

< 修 士 論 文 >

論文題目

コロナ禍における EBPM に資する
倒産・失業関連指標の予測速報モデル

滋 賀 大 学 大 学 院
デ ー タ サ イ エ ン ス 研 究 科
デ ー タ サ イ エ ン ス 専 攻

修了年度：2020 年度

学籍番号：6019105

氏 名：川上 幹男

指導教員：梅津 高朗

提出年月日：2021 年 1 月 20 日

目次

第1章 序論	1
1.1 研究背景と目的	1
1.2 方法	1
1.3 関連研究	4
1.4 本論文の構成	4
第2章 倒産件数予測	5
2.1 使用データ	5
2.1.1 企業倒産件数データ	5
2.1.2 日本銀行短観データ	7
2.1.3 実質実効為替レート	9
2.1.4 貸出約定平均金利	9
2.1.5 円インデックス	9
2.1.6 長期プライムレート	10
2.1.7 国債利回り	10
2.2 推定方法	10
2.2.1 倒産件数予測モデル構築の問題点	10
2.3 サブモデル	11
2.3.1 貸出約定平均金利サブモデル	12
2.3.2 実質実効為替レートサブモデル	13
2.3.3 倒産件数サブモデル	18
2.4 推定結果	21
2.4.1 説明変数の選択結果	21
2.4.2 倒産件数1 四半期先予測モデル	22
2.4.3 倒産件数2 四半期先予測モデル	24
2.5 予測結果	26
2.5.1 倒産件数1 四半期先予測結果	26
2.5.2 倒産件数2 四半期先予測結果	27
2.6 予測の要因分解	28
2.7 結果	28
第3章 完全失業率予測	30
3.1 使用データ	30
3.1.1 完全失業率データ	30
3.2 推定方法	31
3.2.1 完全失業率予測モデル構築の問題点	31
3.3 サブモデル	32
3.3.1 雇用人員 DI サブモデル	32
3.3.2 完全失業率サブモデル	35

3.4	推定結果	37
3.4.1	説明変数の選択結果	37
3.4.2	完全失業率1四半期先予測モデル	38
3.4.3	完全失業率2四半期先予測モデル	41
3.5	予測結果	43
3.5.1	完全失業率1四半期先予測結果	43
3.5.2	完全失業率2四半期先予測結果	44
3.6	予測の要因分解	45
3.7	結果	45
第4章	就業者数予測	47
4.1	業種別の就業者推移	47
4.2	飲食・宿泊サービス業就業者数予測モデル	48
4.3	使用データ	48
4.3.1	就業者数データ	48
4.3.2	住宅宿泊事業者数データ	49
4.3.3	民泊解禁効果データ	49
4.3.4	季節ダミーデータ	50
4.4	推定方法	50
4.4.1	飲食・宿泊サービス業就業者数予測モデル構築の問題点	50
4.5	サブモデルおよびシナリオ分析	50
4.5.1	生産設備DIサブモデル	51
4.5.2	飲食・宿泊サービス業就業者数サブモデル	53
4.5.3	住宅宿泊就業者数シナリオ予測	56
4.6	推定結果	57
4.6.1	説明変数の選択結果	57
4.6.2	飲食・宿泊サービス業就業者数1四半期先予測モデル	58
4.6.3	飲食・宿泊サービス業就業者数2四半期先予測モデル	60
4.7	予測結果	62
4.7.1	飲食・宿泊サービス業就業者数1四半期先予測結果	62
4.7.2	飲食・宿泊サービス業就業者数2四半期先予測結果	63
4.8	予測の要因分解	64
4.9	結果	64
第5章	考察と今後の課題	66
	謝辞	68
付録A	補論	69
A.1	重回帰分析	69
A.1.1	回帰分析の基礎概念	69
A.1.2	系列相関	70
A.1.3	多重共線性	72
A.1.4	影響点	73
A.2	一般化最小二乗法 (GLS)	76
A.2.1	GLSによるパラメータ推定	76

A.2.2	実行可能な GLS	77
A.3	頑健回帰推定	78
A.3.1	M 推定	79
A.3.2	M 推定量の不偏性と漸近的特性	82
A.3.3	崩壊点	83
A.3.4	崩壊点と調整定数	84
A.3.5	σ の推定	84
A.3.6	MM 推定	85
A.4	季節調整法	86
A.4.1	X-12-ARIMA	86
	参考文献	87

目 次

2.1	TDB 倒産件数の推移 (四半期)	6
2.2	TDB 倒産集計方式変更	6
2.3	補正後 TDB 倒産件数の推移 (四半期)	7
2.4	1 四半期先モデルのデータ時点制約概要	10
2.5	2 四半期先モデルのデータ時点制約概要	10
2.6	倒産件数予測モデル構築概要	11
2.7	季節調整済み貸出約定平均金利の月次推移	12
2.8	内外インフレ率差の月次推移	14
2.9	AR 特性方程式の固有値プロット	16
2.10	MA 特性方程式の固有値プロット	16
2.11	名目実効為替レートの月次推移	16
2.12	VAR 特性方程式の固有値プロット	21
2.13	倒産件数 1 期先モデル L-R プロット	23
2.14	倒産件数 1 期先予測モデル MM 推定の各ウェイト	23
2.15	倒産件数 2 期先モデル L-R プロット	25
2.16	倒産件数 2 期先予測モデル MM 推定の各ウェイト	25
2.17	倒産件数 1 期先モデルの予測結果	27
2.18	倒産件数 2 期先モデルの予測結果	27
2.19	倒産件数 1 期先モデルの要因分解図	28
2.20	倒産件数予測推移	29
3.1	完全失業率の推移 (四半期)	31
3.2	完全失業率予測モデル構築概要	32
3.3	雇用人員 DI の四半期推移	32
3.4	VAR 特性方程式の固有値プロット	34
3.5	AR 特性方程式の固有値プロット	37
3.6	MA 特性方程式の固有値プロット	37
3.7	完全失業率 1 期先モデル L-R プロット	39
3.8	完全失業率 1 期先予測モデル MM 推定の各ウェイト	40
3.9	完全失業率 2 期先モデル L-R プロット	42
3.10	完全失業率 2 期先予測モデル MM 推定の各ウェイト	42
3.11	完全失業率 1 期先モデルの予測結果	44
3.12	完全失業率 2 期先モデルの予測結果	45
3.13	完全失業率 1 期先モデルの要因分解図	45
3.14	完全失業率予測推移	46
4.1	業種別前年同月比	47
4.2	業種別前年同月数	48

4.3	飲食・宿泊サービス業就業者数の推移(四半期)	49
4.4	飲食・宿泊サービス業就業者数予測モデル構築概要	50
4.5	生産設備DIの四半期推移	51
4.6	VAR特性方程式の固有値プロット	53
4.7	VAR特性方程式の固有値プロット	56
4.8	住宅宿泊事業者数シナリオ予測	57
4.9	飲食・宿泊サービス業就業者数1期先モデルL-Rプロット	59
4.10	飲食・宿泊サービス業就業者数1期先予測モデルMM推定の各ウェイト	59
4.11	飲食・宿泊サービス業就業者数2期先モデルL-Rプロット	61
4.12	飲食・宿泊サービス業就業者数2期先予測モデルMM推定の各ウェイト	61
4.13	飲食・宿泊サービス業就業者数1期先モデルの予測結果	63
4.14	飲食・宿泊サービス業就業者数2期先モデルの予測結果	64
4.15	飲食・宿泊サービス業就業者数1期先モデルの要因分解図	64
4.16	飲食・宿泊サービス業就業者数予測推移	65
A.1	サンプルデータの散布図	76
A.2	L-Rプロットの例	76
A.3	Ψ 関数のプロット例	81
A.4	ウェイトのプロット例	81

第1章 序論

1.1 研究背景と目的

新型コロナウイルス (COVID-19) は、初の症例が2019年末に中国で確認されて以来、短期間で世界中に感染が拡大した。こうした中、殆どの国で感染拡大防止を目的とした渡航制限や外出制限等が実施されたことにより、人流・物流が大幅に抑制され、世界経済は急激な後退に見舞われた。こうした中、コロナ感染第1波収束後、日本を含む殆どの国は感染拡大防止と経済活動維持の両立を企図した政策を採用している。すなわち、感染状況が拡大すると、緊急事態宣言や営業時間短縮要請など人流・物流の抑制策を実施する。その結果、経済活動は停滞を余儀なくされ、倒産増大や失業増加を招き、生活困窮に伴う自殺者増加にも繋がる。そこで、感染状況が落ち着くと、人流・物流制限を緩和しGoToキャンペーンのような経済政策を実施する。しかしその結果、人的接触増加が感染を再拡大する。つまり、両立を企図した感染拡大防止と経済活動維持は、トレードオフの関係にあり、両立は非常に困難である。従って、両立政策の実施には政策決定過程の課題として昨今指摘されている evidence based policy making (EBPM) が必須である。

ここで、感染拡大防止と経済活動維持を両立させる政策決定 (EBPM) の一連の過程を単純化して説明する。まず、第1段階として、感染拡大と経済問題の何れが深刻であるのかを見極めるために、両問題が現状のまま推移した場合、この先どの程度まで悪化し得るかを見極めるため、能うる限り正確な予測を行う。何れの問題がより深刻であるか見極めた後、第2段階として、同問題が感染拡大あるいは経済問題における全体的問題であるのか、局所的問題であるのかを見定め、局所的問題の際は問題の発生箇所を特定する。次に、第3段階として、同問題への対策を起案するため、同問題の主な原因を特定する。最後に、第4段階として、分析結果に基づき、対策を具体化し政策立案を行い、決議する。残念ながら、現状の日本の政策決定過程ではEBPMに資する分析が十分に実施されているとは言い難いのが実情である。

EBPMに資する分析とは、次の要件を満たすものと考えられる。(1) 正確性、EBPM第1段階では可能な限り予測が正確であることを要求する。(2) 速報性、政策効果の顕在化には一定の期間を要することを考慮すると、EBPMの全段階は可能な限り迅速に実施する必要がある。(3) 説得力、EBPM第3段階では、第1・2段階で特定した問題の主たる原因を特定できることが望ましい。(4) 政策示唆力、EBPM第4段階を速やかに実施するため、問題の主たる原因が政策上制御可能なであることが望ましい。

本研究では経済問題の内、特に深刻な影響が懸念されている倒産・失業問題に注目し、EBPM過程に資する倒産・失業関連指標予測モデルの開発を目的とする。

1.2 方法

上記の研究目的を満たす予測モデルの開発には、重回帰分析を用いる。昨今、予測を目的とした分析モデルの作成には、SVMやニューラルネットに代表される機械学習分野の手法が選択されることが多い。殆どの機械学習手法は、予測が改善するようにデータから反復

的に学習を行うことで高い予測精度を実現する。こうした機械学習手法は、EBPMに資する分析の要件の内、正確性と速報性を満たすことが想定される。しかし、殆どの機械学習手法は、高い予測精度を達成するため、学習データのあらゆる情報を組み合わせて予測を行うので、人間が機械学習の予測の過程を解釈するのが困難な場合が少なくない。従って、機械学習手法はEBPMに資する分析の要件のうち、説得力と政策示唆力を満たすことが難しい。

一方、重回帰分析は目的とする変数の予測だけでなく目的変数と説明変数間の関係性を重回帰式の係数から捉えることができる。本研究では、予測のための情報量基準 AIC だけでなく説明変数に経済理論と整合的な符号条件(例：倒産件数と業況悪化には正の相関関係)を課してモデル選択することにより、説得力を確保する。さらに、政策上制御可能な説明変数を予測に利用することで政策示唆力を担保する。また、高い決定係数の追求や予測区間幅を持った予測により正確性も保証する。

