステップで求める. 以下に準ニュートン法の基本アルゴリズムについて示す.

step1. 初期化

θ_1 を初期化し、 $H_1 = I, k = 1$ とする.

step2. 探索方向の計算

探索方向 $\delta\theta_k = -H_k \nabla E(x; \theta_k)$ を計算する. 停止条件を満たせば、反復を終了させる.

step3. 最適探索幅の計算

$E(x; \theta_k + \alpha \delta \theta_k)$ を最小にする α_k を求める.

step4. 結合重みの修正

結合重みを $\theta_{k+1} = \theta_k + \alpha_k \delta \theta_k$ に修正する.

step5. H の更新

$k = 0 \pmod{M}$, M : 全パラメータ数) ならば $H_{k+1} = I$ とし, それ以外の場合には H_{k+1} を更新する.

$k \leftarrow k + 1$ として step2 へ戻る.

step5 での H_{k+1} の計算にはいくつかの方法があるが, 制約のない非線形最適化では Broydon-Fletcher-Goldfarb-Shanno(BFGS) 公式が最も優れた方法であるとされている. 以下に BFGS 公式について示す.

$$\begin{aligned} H_{k+1} &= H_k + \left(1 + \frac{q^T H_k q}{pq} \frac{pp^T}{p^T p} - \frac{pq^T H_k + H_k qp^T}{p^T q} \right) \\ p &= \lambda_k \delta \theta_k \\ q &= \nabla E(x; \theta_{k+1}) - \nabla x; \theta_k \end{aligned} \tag{3.14}$$

多層パーセプトロンの学習には, BPQ 法以外の方法を用いることも考えられる. 最小二乗法の目的関数におけるヘッセ行列を近似的に求めてニューロン法を行うガウス・ニュートン法や, その拡張であるレーベンバーグ・マーカート法などがある. これらの方法を用いて学習をすることもできるが, 収束の速度や性能にはあまり差がないとされている.

最適中間ユニット数の選択

式 (3.8) における多項式の項数は J であり, これは 3 層パーセプトロンの中間ユニット数にあたる. 一般に, 与えられたデータ集合に対して, 最適な中間ユニット数をあらかじめ知ることはできない. そのため, 中間ユニット数を変えて発見した候補を適切に評価するための尺度が必要である. その手法の一つとして Bayesian Information Criterion(BIC) というものがある. BIC では, ある J の値に対して一回学習をすればよい. そのため, 少ない計算量で最適ユニット数を決定できる. BIC は以下の式で計算される.

$$BIC = -L(\hat{\theta}) + \frac{M}{2} \log N \tag{3.15}$$

ここで, $L(\hat{\theta})$ は最大対数尤度, M は総パラメータ数, N は総サンプル数を表す. いま, 目的変数の誤差 ϵ が平均 0, 標準偏差 σ の正規分布に従うとする.

$$p(y|x; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(f(x; \theta) - y)^2}{2\sigma^2}\right) \tag{3.16}$$



図 3.3: EMA, BBAND, TRENDLINE



図 3.4: 可視化された MACD, STOCH, ATR

学習によって得られたパラメータベクトルを $\hat{\theta}$ とすると、最大対数尤度は以下のようにになる。

$$\begin{aligned}
 L(\hat{\theta}) &= \log \prod_{\mu=1}^N p(y^\mu | x^\mu; \hat{\theta}) \\
 &= \sum_{mu=1}^N \log p(y^\mu | x^{mu}; \hat{\theta}) \\
 &= - \sum_{mu=1}^N \text{frac}f(x^\mu; \hat{\theta} - y^\mu)^2 2\sigma^2 - N \log \sqrt{2\pi\sigma} \\
 &= -\frac{1}{\sigma^2} \sum_{\mu=1}^N \frac{f(x^\mu; \hat{\theta} - y^\mu)^2}{2} - \frac{N}{2} \log(\sigma^2) - N \log \sqrt{2\pi}
 \end{aligned} \tag{3.17}$$

これにより、式 (3.17) を最大化する σ^2 を求めると以下となる。

$$\frac{\partial L(\hat{\theta})}{\partial \sigma^2} = \frac{1}{(\sigma^2)^2} \sum_{\mu=1}^N \frac{(f(x^{mu}; \hat{\theta}) - y^{mu})^2}{2} - \frac{N}{2\sigma^2} = 0 \tag{3.18}$$

$$\iff \hat{\sigma}^2 = \frac{1}{N} \sum_{\mu=1}^N (f(x^\mu; \hat{\theta}) - y^\mu)^2 \tag{3.19}$$

これを式 (3.17) に代入することによって $L(\hat{\theta})$ を計算できる。モデル選択に無関係の項を除外すると以下のようになる。

$$BIC(J) = \frac{N}{2} \log \left(\frac{1}{N} \sum_{\mu=1}^N (f(x^{mu}; \hat{\theta}_J) - y^\mu)^2 \right) + \frac{M}{2} \log N \tag{3.20}$$

$J = 1, 2, \dots$ と変えてパーセプトロンの学習を行い、 BIC_J を J^* として選択する。ただし、 $\hat{\theta}_J$ は J 個の中間ユニットのパーセプトロンの学習後に得られたパラメータベクトルを表す。

§ 3.3 グラフ・ネットワークによる経済の可視化

2 章 2 節でも述べた通り、経済は複雑で相互に関連する多くの要素から成り立っている。企業、市場、国々、金融機関などが絡み合い、その複雑性を理解し、適切な意思決定を行う



図 3.5: EMA, BBAND, TRENDLINE



図 3.6: 可視化された MACD, STOCH, ATR

ことは困難である。そこで、最近では新たなデータサイエンスの手法としてグラフネットワークが経済分析に活用されつつある。ここでは、グラフネットワークに焦点を当て、その経済分析における重要性、具体的な応用例について述べる。

グラフ・ネットワーク

グラフネットワークは、複雑なデータ構造を複数個のノードと、それらをつなぐエッジで表現し、これらの要素がどのように作用しているかをグラフとして視覚化する手法である。グラフ・ネットワークは様々な分野で使用されており、分野によって異なる種類が使われることになる。そのうちの1つは無向ネットワークであり、これは人間関係や道路などを表すときに用いられる。無向ネットワークの特徴として、エッジに方向がないというものがある。そのため、無向グラフは相互に関係する要素が互いに同程度の影響を持つ場合に使われる。また、もう一つ重要なネットワークとして有向グラフというものがある。有向グラフは無向グラフと異なり、エッジに方向を持っている。そのため因果関係や確率関係などを描画するときに非常に有用であり、様々な分野で注目を集めている。また、重み付きグラフというものもある。重み付きグラフはエッジに影響の大きさを表す値をグラフネットワークは経済の分野でも注目を集めており、例えば経済の例でいうと、企業、産業、市場、国家などがノードとして配置され、それらの関係性がエッジで結ぶことができる。これにより、経済の構造や相互関係を直感的に理解することが可能になります。

経済への応用

グラフネットワークを経済に用いることによる利点はいくつかある。まず、グラフネットワークによる可視化は利用者が情報を直感的に把握することができる。グラフネットワークは情報を直感的かつ視覚的に表現する手法であるため、要素同士の相互作用が明示的になり、経済変数間の伝播など複雑なデータ構造を理解しやすくなる。もう一つの利点として、グラフネットワークを用いることで経済の変動やリスクがどのようにシミュレーションすることができる。ベイジアンネットワークを用いた経済の伝播を予測する研究などがあり[本邦]、ストレス事象が発生したときの伝播を予測している。

グラフネットワークの経済分析への応用例として、産業連関図の構築がある。

グラフネットワークを用いて産業連関図を構築することで、異なる産業や企業の相互の依存関係を可視化することができる。例えば、自動車産業が鋼鉄産業に依存している場合、これをエッジで結ぶことで、製造業界全体のリスクや変動を理解する手助けとなる。これにより、サプライチェーン全体の複雑な構造を理解し、リスクヘッジの方針を検討することができる。

他の例として、国際経済のネットワーク分析がある。グラフネットワークを用いて国家や地域の経済的な相互関係を分析することで、国際経済のネットワークがどのように形成されているかを理解できる。国際貿易、外国直接投資、金融取引などの要素をノードとして結び、これらのエッジを通じて国家間の経済的な相互依存関係を可視化する。これは、グローバルな金融危機や貿易紛争の影響評価において役立つことがある。

また、金融市場の相関関係の解明についてグラフネットワークを用いたものもある。グラフネットワークを用いて金融市場内の企業、投資家、資産クラスの関係性を分析することで、市場の相関関係やリスク伝播のメカニズムを理解することができる。企業の財務データや市場指標がノードとなり、これらのエッジが相関や取引の関係を示すことで、市場の変動がどのようにして異なる要因に影響を与えるかを分析する。これにより、ポートフォリオの構築やリスク管理の向上が期待される。