ただ、古典的分析手法である重回帰分析は的確な推定が困難な手法であることに留意されたい。これは、OLS が誤差項の均一分散、系列無相関、説明変数間の低相関、影響点の非存在という理想的諸仮定の下でのみ正当化される推定量であり、これらの仮定を適切に検定し、帰無仮説が棄却された場合は、OLS とは異なる適切な推定量を導出する必要があるからである。残念ながら、適切な検定と、帰無仮説棄却時の適切な推定量の導出が行われていない実証分析が散見されるので、典型例を示しておく。誤差項の系列相関の対処では、まず、DW 検定を用いて誤差項の系列相関有無の判断を行う。従来の方法では、DW 分布の棄却臨界値は分析ごとに計算を要するため、判定不能域を持つ DW 統計量の分布表を用いて検定を行う。従って、判定不能域に落ちた場合、判定できないという欠点がある。また、系列相関が検出された場合も、データ数を 1 期減少させる非効率なコクラン=オーカット法が重回帰分析の推定に用いられることが多い。多重共線性の対処では、まず、VIF により検出を行う。VIF により多重共線性が発見された際の最も非効率な対処法は恣意的な変数除去である。これは分析目的にとって重要な変数を分析の枠組みから削除してしまう恐れがある。恣意的な変数除去を避けて多重共線性に対処する方法としては、L2 正則化による Ridge 推定が用いられる。ただし、Ridge 推定量は不偏性を大きく損なう恐れがあり対処としては必ずしも適切ではない。また、恣意性を軽減した変数除去方法として、L1 正則化による Lasso 回帰が用いられるが、こちらは過度な変数除去を行う恐れがある。影響点(外れ値と高い作用点)の対処法では、まず、クックの D が 1 以上の基準で影響点を検出される。しかし、蓑谷 [38] によるとこの検出基準では影響点をほぼ検出できないことが指摘されており適切な方法とはいえない。影響点が存在する場合の最も非効率な対処法は恣意的な影響点の除去である。これは分析上、重要なデータ点を分析の枠組みから削除してしまう恐れがある。恣意的なデータ点除去を避けて影響点に対処する方法としては、頑健推定(M 推定)を用いて外れ値が推定結果に及ぼす影響を軽減させる。しかし、頑健推定は、影響点の内、外れ値(Y 方向の影響点)にした対処しない方法であり高い作用点(X 方向の影響点)が及ぼす影響には対応していない。こうした不適切な重回帰分析が不正確な予測を導いている例が散見される。重回帰分析は正確性の観点で機械学習手法に劣ると見られているが、それは、こうした多くの不適切な重回帰分析によりもたらされている面が少なくない。

しかし、近年の重回帰分析の理論的發展と R 言語を用いた最新手法のパッケージプログラム開発という技術的發展によって、より適切で高度な重回帰分析が手軽に行えるようになってきている。本研究では、それら発展による恩恵を十全に享受した重回帰分析方法により予測モデルを開発する。

本研究で行った重回帰分析を具体的に説明する。まず、誤差項の系列相関への対処には、計算機上で DW 分布の棄却臨界値を分析ごとに計算し判定不能域を持たない DW 検定を行う。

DW 検定の帰無仮説が棄却された場合は、コ克蘭=オーカット法を改善した Prais-Winsten 変換による一般化最小二乗法 (GLS) により推定を行う。この推定は、コ克蘭=オーカット法におけるデータ数を 1 期減少させる欠点を克服する。多重共線性は、VIF を用いて検出する。多重共線性が発見された際は、不偏性を満たさない L2 正則化と過度な変数除去を行う恐れのある L1 正則化の両欠点を克服する Elastic Net を用いて対処する。影響点は、クックの D が 1 以上よりも適切な検出方法として蓑谷 [38] で述べられている修正されたクックの D とその切断点を用いて検出を行う。さらに、L-R プロットを描画することで検出された影響点が適切であるか確認する。影響点が検出された場合は、頑健推定を拡張し高い作用点 (X 方向の影響点) による推定結果への影響を軽減する有界影響頑健推定 (MM 推定) を用いる。

上記で説明した本研究の重回帰分析の各手法は R 言語の関数として提供されている。しかし、系列相関に対処するために GLS を行った後、影響点への対処として MM 推定を行いたい場合のように複数の問題へ同時に対処が必要となった時、公開されている R 言語の関数では実行できない。従って、本研究では複数の手法を組み合わせる重回帰分析が行えるように自作プログラムをコーディングしている。

こうした適切な重回帰分析に基づく予測モデルは正確性、説明力、政策示唆力を満たすことが期待できる。残る速報性の要件を満たすには 1 点問題が残る。それは、予測モデル開発に利用する大半のマクロ経済指標は翌月もしくは翌々月に公表される。それに伴って予測が遅れることで速報性が満たされなくなる。例えば、四半期ごとの予測で 2021 年第 1 四半期の予測には、2020 年第 4 四半期までのデータが必要だが 2020 年 12 月データは翌月公表となりデータを待つ必要が発生する。そこで本研究では、重回帰分析による予測モデルをメインモデルとした上で、補助的に時系列分析によって公表が遅れる変数を予測し補完するサブモデルを開発し据え置く。

サブモデルの開発は、まず、対象の変数を ARIMA(p,d,q) モデルで選択する。モデル選択の基準には、予測を考慮して情報量基準 AIC で行う。もしも、AIC 基準で AR(p) 過程が選択された場合は、さらなる予測性能の追求を目的に他の変数系列を組み入れた VAR(p) モデルを推定する。VAR(p) モデルも情報量基準 AIC でモデル選択を行う。

昨今、時系列分析も重回帰分析と同様に理論的發展と R 言語を中心とした技術的發展に支えられ高度な分析を行うことが可能となっている。まず、誤差項の自己相関の有無を検出するためには、Ljung-Box 検定を行う。Ljung-Box 検定のラグ選択には、沖本 [42] で提案されているデータサンプル数からモデル次数を引いた値の対数値を採用する。次に誤差項が独立同一分布に従うか検出するためには、転換点に関する検定を用いる。こちらは、誤差項に現れる転換点の数が独立同一分布を仮定した誤差項と整合的であるか検定を行う手法である。時系列モデルが AR 過程を含む場合は、定常性のチェックが必要となる。AR 過程の定常性は、AR 特性方程式の同伴行列の固有値が単位円内に収まる場合、定常性は満たされる。従って、同伴行列の固有値を算出し、単位円に描画する。時系列モデルが MA 過程を含む場合は、反転可能性のチェックが必要となる。MA 過程の反転可能性は、AR 過程の定常性チェックと同様に同伴行列の固有値を算出し、単位円に描画することで行える。固有値が単位円内に収まる場合、反転可能性は満たされる。これらの時系列分析は、R 言語の関数として実装されており、時系列モデルの仮定に従ったサブモデルが開発できる。

以上より、適切な重回帰分析に基づく予測モデルをメインモデルとし、適切な時系列分析に基づくモデルをサブモデルとした予測モデルを開発する。

1.3 関連研究

まず、倒産予測の関連研究をいくつか紹介する。Chen[20]は複数の倒産予測モデル研究についてまとめ各提案手法を台湾証券取引所から集めた企業データに対して適用し比較検討を行なっている。Chen[20]によると、倒産予測モデルの研究は統計手法と人工知能手法に大きく分けられる。統計手法としては、1変量ごとに財務比率の倒産予測能力を検討した研究[18]が嚆矢とされており、その後、判別関数(discriminant function)を用いて5つの財務比率を変数とした倒産判別モデルを構築した研究[17]がある。線形判別分析(LDA)の他に倒産予測モデルとしてはロジスティック回帰を用いた研究[29]がある。人工知能手法としては、決定木、自己組織化マップ(SOM)とLVQを用いたニューラルネットによる分類、遺伝的アルゴリズム、PSOアルゴリズム、SVMが検討されている[20]。Chen[20]によると、台湾の個別企業倒産予測にはマクロ指標や非財務比率のデータが予測に寄与せず、主に財務比率データが予測精度に寄与すると述べられている。また、統計的モデルや決定木は遺伝的アルゴリズムと比較して長期予測において予測精度が下がる傾向にあると報告されている。但し、これらは個別企業の倒産予測を行なった結果であり倒産件数を予測するモデル構築に関するものではない。倒産の定義は企業や国ごとに様々なものが存在する。本稿では、帝国データバンク社の倒産の定義に従った倒産件数の予測を日銀短観や為替レートのようなマクロ指標を用いて行う。

同様に失業率予測の関連研究をいくつかまとめる。失業率予測を実証分析に基づき行なった研究は国内ではあまり見られず海外事例も踏まえて紹介する。失業率予測の関連研究は、使用データを改善した研究とモデリング方法を改善した研究に大きく分けられる。使用データの改善としては、消費者調査のような市民の今後の景況感の期待を捉える質的データを検討したもの[22]やGoogleの検索クエリのデータを組み入れたもの[30][34][24]が存在し、いずれも失業率予測精度の改善に有用であることが示唆されている。モデリング方法の改善としては、オークンの法則やフィリップス曲線などの経済理論に基づきGDP成長率、失業率、インフレ率から状態空間モデルを用いて予測した研究[21]、複数の時系列モデル(ARIMA, SARIMA, TAR, MSA, VAR)を用いて予測を試みた研究[32]、失業率推移の線形と非線形性をそれぞれARIMA, ARNNモデルで表現するアプローチを提案した研究[19]、線型性を考慮した時系列モデルと非線形性を機械学習手法(ニューラルネットワーク, SVM, スプライン回帰)で考慮した研究[28]などが挙げられる。これら手法は、予測精度の向上(正確性)に重きが置かれた研究となっており、本稿の研究では予測に加えて説明力も重視しEBPMに資する要件を満たすモデルを提案している点がこれら関連研究との大きな相違点である。

1.4 本論文の構成

第2章で、倒産件数予測の使用データと推定方法および予測結果を説明する。第3章で、完全失業率で同様の結果を述べる。続く第4章で、業種別就業者数の前年同期比較から特にコロナ禍で影響が大きかった業種を見極めた後、当該業種の就業者数について同様の結果を述べる。最後に第5章で結論と今後の課題をまとめる。

第2章 倒産件数予測

企業は、景況感が悪化すると倒産する前にまず従業員の解雇を行なうことが想定され、倒産件数は倒産・失業問題の深刻さを表すと言える。そこで、まず、企業倒産件数を対象とした予測モデルを開発する。本章では、コロナ禍での企業倒産件数予測モデルについて述べる。はじめにモデル作成に用いたデータの詳細を説明し、推定結果および第2四半期先までの予測結果、推定量の要因分解による予測に寄与する要因の考察を行う。

2.1 使用データ

2.1.1 企業倒産件数データ

データ概要

本研究では、(株)帝国データバンク社(以下、TDB)が毎月公表している倒産集計を利用する。この倒産集計(以下、TDB倒産件数)は、倒産4法(会社更生法、民事再生法、破産法、特別生産)による法的整理を申請した負債額1,000万円以上の法人および個人経営を対象としている[13]。但し、任意整理(銀行取引停止、内整理など)は集計対象に含まない。なお、倒産は法律用語でないため、定義が曖昧な用語である。そこで本稿では、帝国データバンク社の倒産企業の定義に従うこととする。以下に帝国データバンク社が集計している倒産企業の定義をまとめる[12]。

- 銀行取引停止処分を受ける
- 内整理する
- 裁判所に会社更生手続き開始を申請する
- 裁判所に民事再生手続き開始を申請する
- 裁判所に破産手続き開始を申請する
- 裁判所に特別清算開始を申請する

倒産件数予測モデルでは、このデータを予測対象としてモデルを構築する。以下に四半期ごと2000年第1四半期(2000-Q1)から2020年第3四半期(2020-Q3)TDB倒産件数推移を示す。

これは帝国データバンク社が毎月発行する全国企業倒産集計の月次データを四半期ごとに足し上げ作成した。なお、シャドー部分は内閣府が発表する景気基準日付[6]を元にした景気後退局面である。2018年第4四半期(2018-Q4)以降の景気後退局面は暫定であることに注意されたい。

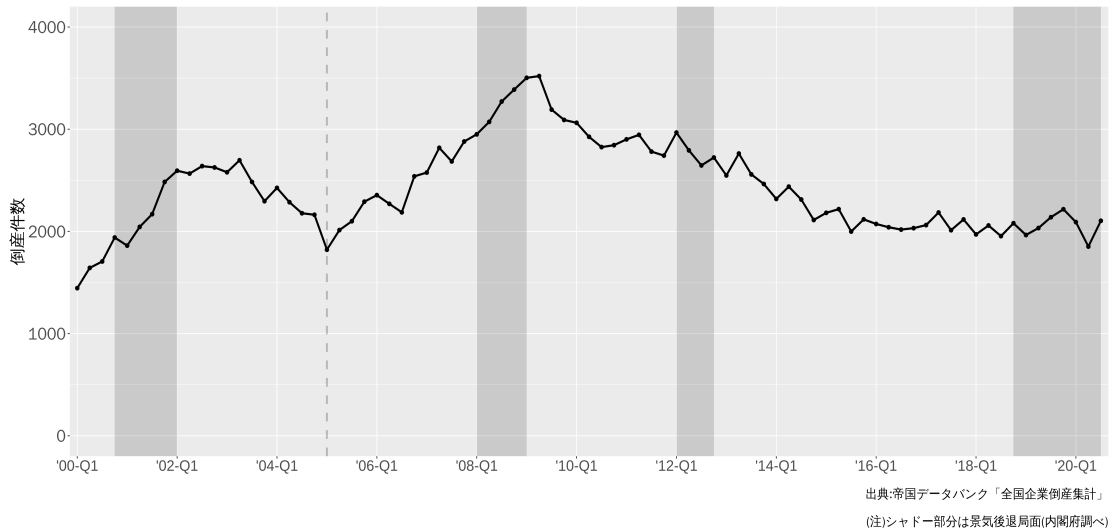


図 2.1: TDB 倒産件数の推移 (四半期)

倒産件数データの問題点

TDB 倒産件数を分析に使用するに当たり、2点問題が発見された。そこで問題点とその対処法について言及する。

問題点 1：2005 年 4 月を境に集計方式の変更

1つ目は、TDB が公表する倒産集計方式が 2005 年 4 月以前と以降で変更された点である [10]。

変更前は、倒産集計の対象が任意整理と法的整理であった。一方で変更後は、法的整理に加えて任意整理から法的整理への移行が判明したものを集計している。

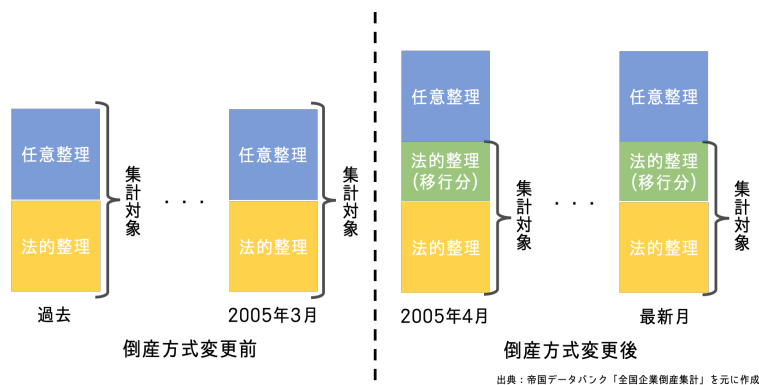


図 2.2: TDB 倒産集計方式変更

図 2.1 における 2005 年第 1 四半期の縦点線は集計方式が変更された時点を表している。図 2.1 の集計方式変更時点 2005 年第 2 四半期以降を見ると景気拡大下で倒産件数が増大していることが分かる。本研究の倒産件数予測モデル作成においては、この期間をデータ計測の機能不全とみなし重回帰モデルのパラメータ推定対象期間を 2007 年第 3 四半期以降とする。

問題点 2：2020 年 5 月の倒産件数過少計測

現時点での 2020 年 1 月から 11 月までの倒産集計結果は、表 2.1 となっている。

表 2.1: 2020 年度 TDB 倒産集計

月	1	2	3	4	5	6	7	8	9	10	11
倒産集計件数	713	634	744	758	288	806	847	655	602	647	563

ここで 2020 年 5 月の報告が 288 件と他月と比較して少なくなっている。原因を帝国データバンク社の社員に聞き取りを行ったところ、コロナ禍による裁判所業務縮小に伴う法的整理数の減少と推測されるとの回答をいただいた。従って、本研究では、5 月に計測されなかった倒産件数は 6 月、7 月の件数に法的整理 (移行分) として反映されていると仮定し 2020 年 5 7 月の 3 ヶ月間の平均値で補正した。補正後の倒産件数は、表 2.2 となる。

表 2.2: 2020 年度 TDB 倒産集計 (補正後)

年月	1	2	3	4	5	6	7	8	9	10	11
倒産集計件数	713	634	744	758	647	647	647	655	602	647	563

以上より、分析対象とする 2007 年第 3 四半期以降の補正後倒産件数の推移は図 2.3 となる。

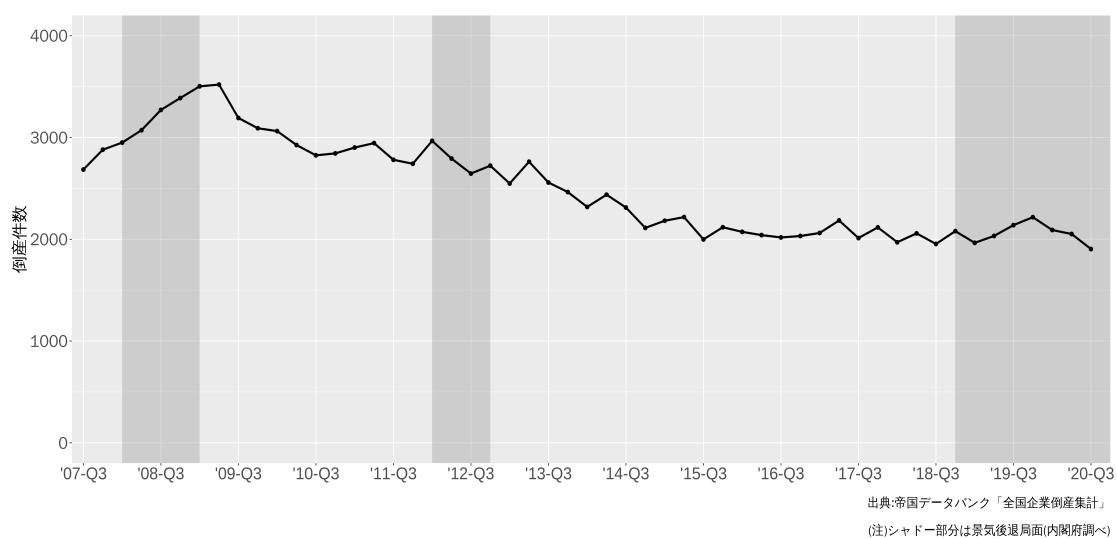


図 2.3: 補正後 TDB 倒産件数の推移 (四半期)

図 2.3 の倒産件数推移から、世界金融危機時の高水準とコロナ禍での低水準傾向が読み取れる。

2.1.2 日本銀行短観データ

日本銀行短観 (全国企業短期経済観測調査) とは、日本銀行調査統計局が四半期ごとに作成している統計調査である [5]。その目的は、全国の企業動向を把握し、金融政策の適切な運営に資することとされている。

調査対象としては、「金融機関」および「経営コンサルタント業、純粋持株会者」を除いた全国の資本金2,000万円以上の民間企業の中から調査対象企業を抽出している。公表時期は、毎年4月、7月、10月、12月にそれぞれ当該月の調査結果を公表している。

13項目の設問について、3つの選択肢から回答企業が現状に最も近い番号を選ぶ形式でオンラインまたは書面で集計される。

回答結果は、ディフュージョン・インデックス(以下、DI)と呼ばれる指標に加工集計し公表される。DIの定義式は以下である。

$$DI(\%ポイント) = \text{第1 選択肢の回答社数構成比}(\%) - \text{第3 選択肢のお回答社数構成比}(\%)$$

本研究で日本銀行短観データを用いる際は、このDI指数を使用する。以降に、予測モデル作成に用いた日銀短観データの項目とその設問詳細を述べる。また、表2.3に各設問の選択肢項目の詳細を表形式でまとめる。

業況 DI

回答企業の収益を中心とした、業況についての全般的な判断。

雇用人員 DI

回答企業の雇用人員の過不足についての判断。

資金繰り DI

回答企業の手元流動性水準、金融機関の貸出態度、資金の回収・支払い条件などを統合した資金繰りについての判断。

貸出態度 DI

回答企業からみた金融機関の貸出態度についての判断。

生産設備 DI

回答企業の生産設備、営業用設備の過不足についての判断。

販売価格 DI

回答企業の主要製商品の販売価格、または主要サービスの提供価格についての判断。

仕入価格 DI

回答企業の主要原材料購入価格(外注加工費を含む)または主要商品の仕入れ価格についての判断。

表 2.3: 日銀短観設問選択項目の詳細

	1	2	3
業況 DI	良い	さほど良くない	悪い
雇用人員 DI	過剰	適正	不足
資金繰り DI	楽である	さほど苦しくない	苦しい
貸出態度 DI	緩い	さほど厳しくない	厳しい
生産設備 DI	過剰	適正	不足
販売価格 DI	上昇	もちあい	下落
仕入価格 DI	上昇	もちあい	下落

2.1.3 実質実効為替レート

実質実効為替レートは、日本銀行調査統計局が作成し月次で公表している [3]。これは、貿易財の対外競争力、すなわち、一国の対外競争力を「実質」と「実効」を考慮に入れた為替レートで計った指標である。

「実質」とは、名目の為替レートを自国と競合国の製品価格で調整することである。また、「実効」とは、複数の為替レートを加重平均することである。日本銀行は、国際決済銀行 (Bank for International Settlements, BIS) 公表の Broad ベースの実効為替レートを利用している。BIS は、Narrow ベースと Broad ベースの 2 種類の為替レートを公表している。Broad ベースとは、為替レートを実効化する際にどの国の通過を対象にするかというカバレッジが広い方を指す [8]。2020 年 12 月時点では、Broad ベースの対象国は 61ヶ国となっている [1]。

2.1.4 貸出約定平均金利

貸出約定平均金利は、日本銀行金融機構局金融データ課預貸金統計グループが作成し月次で公表している [4]。これは、銀行や信用金庫が個人や企業に資金を貸す際の金利を平均した指標である。金利の面から借入環境を計る指標として予測モデル構築に利用する。本研究で使用した、貸出約定平均金利は、「国内銀行」で「新規」かつ貸出期間が「総合」もしくは「長期」のものを対象とした。

「国内銀行」とは、銀行本体の設立根拠が日本の銀行法に準拠している銀行のうち、日本銀行と当座預金取引契約をしている銀行を指す。「新規」とは、当該月末貸出残高のうち当月中に実行した貸出を指す。「総合」とは、全ての期間の貸出を対象とする。「長期」は、約定時の貸出期間が 1 年以上のものを集計対象としたものである。また、銀行勘定の円貸出のうち金融機関向けの貸出を除外したものが集計対象となる。他の銀行貸出金利動向を把握するデータのとしては、後述する長期プライムレートがあげられる。

2.1.5 円インデックス

円インデックスは、名目実効為替レートとも呼ばれており相対的な通貨の実力を計るための総合的な指標である [2]。

作成方法は、Broomeberg が提供する為替レートを使用し各国の貿易額で加重平均を用いて作成される。なお、指標は 1999 年 1 月 4 日を 100 として指数化されている。実質実効為替レートや名目為替レートとの違いは、日次データで公表されており、利用する為替レート

のソースや時点が異なる特徴がある。本研究では、円インデックスを名目実効為替レートの推定に用いる。

2.1.6 長期プライムレート

長期プライムレートは、銀行が1年を超える期間で大企業向けに融資する際の指標となる金利 [11]。日本銀行が提供する長期プライムレートは、みずほ銀行が決定・公表した値を使用している。本研究では、日本銀行公表値を利用した。