グラフネットワークによる経済分析は、従来の手法に比べて新しい洞察をもたらすと期待されている。

グラフネットワークによる経済分析は、経済の複雑性に迅速かつ直感的に対応する新しい手法として注目されている。産業連関図の構築や国際経済のネットワーク分析、金融市場の相関関係の解明など、多岐にわたる応用例が存在する。これにより、経済主体や政策立案者が迅速な意思決定を行い、持続可能な発展を促進するための戦略を構築する際に有益な情報を提供することができる。将来的には、この手法が進化し、経済分析における新たな次元を切り開くことが期待されている。

提案手法

§ 4.1 数法則発見法による経済変数のモデル化

本研究では経済変数がそれぞれどのように作用しあっているのかを把握するために、数法則発見法の一つである、3層パーセプトロンを用いた多変量多項式回帰法、RF5法を用いて経済関係のモデル化を行う。数法則発見法にはいくつか種類があり、質的変数を考慮したRF6.4などがあるが、今回扱う変数は経済時系列データであり、質的変数というものがそもそもあまり存在しない。そのため、本研究では質的変数を考慮せず量的変数のみを考慮するRF5法を用いてモデル化を行うこととする。まず、数法則発見法RF5を用いるメリットはいくつかあるのでそれについて説明する。RF5では自動的に変数の選択を行い、必要のない変数はモデル化された式に加わらないようになっている。要するに自動的に変数選択を行うことができる。

用いるデータ

先にも述べたようにRF5法では質的変数を考慮せず量的変数のみを対象として回帰を行う。そのため、用いるデータは全て量的変数でなければいけない。また、RF5では時系列を考慮することができていないため、日付を除いてスクレイピングを行う必要がある。用いるデータの例を図に示す。今回はまずRF5によるモデル化が意味を持つものなのかを確かめるために、デモデータによる実験を行った。デモ実験では従来研究[RF]を参考に y に影響を与える変数 x_1, x_2, x_3, x_4, x_5 と y に何も影響を与えない変数 x_6, x_7, \dots, x_{10} を説明変数とし、デモ実験に用いたデータを示す。

データの例

§ 4.2 VAR-LiNGAMによる因果関係の導出

本研究では因果探索を行い、因果探索を行う手法はいくつかあるが、本研究では因果探索の手法として時系列を考慮した因果性を求めることができるVAR-LiNGAMを採用する。

スクレイピングによるデータの収集

VAR-LiNGAMで用いるデータを自動的に収集するためにスクレイピングを実装する。本研究では様々なサイトからスクレイピングを行い、時系列データを取得する必要がある。本研

A	B	C	D	E	F	G
1 データコ-PR01'PRCPR01'PRCPR01'PRCPR01'PRCPR01'PRCPR01'PRCPR01'PRC						
2 系列名称 [国内企業/大類別/工 類別/_飲食類別/_織物類別/_木林類別/_パ						
3 Jan-18	100	100	97.6	96.3	100.5	92.3
4 Feb-18	100	100	97.6	96.9	100.9	92.6
5 Mar-18	99.9	99.8	97.7	97.4	101.3	92.6
6 Apr-18	100.3	100.2	98.1	97.3	101.5	93
7 May-18	100.8	100.6	97.9	97.5	101.4	93.1
8 Jun-18	101.1	100.9	98.1	97.6	101.6	93
9 Jul-18	101.5	101	98.3	97.3	101.3	93.3
10 Aug-18	101.5	101	98.3	97.7	101.1	93.3
11 Sep-18	101.7	101.2	98.4	98	101	93.1
12 Oct-18	102.1	101.8	98.9	98.4	101.1	93.1
13 Nov-18	101.7	101.4	98.7	98.7	101.1	93.9

図 4.1: スクレイピングによって取得した時系列データ



図 4.2: 可視化された MACD,STOCH,ATR

A	B	C	D	E	F	G
1 データコ-PR01'PRCPR01'PRCPR01'PRCPR01'PRCPR01'PRCPR01'PRCPR01'PRC						
2 系列名称 [国内企業/大類別/工 類別/_飲食類別/_織物類別/_木林類別/_パ						
3 Jan-18	100	100	97.6	96.3	100.5	92.3
4 Feb-18	100	100	97.6	96.9	100.9	92.6
5 Mar-18	99.9	99.8	97.7	97.4	101.3	92.6
6 Apr-18	100.3	100.2	98.1	97.3	101.5	93
7 May-18	100.8	100.6	97.9	97.5	101.4	93.1
8 Jun-18	101.1	100.9	98.1	97.6	101.6	93
9 Jul-18	101.5	101	98.3	97.3	101.3	93.3
10 Aug-18	101.5	101	98.3	97.7	101.1	93.3
11 Sep-18	101.7	101.2	98.4	98	101	93.1
12 Oct-18	102.1	101.8	98.9	98.4	101.1	93.1
13 Nov-18	101.7	101.4	98.7	98.7	101.1	93.9