2.1.7 国債利回り

流通市場における固定利付国債の実勢価格に基づいて算出した主要年限毎の半年複利金利 [7]。本研究では、10年物国債を貸出約定平均金利の推定に用いる。

2.2 推定方法

2.1節で述べたデータを用いて倒産件数を予測するための重回帰モデルを構築する。倒産件数予測モデルは、1四半期先倒産件数予測モデルと2四半期先倒産件数予測モデルの2種類を作成した。

2.2.1 倒産件数予測モデル構築の問題点

倒産件数予測モデル構築に当たり、2点の問題を挙げる

問題点1：説明変数の時点制約

倒産件数を被説明変数として2四半期先までの予測用重回帰モデルを作成する際、説明変数は予測対象期間よりも以前の時点のみを使用しなければならない。

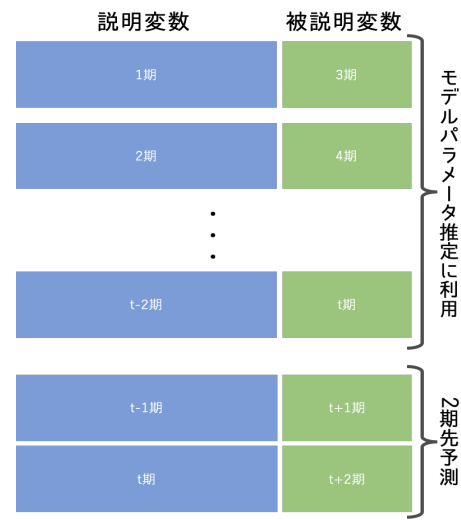
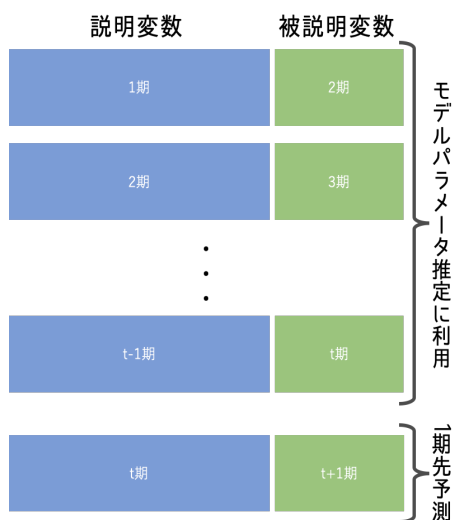


図 2.4: 1 四半期先モデルのデータ時点制約概要 図 2.5: 2 四半期先モデルのデータ時点制約概要

図 2.4, 図 2.5 は, t 期における 2 期間先予測までに使用する説明変数と被説明変数の対応する時点を表している. 例えば, 1 期先モデルで $t+1$ 期の倒産件数を予測する場合, 重回帰モデルのパラメータ推定には 1 期から $t-1$ 期までの説明変数データを利用する. そして, t 期の説明変数と推定したパラメータにより $t+1$ 期の倒産件数を予測する. 2 期先も同様である.

但し, 日本銀行短観データの業況 DI, 雇用人員 DI, 生産設備 DI, 販売価格 DI, 仕入価格 DI においては 1 期先の予測値が日本銀行から公表されているため推定, 予測ともに被説明変数と同じ時点のデータを利用できる. すなわち, $t+1$ 期の予測を行う際に上述の短観データは $t+1$ 期の説明変数を利用する.

問題点 2: 公表の速報性

本予測モデルは, 四半期ごとに予測値を算出する構造である. しかし, 説明変数として利用するデータには四半期間隔で発表されず約 1 ヶ月遅れて公表されるものもある. 例えば, 実質実効為替レートは翌月の 20 日ごろ発表されている. これでは, 1 月に 2 四半期先の予測値を公表するために 1 月 20 日まで待つ必要があり, 景況感が 1 ヶ月遅れた値を算出することになる. そこで本研究では, 公表が遅れる説明変数に対しては 1 ヶ月先予測モデル (以下, サブモデル) を構築し, 予測値公表の速報性を高めることとした. 2.3 節では, サブモデルを必要とするデータとそのモデリング方法の概要を述べる. なお, 本稿では倒産件数予測を 12 月中旬時点 (2020 年第 4 四半期) で 2 期先予測を行なった場合を例に挙げて説明と結果を述べる.

ここで倒産件数予測の全体像を俯瞰するために, モデル構築の概要を図示する.

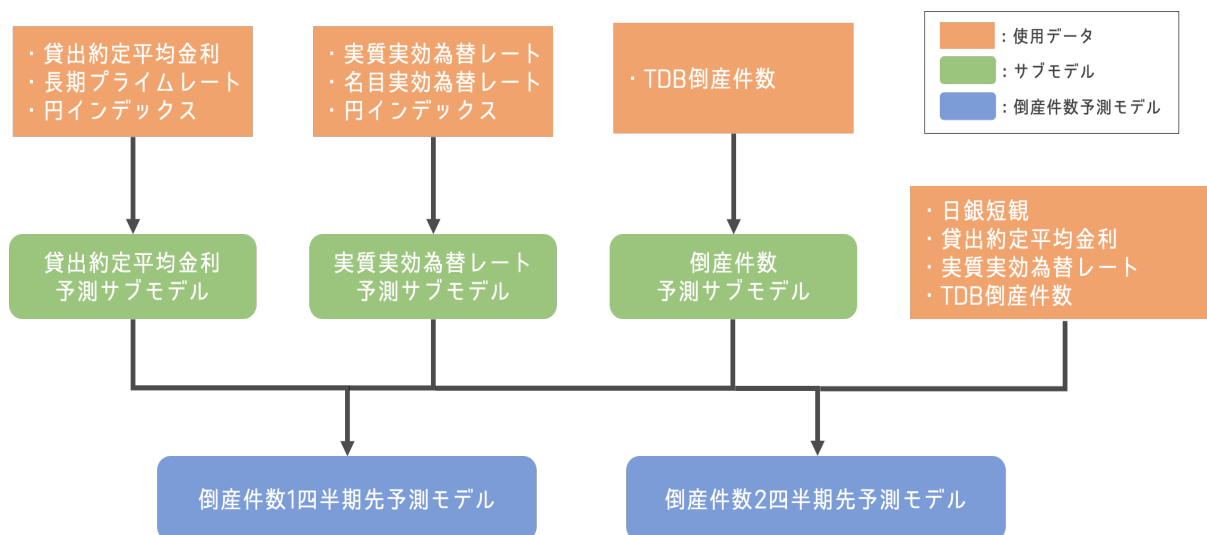


図 2.6: 倒産件数予測モデル構築概要

2.3 サブモデル

本節では, 倒産件数予測公表の速報性を確保するために作成した貸出約定平均金利, 実質実効為替レート, 倒産件数のサブモデルを説明する.

2.3.1 貸出約定平均金利サブモデル

貸出約定平均金利は、翌月下旬または翌々月の上旬に公表されるため4半期ごとの倒産件数予測を行うには2ヶ月先の貸出約定平均金利を予測する必要がある。被説明変数に貸出約定平均金利、説明変数に国債10年利回り、長期プライムレートを使用した重回帰モデルを構築する。なお、データは月次ごとに集計値を使用し、貸出約定平均金利はX12-ARIMAにより季節調整を行う。データの推定期間対象は、2007年1月2020年10月とした。季節調整済み貸出約定平均金利の月次時系列推移を図2.7に示す。

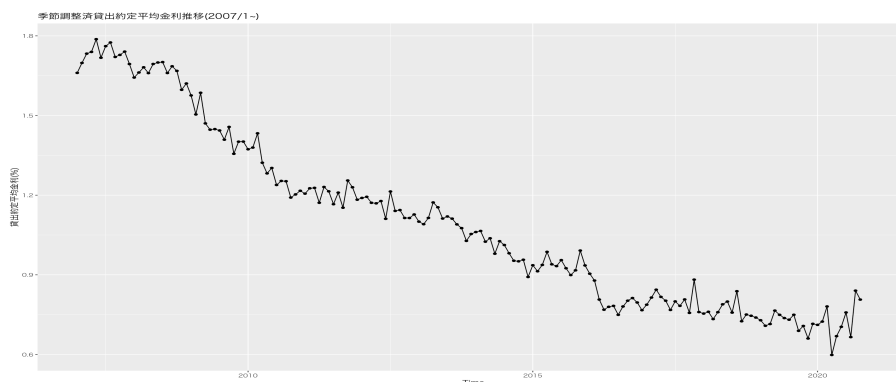


図 2.7: 季節調整済み貸出約定平均金利の月次推移

図 2.7 より、貸出約定平均金利は右肩下がりの傾向が見られる。この貸出約定平均金利を被説明変数とし、最小二乗法 (OLS) による重回帰分析を行なった結果を表 2.4 に示す。

表 2.4: 貸出約定平均金利サブモデル OLS 推定結果

説明変数	時点	符号条件	被説明変数:	
			貸出約定平均金利	VIF
国債 10 年利回り	当期	正	0.295*** (0.019)	6.151026
長期プライムレート	当期	正	0.315*** (0.024)	6.151026
定数項	無	無	0.462*** (0.023)	
観測数	166	DW 検定	0.77913***	
決定係数	0.967		(3.562e-16)	
自由度修正済決定係数	0.967	BP 検定	4.3481	
Residual Std. Error	0.060		(0.1137)	
F Statistic	2,405.270***			

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

OLS 結果より，DW 検定の 1 階の系列相関無しという帰無仮説が 1% 有意水準で棄却されていることがわかる．従って誤差項が AR(1) に従う疑いがあるため GLS(Prais-Winsten 変換) による GLSE を求める．GLS による推定結果を表 2.5 に示す．

表 2.5: 貸出約定平均金利サブモデル GLS 推定結果

説明変数	時点	符号条件	被説明変数:	
			貸出約定平均金利	VIF
国債 10 年利回り	当期	正	0.28334*** (2e-16)	3.497153
長期プライムレート	当期	正	0.30807*** (1.58e-15)	3.497153
定数項	無	無	0.47931*** (2e-16)	
観測数	166	DW 検定	2.1925	
決定係数	0.9072423		(0.8768)	
自由度修正済決定係数	0.9061042	BP 検定	0.68177	
Residual Std. Error	0.048		(0.7111)	
F Statistic	4,572.483***			

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

表 2.5 の GLS 推定結果より，多重共線性，系列相関，不均一分散の問題に対処した推定値が得られた．従って，本結果を貸出約定平均金利予測サブモデルとして利用する．貸出約定平均金利予測サブモデルによる，2020 年 11 月と 12 月の予測結果は以下となる．

表 2.6: 貸出約定平均金利予測結果

年月	2020/08 実績値	2020/09 実績値	2020/10 実績値	2020/11 予測値	2020/12 予測値
貸出約定平均金利 (季節調整済)	0.6654978	0.8404200	0.8067659	0.7955891	0.7938081

倒産件数予測モデル推定のデータとして，表 2.6 の貸出約定平均金利 2020 年 11 月，12 月予測値を利用する．

2.3.2 実質実効為替レートサブモデル

実質実効為替レートは，翌月 20 日頃に公表される．12 月中旬時点で倒産件数の 2 期先予測を行うには翌年 1 月 20 日頃に公表される 12 月分実質実効為替レートを予測する必要がある．そこで，実質実効為替レート予測サブモデルを作成した．予測方法を以下にまとめる．

1. 名目実効為替レートから実質実効為替レートを引いた値を国内外のインフレ率差と定義し、この値の予測を ARIMA(p,d,q) モデルを推定し行う。
2. 名目実効為替レートを、説明変数に円インデックスを用いた単回帰モデルで予測する。
3. 2. で予測した名目実効為替レートから 1. で予測した内外インフレ率差を差し引いて実質実効為替レートの予測値を算出する。

内外インフレ率差、名目実効為替レートそれぞれの予測モデルの詳細と推定結果を以下で述べる。なお、名目実効為替レート、実質実効為替レートは X12-ARIMA により季節調整を行なったものを使用する。

内外インフレ率差サブモデル

内外インフレ率差を、ARIMA(p,d,q) 過程によりモデリングし予測する。推定期間は、倒産件数予測モデルの推定対象期間に合わせた 2007 年 12 月から 2020 年 11 月とする。内外インフレ率差の月次推移を図 2.8 に示す。

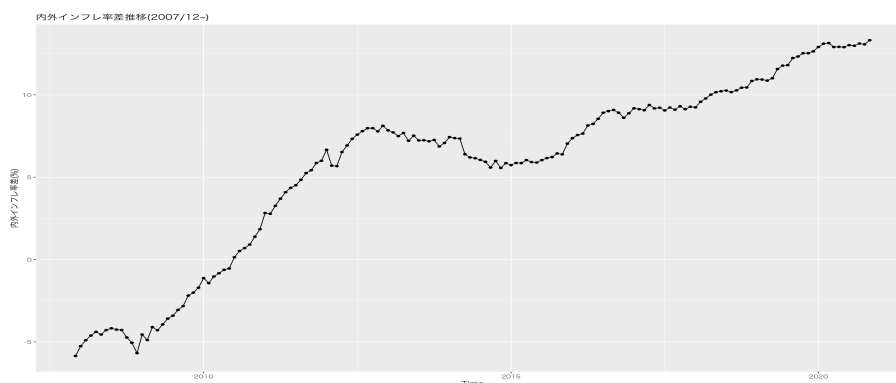


図 2.8: 内外インフレ率差の月次推移

内外インフレ率差の定常性を満たす ARIMA(p,d,q) モデルを、修正済み赤池情報量基準 (以下、AICC) 最小となるように探索した。探索結果が表 2.7 である。なお、最大探索範囲は p,d,q 全て 12 までとした。

表 2.7: ARIMA(p,d,q) 探索結果

p	d	q	AICC
0	0	0	1086.662
0	0	1	889.0544
0	0	2	722.3679
4	0	1	76.40121

表 2.7 より、AICC 基準で最小となるモデル ARMA(4,1) の推定結果を以下にまとめる。

表 2.8: 内外インフレ率差サブモデル ARMA(4,1) 推定結果

被説明変数:	
説明変数	内外インフレ率差
AR1	1.8003*** (0.0000)
AR2	-0.4955*** (0.0079)
AR3	-0.4428*** (0.0097)
AR4	0.1377 (0.1708)
MA1	-0.8475*** (0.0000)
観測数	156
$Q^*(10)$	6.019444 (0.3043)
転換点に関する検定	-1.2733 (0.203)

() 内は p 値

Note: *p<0.1; **p<0.05; ***p<0.01

表 2.8 の Ljung and Box のカバン検定 $Q^*(10)$ と転換点に関する検定結果より、どちらも帰無仮説が有意水準 5% で棄却されず誤差項が自己相関をラグ 10 期まで持たず i.i.d. に従うといえる。さらに AR 構造の定常性と MA 構造の反転可能性の診断を特性方程式の同伴行列の固有値をもとに行なった。以下に、特性方程式の同伴行列の固有値を複素平面上にプロットして示す。

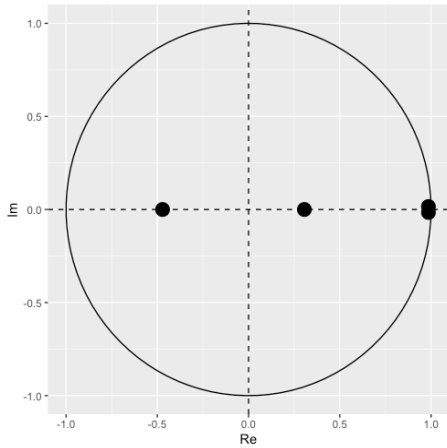


図 2.9: AR 特性方程式の固有値プロット

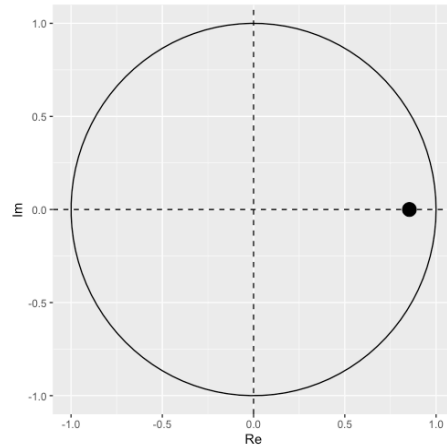


図 2.10: MA 特性方程式の固有値プロット

図 2.9, 図 2.10 より, 特性方程式の同伴行列の固有値が単位円内 (固有値が 1 以下) にあるため, 推定モデル ARMA(4,1) は AR 構造の定常性および MA 構造の反転可能性が満たされている。

上記診断結果より, 内外インフレ率差の予測サブモデルとして ARMA(4,1) を使用する。

名目実効為替レートサブモデル

実質実効為替レートに対応する名目実効為替レートは月次で公表されているが, 円インデックスは日次で公表される。そこで, 円インデックスを説明変数として名目実効為替レート予測モデルを単回帰モデルにて作成した。推定期間は, 2007年12月から2020年11月とする。円インデックスは月ごとの算術平均値を使用する。なお, 名目実効為替レート, 円インデックスともに X12-ARIMA による季節調整を行う。名目実効為替レートの月次推移は以下となる。

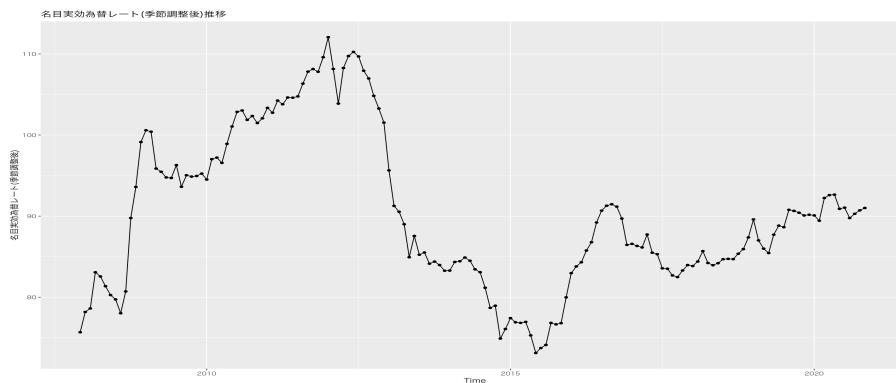


図 2.11: 名目実効為替レートの月次推移

名目実効為替レートサブモデルの OLS 推定結果を表 2.9 に示す。

表 2.9: 名目実効為替レートサブモデル OLS 推定結果

説明変数	時点	符号条件	被説明変数:
			名目実効為替レート
円インデックス	当期	正	0.855601*** (2e-16)
定数項	無	無	-0.038817 (0.916)
観測数	156	DW 検定	1.0345***
決定係数	0.998		(3.901e-10)
自由度修正済決定係数	0.997	BP 検定	0.037237
Residual Std. Error	0.478		(0.847)
F Statistic	61,573.930***		

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

OLS 結果より, DW 検定の 1 階の系列相関無しという帰無仮説が 1% 有意水準で棄却されていることがわかる. 従って誤差項が AR(1) に従う疑いがあるため GLS(Prais-Winsten 変換) による GLSE を求める. GLS による推定結果を表 2.10 に示す

表 2.10: 名目実効為替レートサブモデル GLS 推定結果

説明変数	時点	符号条件	被説明変数:
			名目実効為替レート
円インデックス	当期	正	0.854077*** (2e-16)
定数項	無	無	0.129538 (0.823)
観測数	156	DW 検定	1.8801
決定係数	0.9939556		(0.2033)
自由度修正済決定係数	0.9939163	BP 検定	0.0060988
Residual Std. Error	0.4214		(0.9378)
F Statistic	1.051e+06***		

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

表 2.10 の GLS 推定結果より, 系列相関, 不均一分散の問題に対処した推定値が得られた.

従って、本結果を名目実効為替レート予測サブモデルとして利用する。

実質実効為替レート予測結果

内外インフレ率差サブモデルおよび名目実効為替レートサブモデルの予測結果から、実質実効為替レート予測値を算出する。算出方法としては、名目実効為替レートから内外インフレ率差を差し引いて求める。内外インフレ率差、名目実効為替レート、実質実効為替レートの2020年12月予測結果は以下となる。

表 2.11: 実質実効為替レート予測結果

年月	2020/08 実績値	2020/09 実績値	2020/10 実績値	2020/11 実績値	2020/12 予測値
内外インフレ率差 (季節調整済)	12.982902	13.115675	13.068123	13.319373	13.30610
名目実効為替レート (季節調整済)	89.76295	90.28859	90.71052	91.00650	90.67101
実質実効為替レート (季節調整済)	76.78005	77.17291	77.64240	77.68713	77.36491

倒産件数予測モデル推定のデータとして、表 2.11 の実質実効為替レート 2020 年 12 月予測値を利用する。

2.3.3 倒産件数サブモデル

TDB 倒産集計は、翌月の月上旬に公表される。12 月中旬時点で倒産件数の 2 期先予測を行うには翌年 1 月上旬に公表される 12 月分 TDB 倒産件数を予測する必要がある。そこで、倒産件数予測サブモデルを作成した。推定期間は、倒産件数予測モデルの推定対象期間に合わせた 2007 年 12 月から 2020 年 11 月とする。なお、倒産件数は X12-ARIMA により季節調整を行なった値を使用する。

定常性を満たす ARIMA(p,d,q) モデルを、AICC 最小となるように探索した。探索結果が表 2.12 である。なお、最大探索範囲は p,d,q 全て 12 までとした。

表 2.12: ARIMA(p,d,q) 探索結果

p	d	q	AICC
0	0	0	2034.141
0	0	1	1923.867
0	0	2	1890.082
0	0	3	1831.815
0	0	4	1819.02
0	0	5	1799.292
1	0	0	1778.716
1	0	2	1713.875
1	0	3	1713.887
1	0	4	1715.185
2	0	0	1736.987
2	0	2	1711.179
2	0	3	1713.921
3	0	0	1709.074
3	0	1	1711.061
3	0	2	1712.355
4	0	0	1711.067
5	0	0	1711.891

表 2.12 より，AICC 基準で最小となるモデルは AR(3) となる．倒産件数推移を AR 過程で表現できる可能性が示唆されたことより VAR(p) によるモデル推定を行う．ここでは，倒産件数，実質実効為替レート，貸出約定平均金利（新規・長期・国内銀行）の 3 変数 VAR モデルから AIC 基準最小となるモデルを選択する．VAR(p) モデルの探索結果は表 2.13 である．

表 2.13: 3 変数 VAR(p) 探索結果

倒産件数	実質実効為替レート	貸出約定平均金利	選択次数	AIC
○			3	8.088306
○	○		3	9.278773
○		○	3	1.624569
○	○	○	3	2.814107

表 2.13 より，倒産件数と貸出約定平均金利の 2 変数 VAR(3) が AIC 基準で最小となる．従って本モデルを倒産件数予測サブモデルとする．2 変数 VAR(3) の推定結果を以下にまとめる．

表 2.14: 倒産件数サブモデル 2 変数 VAR(3) 推定結果

説明変数	被説明変数:	
	倒産件数	貸出約定平均金利
倒産件数_lag1	0.09716 (0.20864)	-9.708e-05 (0.1064)
貸出約定平均金利_lag1	110.69839 (0.29300)	4.242e-01*** (6.41e-07)
倒産件数_lag2	0.19681** (0.00870)	1.544e-05 (0.7886)
貸出約定平均金利_lag2	-65.06533 (0.54728)	3.678e-01*** (2.16e-05)
倒産件数_lag3	0.33559*** (8.45e-06)	3.726e-05 (0.5103)
貸出約定平均金利_lag3	156.87836 (0.14228)	2.039e-01* (0.0147)
定数項	89.34814** (0.00236)	3.097e-02 (0.1694)
観測数	156	
誤差項の独立性:F 値 (lag=12)	1.1243 (0.3239)	
誤差項の不均一分散:ARCH-LM(lag=12)	106.9 (0.5119)	

() 内は p 値 Note:*p<0.1; **p<0.05; ***p<0.01

表 2.14 の 2 変数 VAR(3) 推定結果より, 系列相関, 不均一分散の問題に対処した推定値が得られた. さらに VAR 構造の定常性の診断を特性方程式の同伴行列の固有値をもとに行なった. 以下に, 特性方程式の同伴行列の固有値を複素平面上にプロットして示す.