図 4.3: スクレイピングによって取得した時系列データ



図 4.4: 可視化された MACD,STOCH,ATR

究のシステムは Python を用いて実装するので、Python でのライブラリである「Selenium」を用いてデータの収集を行う。「Selenium」の特徴としてクリックなどのユーザーアクションを模倣できるというものがある。本研究では 2.2 節の経済情報メカニズムを参考に、またスクレイピングによるデータ収集が可能な金融時系列と経済情報について分析を行うこととした。経済時系列情報は主に「日本銀行時系列検索サイト」、金融時系列情報は「日経プロフィル」から取得する。取得したデータについては表 4.1 にまとめる。これらの単位は全て円である。

データの前処理について

VAR-LiNGAM は時系列を考慮し、因果探索を行う手法であるので、時系列データを入力として用いる必要がある。その際にはデータの前処理が必要になってくる。そのため、ここでは今回の研究で用いたデータの前処理手法について説明する。

時間足の設定

VAR-LiNGAM には多変量の時系列データを目的変数、あるいは説明変数として用いる必要がある。その際、変数によって時間足が異なるモノが存在する。例えば、株価などの情報は日々更新されており、日足でデータを取得することが可能である。しかし、金利やマネタリーベースなどのデータは月足であったり場合によっては四半期足、年足になってしま

表 4.1: 日経業種間データ

データ項目	単位	データ項目	単位
水産	商社	窯業	海運
鉱業	小売業	鉄鋼	空運
建設業	銀行	非鉄・金属	倉庫
食品業	その他金融	機械	通信
繊維業	証券	電気機器	電力
パルプ・紙業	造船	ガス	サービス
科学業	保険	自動車	ゴム
医薬品	不動産	輸送用機器	陸運
石油	鉄道・バス	精密機器	その他製造

こともある。本研究ではそういった変数が存在したときに一ヵ月間同じ値をとり続けると仮定し、日足のデータに揃えるようにした。また、市場データの殆どのものは取引が行われる平日のみのデータしか存在しない。そのため、土日祝日の時間にはデータが入らないようデータセットを作成する必要がある。Python のライブラリ「pandas-market-calender」を用いて土日祝日はデータセットから除外する。「pandas-market-calender」では各国の金融市场の営業日や取引時間などに関するデータが提供されている。それによって日足にそろえられたデータの例を図に示す。なおデータは日本銀行時系列検索サイトから取得したものと示す。

データの正規化

本研究で扱う目的変数および説明変数はそれぞれの値の間に大きな差があり、取得したデータそのままで VAR-LiNGAM による因果探索を行うと明らかに大きい値が生まれてしまう。例えばマネーストックは基本的に業種別株価などと比べて圧倒的に大きい値をとる。これをそのまま因果探索するデータとして用いてしまうと、因果の強さが他のモノに比べて圧倒的に大きく推定されてしまう。それにより、本来大きいはずの因果関係が小さいものとして推定されてしまう、あるいはその逆に小さい因果関係が大きいものとして推定されてしまうことが懸念される。そのため、本研究では因果探索を行う前のステップとしてデータの正規化を行う。データの正規化を行うことで異なる尺度や範囲を持つデータの範囲を一定の値に統一することができる。正規化の手法として、今回はデータの最小値を 0、最大値を 1 にスケーリングする Min-Max normalization を採用する。Min-Max normalization の定式化を以下に示す。

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4.1)$$

これらの手法によって処理されたデータと処理前のデータを図に示す。

VAR-LiNGAM による因果探索

Python のモジュール「LiNGAM」を用いて VAR-LiNGAM による時系列を考慮した因果探索を行う。VAR-LiNGAM によって求められた因果性のデータ形態、また因果性を 2 次元ネットワークとして出力したものを図に示す。