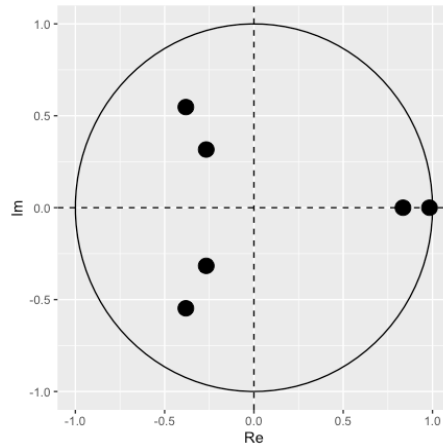


図 2.12: VAR 特性方程式の固有値プロット

図 2.12 より，特性方程式の同伴行列の固有値が単位円内 (固有値が 1 以下) にあるため，推定モデル 2 変量 VAR(3) は AR 構造の定常性が満たされている。

上記診断結果より，倒産件数予測サブモデルとして倒産件数と貸出約定平均金利の 2 変量 VAR(3) を使用する。倒産件数予測サブモデルによる，2020 年 12 月の予測結果は以下となる。なお予測値はサブモデルの予測結果から原系列への変換後のものである。

表 2.15: 倒産件数予測結果

年月	2020/08 実績値	2020/09 実績値	2020/10 実績値	2020/11 実績値	2020/12 予測値
倒産件数 (原系列)	655	602	647	563	640.32

倒産件数予測モデル推定のデータとして，表 2.15 の倒産件数 2020 年 12 月予測値を利用する。

2.4 推定結果

2.4.1 説明変数の選択結果

倒産件数 1 期・2 期先予測モデルそれぞれの説明変数候補の中から符号条件が合致するモデルの中から AIC 最小基準でモデル選択を行なった。説明変数の候補として使用した変数およびその符号条件，各モデルが最終的に使用する説明変数を以下に表形式でまとめる。

表 2.16: 説明変数選択結果

対象指標	説明変数候補	時点	符号条件	1期先	2期先
業況	業況 DI(中小企業・製造業)	当期	負	○	
		前期	負		○
採算性	販売価格 DI-仕入価格 DI(中小企業・全産業)	当期	負	○	
		前期	負		○
固定費負担	雇用人員 DI(中小企業・全産業)	当期	正		
		前期	正		
	生産設備 DI(中小企業・全産業)	当期	正		
		前期	正		
借入環境	資金繰り DI(中小企業・全産業)	前期	負		
		前々期	負		
	貸出態度 DI(中小企業・全産業)	前期	負	○	
		前々期	負		
金利面	貸出約定平均金利(新規・総合・国内銀行)	前期	正		
		前々期	正		
	貸出約定平均金利(新規・長期・国内銀行)	前期	正	○	
		前々期	正		○
対外競争力	実質実効為替レート	前期	正	○	
		前々期	正		○

表 2.16 より、1 期先倒産件数予測モデルは、業況 DI・販売価格 DI-仕入価格 DI・貸出態度 DI・貸出約定平均金利・実質実効為替レートを含む 5 変数の重回帰モデル、2 期先倒産件数予測モデルは、業況 DI・販売価格 DI-仕入価格 DI・貸出約定平均金利・実質実効為替レートを含む 4 変数の重回帰モデルを予測モデルとして採用する。但し、日銀短観の業況・販売価格・仕入価格・生産設備 DI は、前期調査時の 1 四半期先予測値が日本銀行から公表されている。そこで、1 四半期先モデルの予測の際は「当期」をこの 1 四半期先予測値で代用する。

2.4.2 倒産件数 1 四半期先予測モデル

2.4.1 で選択された 1 期先倒産件数予測用 5 変数重回帰モデルのパラメータ推定結果を示す。ここで、通常の OLS を行ない DW 検定の結果、帰無仮説が有意水準 5% で棄却され誤差項が 1 階の自己相関構造を持つ可能性が示唆された。従って、PW 変換による GLS 推定を行いその L-R プロットを以下に示す。赤点は外れ値 (Y 方向の影響点)、紫点は高い作用点 (X 方向の影響点) を表す。

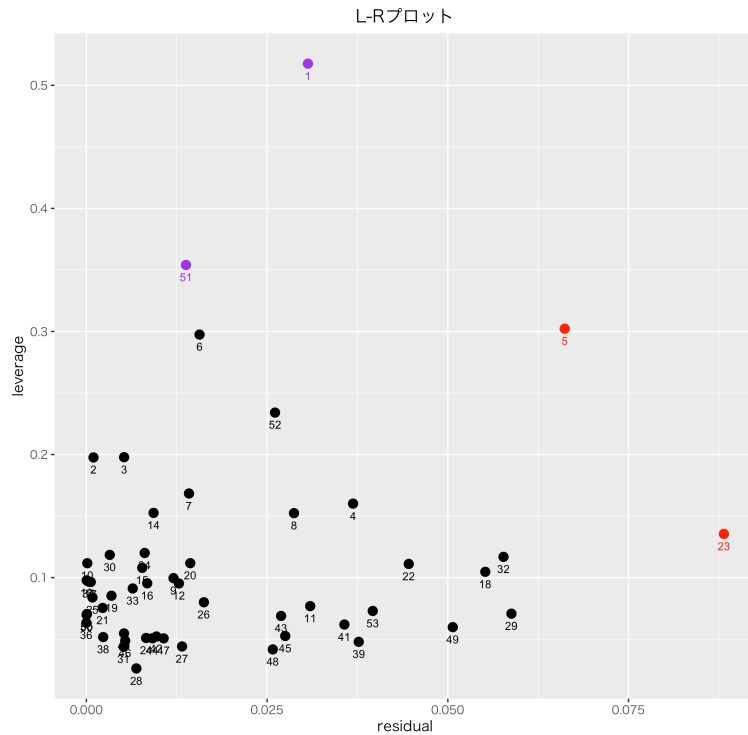


図 2.13: 倒産件数 1 期先モデル L-R プロット

図 2.13 にて、赤点は式 A.39 基準に従って検出した影響点を指す。影響点を 4 点検出したため PW 変換後に MM 推定を行い頑健なパラメータ推定を行なった。1 期先モデルの GLS・MM 推定結果を表 2.17 にまとめる。

MM 推定結果より、多重共線性、系列相関、不均一分散の問題に対処した推定値が得られたと判断。従って、表 2.17 の結果を倒産件数 1 期先予測モデルとして利用する。また、MM 推定による加重最小二乗推定の各データ点へのウェイトを棒グラフで表現し図 2.14 に示す。図 2.14 から、外れ値 (赤棒線) を中心にウェイトを低下させ影響点による推定量への影響を制御していることがわかる。

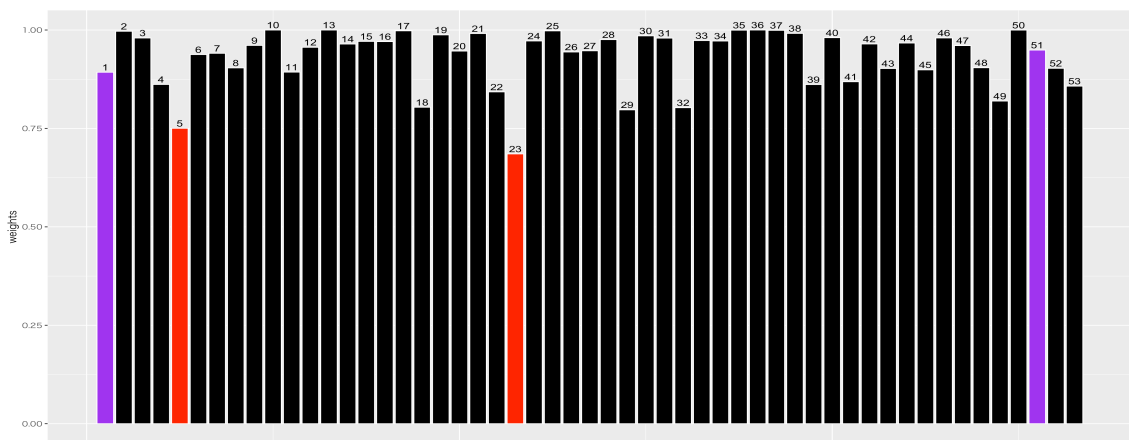


図 2.14: 倒産件数 1 期先予測モデル MM 推定の各ウェイト

表 2.17: 倒産件数 1 期先予測モデル MM 推定結果

説明変数	時点	符号条件	被説明変数:	
			倒産件数	VIF
業況 DI	当期	負	-2.1033 (0.20770)	2.0230
販売価格 DI-仕入価格 DI	当期	負	-3.0543 (0.28460)	3.4197
貸出態度 DI	前期	負	-13.147** (0.03126)	2.8516
貸出約定平均金利	前期	正	735.59*** (0.00022)	4.8955
実質実効為替レート	前期	正	6.5113* (0.08377)	2.4358
定数項	-	-	1,134.2** (0.01149)	-
観測数	53	DW 検定	1.9884	
決定係数	0.9469067		(0.2752)	
自由度修正済決定係数	0.9412585	BP 検定	3.2558	
Residual Std. Error	104.7		(0.6606)	

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

2.4.3 倒産件数 2 四半期先予測モデル

2.4.1 で選択された 2 期先倒産件数予測用 4 変数重回帰モデルのパラメータ推定結果を示す。ここで、通常の OLS を行ない DW 検定の結果、帰無仮説が有意水準 5% で棄却され誤差項が 1 階の自己相関構造を持つ可能性が示唆された。従って、PW 変換による GLS 推定を行いその L-R プロットを以下に示す。赤点は外れ値 (Y 方向の影響点)、紫点は高い作用点 (X 方向の影響点) を表す。

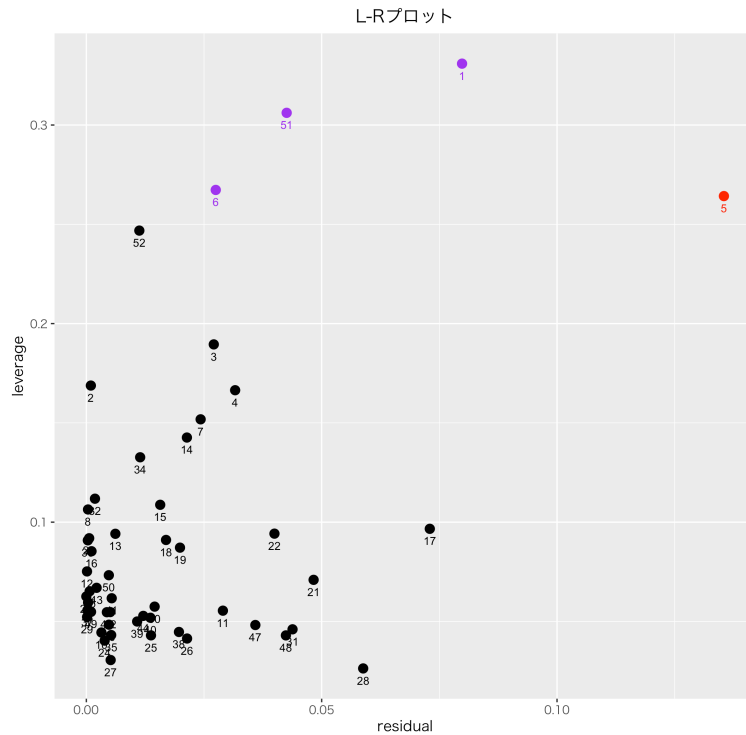


図 2.15: 倒産件数 2 期先モデル L-R プロット

図 2.15 にて、赤点は式 A.39 基準に従って検出した影響点を指す。影響点を 4 点検出したため PW 変換後に MM 推定を行い頑健なパラメータ推定を行なった。2 期先モデルの GLS・MM 推定結果を表 2.18 にまとめる。

MM 推定結果より、多重共線性、系列相関、不均一分散の問題に対処した推定値が得られたと判断。従って、表 2.18 の結果を倒産件数 2 期先予測モデルとして利用する。また、MM 推定による加重最小二乗推定の各データ点へのウェイトを棒グラフで表現し図 2.16 に示す。図 2.16 から、外れ値 (赤棒線) を中心にウェイトを低下させ影響点による推定量への影響を制御していることがわかる。

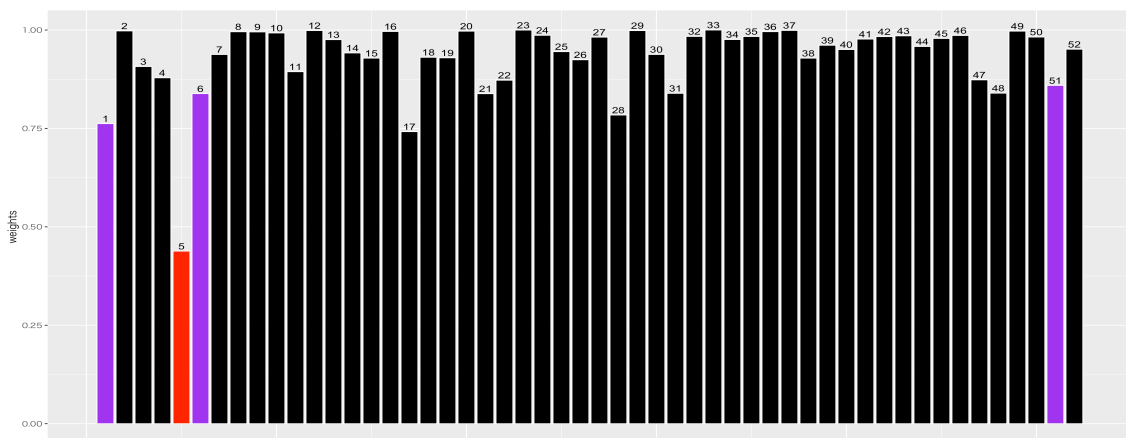


図 2.16: 倒産件数 2 期先予測モデル MM 推定の各ウェイト

表 2.18: 倒産件数 2 期先予測モデル MM 推定結果

説明変数	時点	符号条件	被説明変数:	
			倒産件数	VIF
業況 DI	前期	負	-1.4663 (0.28652)	1.7631
販売価格 DI-仕入価格 DI	前期	負	-4.9435 (0.19051)	3.6313
貸出約定平均金利	前々期	正	961.30*** (2.6e-07)	4.84340
実質実効為替レート	前々期	正	11.496*** (0.00031)	2.0602
定数項	-	-	280.39 (0.16490)	-
観測数	52	DW 検定	1.8411	
決定係数	0.8848281		(0.1444)	
自由度修正済決定係数	0.8752305	BP 検定	8.9177	
Residual Std. Error	125.6		(0.06319)	

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

2.5 予測結果

2.5.1 倒産件数 1 四半期先予測結果

2.4.2 のモデルを使用して 1 期先 (2021 年第 1 四半期) の倒産件数を予測する。予測値とその 95% 信頼区間および実績値推移のプロットを図 2.17 に示す。なお、縦点線は MM 推定量を算出する際に用いたデータ期間を表す (2007 年第 4 四半期から 2020 年第 4 四半期)。図 2.17 の推移より、実績値が推定対象全期間で 95% 信頼区間に収まっており、倒産件数推移の変化を概ね捉えている。本モデルによる 2021 年第 1 四半期予測値を 1 期先予測として使用する。

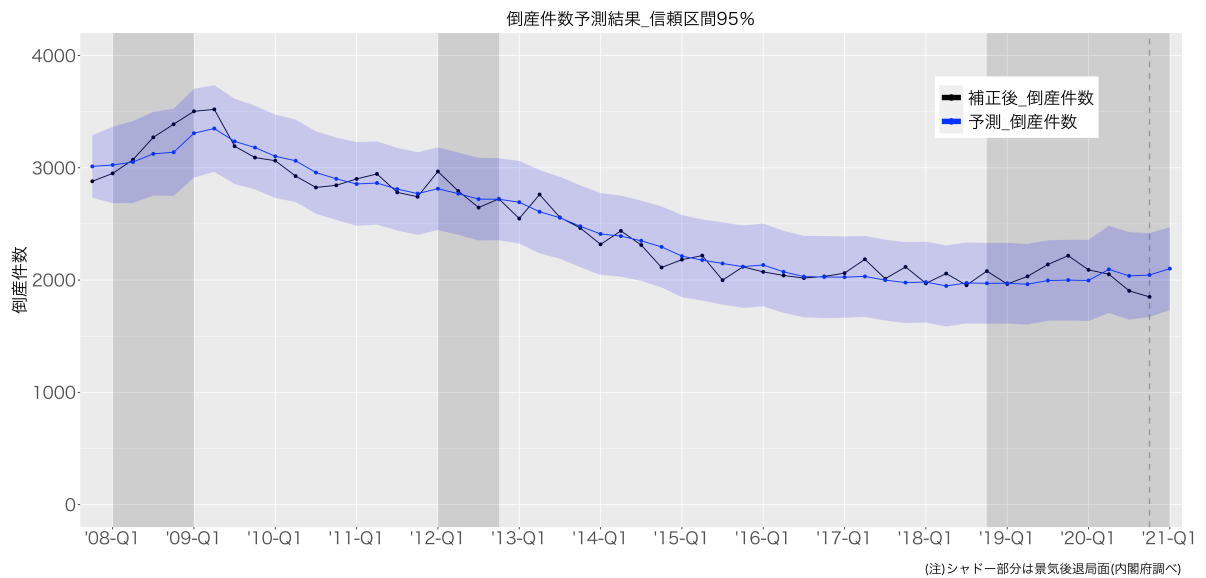


図 2.17: 倒産件数 1 期先モデルの予測結果

2.5.2 倒産件数 2 四半期先予測結果

2.4.3 のモデルを使用して 2 期先 (2021 年第 2 四半期) の倒産件数を予測する。予測値とその 95% 信頼区間および実績値推移のプロットを図 2.18 に示す。なお、縦点線は MM 推定量を算出する際に用いたデータ期間を表す (2008 年第 1 四半期から 2020 年第 4 四半期)。図 2.18 の推移より、実績値が推定対象全期間で 95% 信頼区間に収まっており、倒産件数推移の変化を概ね捉えている。本モデルによる 2021 年第 2 四半期予測値を 2 期先予測として使用する。

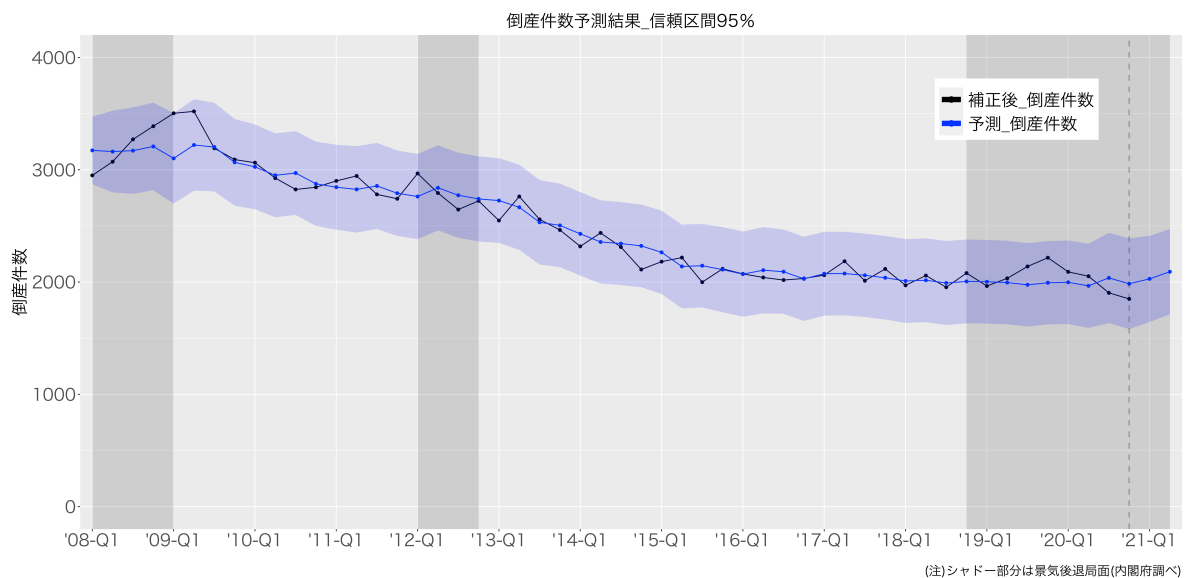


図 2.18: 倒産件数 2 期先モデルの予測結果

2.6 予測の要因分解

1 期先モデルの要因分解図を以下に示す。

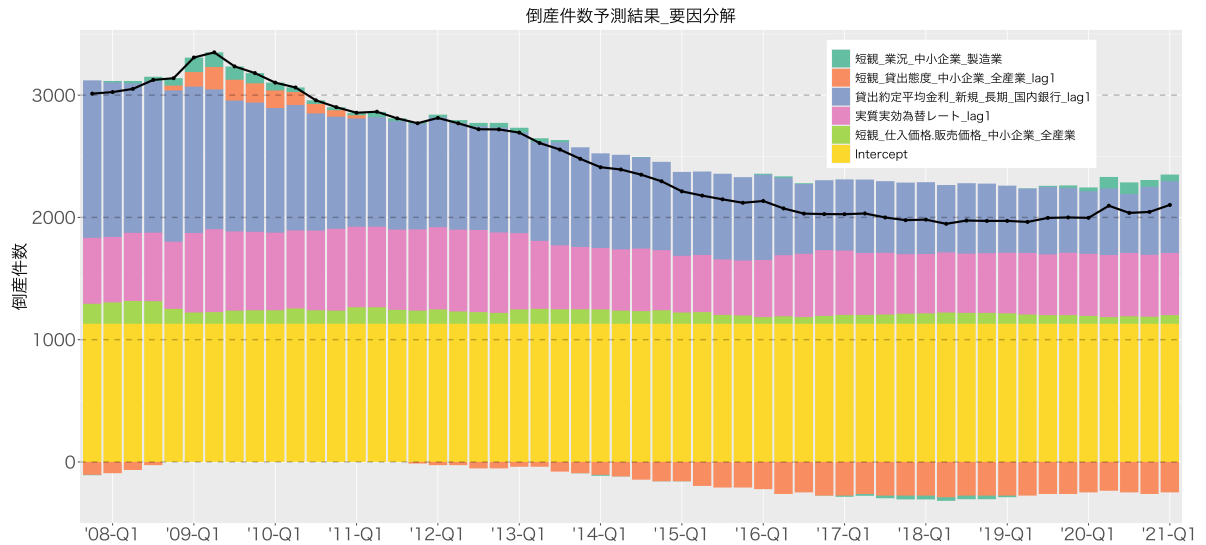


図 2.19: 倒産件数 1 期先モデルの要因分解図

2.7 結果

図 2.19 より、世界金融危機時と比較してコロナ禍の倒産件数が低水準に止まっている主要因は、貸出約定平均金利の低下、貸出態度 DI から推測される貸出態度の軟化、実質実効為替レートの低下であることが示唆された。また、2013 年第 2 四半期以降の倒産件数減少もこれら 3 要因が説明に寄与することから本予測モデルが日本銀行の金融緩和政策による倒産への影響を定量的に評価する可能性を持つことも示唆される。これについては、評価事項を明確にした上で他要因を考慮する必要がありさらなる研究が必要である。倒産件数の 2 四半期先 (2021 年第 2 四半期) までの予測結果を実績値と併せて表にまとめる。また、実績値と予測値の推移を図 2.20 に示す。