図 4.5: 波及メカニズム

§ 4.3 経済データの3Dグラフによる可視化

3D グラフによる可視化では、ネットワーク構造を視覚的に理解しやすくすることができる。また、異常な構造や特異なパターンを素早く発見することができる。本研究では有向重み付き 3D グラフによる因果性の可視化を行うことで変数間の因果性を直感的に理解することができるようとする。3D グラフの作成には Python のライブラリである「Pyvis」を用いる。「Pyvis」は主にネットワークやグラフの可視化に特化しており、対話的なグラフ・ネットワークを作成し、ブラウザ上でそれを表示することができる。また、グラフにクリックアクションを追加し、クリックしたノードの過去の値、RF5 によって得られた同定式を確認することができるようするために Python と JavaScript を繋げ、システムを作成した。作成したグラフとページ遷移については図に示す。「pyvis」によるグラフネットワークの作成をするには csv ファイルの読み込みを行う。csv ファイルは VAR-LiNGAM を実行した段階で作成しており、これによって作成されたグラフが図であり、VAR-LiNGAM を動作させる際に表示される 2D グラフと比べると比較的理 解しやすくなつた。VAR-LiNGAM による因果探索では大小関係なくほとんど全ての場合に因果性を見出してしまう、あまりにも因果性が弱い場合でもエッジを作成し、ネットワークが見づらくなってしまう。そのため、今回のグラフ作成では正の因果性を持つ影響の大きいもの、負の因果性を持つもので影響の大きいものを抽出し、グラフを作成する。具体的には正方向に因果性を持つものから大きなものから順に 120 個、負の因果性を持つものから影響が大きいものを 120 個抽出することとした。



図 4.6: 可視化システムの流れ



図 4.7: 可視化に用いたデータ

実験結果並びに考察

§ 5.1 実験の概要

本章では、実際の経済データを用いて、4章で述べた手法で経済変数の因果性を導出、経済変数のモデル化を行い、その有効性の確認および考察を行う。また、因果性とモデルを用いて経済動向を可視化した3Dグラフを作成し、考察を行う。数値実験では VAR-LiNGAMによる分析の結果と RF5によるモデル化の結果を述べ、システム面についての実験でネットワークによる可視化の有意性について述べていく。

実験で扱うデータ

実験に用いるデータについて述べる。今回の実験で用いるデータは日経プロフィルが公開している株価時系列データと日本銀行時系列検索サイトで公開されている金融データを用いる。株価時系列データでは日経33業種株価データを用いることとした。金融データには、データ項目として金利、マネーストック、マネタリーベース、国内物価指数(総平均)が含まれている。これらのデータをスクレイピングを用いることで収集し、

実験では時間足を日足に設定し、分析を行った。日足で実験を行う際には月足データが1ヶ月間同じ値をとりつづけると仮定する。今回は経済データ、金融データの2つを合わせて40個になった。今回の実験では2021年1月4日から2023年12月29日までのデータを用いた。なお、今回使用するレコードについては、土日祝日に株式市場が空いていないということを考慮して土日祝日はデータから除外し、サンプル数は735となった。データの詳細については表に示す。RFの実験で扱うデータについては上で作成したデータセットから日付のみを除外したものになる。

RFにおいて

§ 5.2 実験結果と考察

VAR-LiNGAMによる分析の結果

検証用のデータを用いて、導出した因果関係の結果を図に示す。

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



図 5.2: 可視化された MACD,STOCH,ATR

図 5.1: EMA, BBAND, TRENDLINE

おわりに

謝辞

本研究を遂行するにあたり、多大なご指導と終始懇切丁寧なご鞭撻を賜った富山県立大学電子・情報工学科情報基盤工学講座の奥原浩之教授, Antònio Oliveira Nzinga Renè 講師に深甚な謝意を表します。最後になりましたが、多大な協力をして頂いた、奥原研究室の同輩諸氏に感謝致します。

2023年2月

清水 豪士

参考文献

- [1] 財務省, “ファイナンス 2022年 6月号 No679”, 2022
- [2] 野村證券金融工学研究センター クオンツ・ソリューション・リサーチ部 饗場 行洋,
山本 裕樹, “データサイエンスと新しい金融工学”, 2018.4
- [3] 木下大輔, “市場間データを活用した高頻度データに対するパラメータ選択と最適なス
トラテジー構築”, 富山県立大学学位論文, 2022
- [4] 平林 明憲, “遺伝的アルゴリズムによる外国為替取引手法の最適化”, 人口知能学会
全国大会論文集 第 22 回
- [5] 山本, 拓「経済における時系列分析：概観と展望」『日本統計学会誌. シリーズ J』35(2).
国立国会図書館デジタルコレクション <https://dl.ndl.go.jp/pid/10823042> (参照 2024-
01-18)
- [6] 金融政策の波及メカニズム