表 2.19: 倒産件数予測結果

	'20-Q1 実績値	'20-Q2 実績値	'20-Q3 実績値	'20-Q4 実績値	'21-Q1 予測値	'21-Q2 予測値
95% 信頼区間上限	-	-	-	-	2,488	2,471
倒産件数	2,091	2,052	1,904	1,850	2,102	2,092
95% 信頼区間下限	-	-	-	-	1,716	1,716

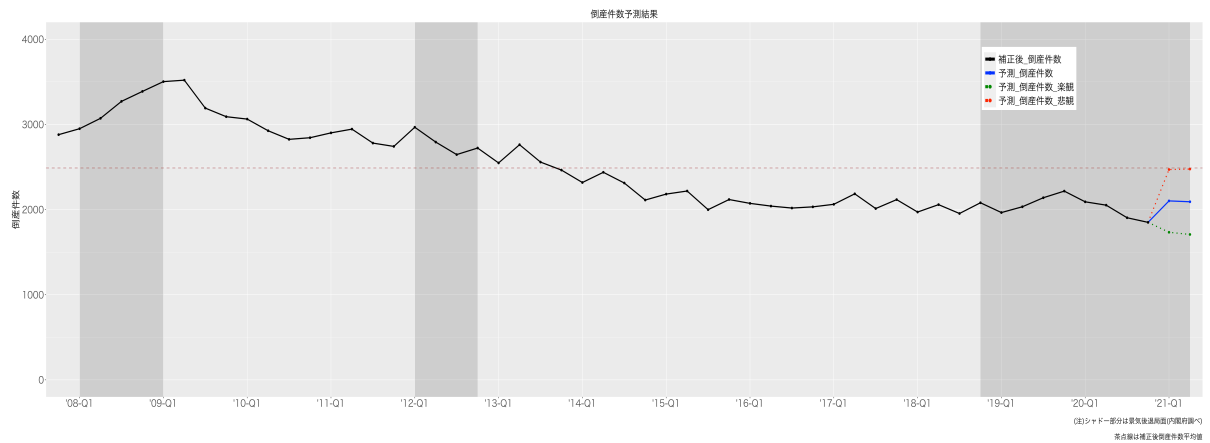


図 2.20: 倒産件数予測推移

図 2.20 より、95% 信頼区間上限の予測であっても推定対象期間平均を下回る予測となっており、本研究による予測では 2021 年第 2 四半期まで倒産件数は低水準で推移すると予測している。ただし、本モデルの予測に用いたデータは主に 11 月調査のものであるため足許の景況感を含めた予測でないことには注意が必要である。

第3章 完全失業率予測

本章では、コロナ禍の雇用問題の深刻さを捉えるため開発した完全失業率予測モデルについて述べる。はじめにモデル作成に用いたデータの詳細を説明し、推定結果および第2四半期先までの予測結果、推定量の要因分解による予測に寄与する要因の考察を行う。

3.1 使用データ

本モデルで使用する日銀短観、貸出約定平均金利、実質実効為替レートは、倒産件数予測で用いたものと同様のデータを使用する。そのため本節では、新たに使用する完全失業率データについて述べる。

3.1.1 完全失業率データ

完全失業率とは、労働力人口に占める完全失業者の割合である。本研究で使用する完全失業率データは、総務省統計局が実施している労働力調査に基づくものである [14]。なお、季節調整値が総務省統計局 HP にて公開されておりこちらを利用した。

総務省統計局 HP によると、労働力人口は15才以上人口の完全失業者と就業者を合わせたものとされている。完全失業者とは次の3つの条件を満たす者を指す。

1. 仕事がなく調査週間中に少しも仕事をしなかった
2. 仕事があればすぐ就くことができる
3. 調査週間中に、仕事を探す活動や事業を始める準備をしていた

また、就業者は調査週間中に賃金、給料、諸手当、内職収入などの収入を伴う仕事を1時間以上した従業者と休業者を合わせたものである。詳細は総務省統計局 HP 「労働力調査用語の解説」 [16] を参照されたい。

完全失業率予測モデルでは、このデータを予測対象としてモデルを構築する。以下に四半期ごと2000年第1四半期(2000-Q1)から2020年第3四半期(2020-Q3)完全失業率推移を示す。

これは総務省統計局が公表する労働力調査の月次データを四半期ごとに算術平均をとり作成した。なお、シャドー部分は内閣府が発表する景気基準日付 [6] を元にした景気後退局面である。2018年第4四半期(2018-Q4)以降の景気後退局面は暫定であることに注意されたい。

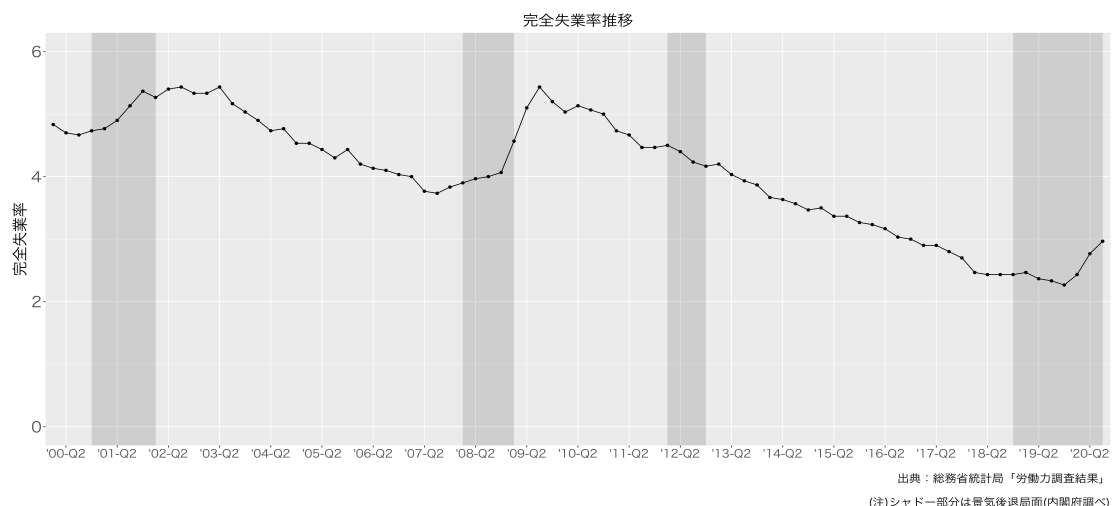


図 3.1: 完全失業率の推移 (四半期)

完全失業率推移の特徴として、2000年代前半と世界金融危機時の高水準と足許は右肩上がりの傾向が見られる。

3.2 推定方法

2.1節と3.1節で述べたデータを用いて完全失業率を予測するための重回帰モデルを構築する。完全失業率予測モデルは、1四半期先完全失業率予測モデルと2四半期先完全失業率予測モデルの2種類を作成した。

3.2.1 完全失業率予測モデル構築の問題点

完全失業率予測モデル構築に当たり、2.2.1小節と同様の問題が2点存在する。

問題点1：説明変数の時点制約

問題点2：公表の速報性

従って、完全失業率予測モデルにおいても説明変数の適切な時点データの使用とサブモデル作成による速報性の担保を行う。

ここで完全失業率予測の全体像を俯瞰するために、モデル構築の概要を図示する。

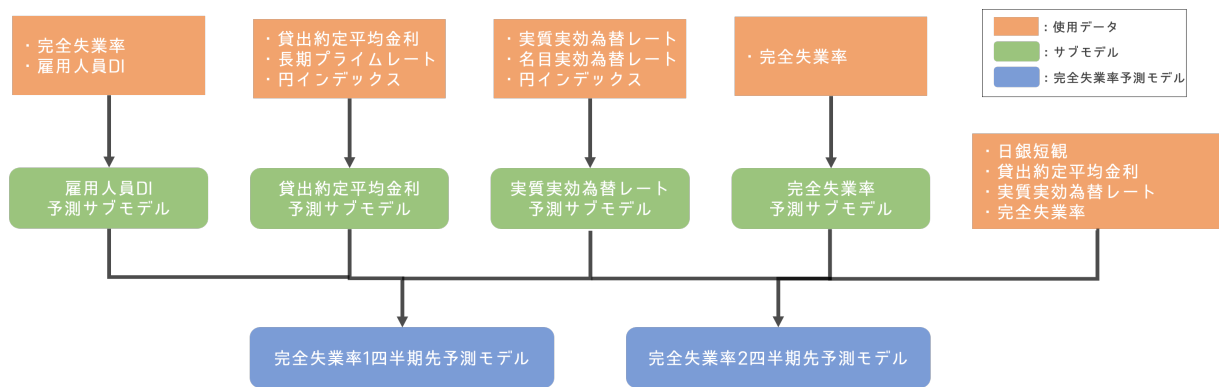


図 3.2: 完全失業率予測モデル構築概要

3.3 サブモデル

本節では、完全失業率予測公表の速報性を確保するために作成した雇用人員 DI、完全失業率のサブモデルを説明する。なお、本モデルで使用する貸出約定平均金利と実質実効為替レートのサブモデルは完全失業率予測モデル時に説明したものと同様のものを使用するため説明を省略する。

3.3.1 雇用人員 DI サブモデル

雇用人員 DI は、日本銀行短観データのため四半期ごとに公表される。完全失業率 2 期先モデルでは雇用人員 DI の当期を使用する。そのため 12 月中旬時点での 2 期先予測のためには 2021 年第 2 四半期の値が必要となる。そこで雇用人員 DI サブモデルを作成した。推定期間は、完全失業率モデルの推定対象期間に合わせた 2002 年第 2 四半期から 2020 年第 4 四半期とした。雇用人員 DI の四半期時系列推移を図 3.3 に示す。

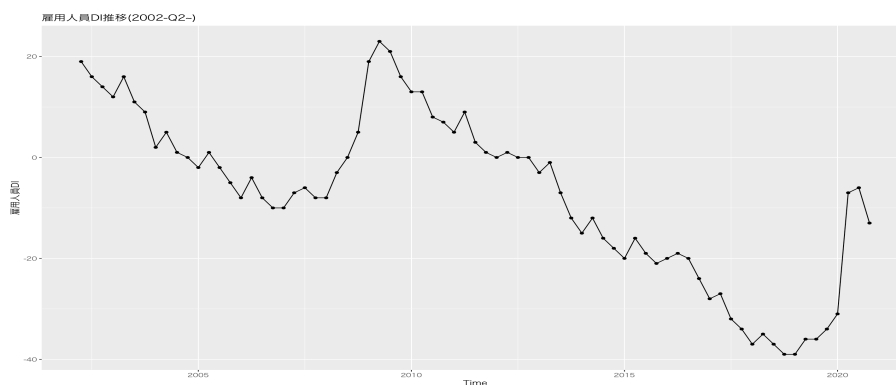


図 3.3: 雇用人員 DI の四半期推移

雇用人員 DI データに対して、定常性を満たす ARIMA(p,d,q) モデルを、AICC 最小となるように探索した。探索結果が表 3.1 である。なお、最大探索範囲は p,d,q 全て 12 までとした。

表 3.1: ARIMA(p,d,q) 探索結果

p	d	q	AICC
0	0	0	648.1895
0	0	1	574.4241
0	0	5	465.7809
1	0	0	444.1389
1	0	1	443.9637
1	0	2	445.6332
1	0	3	443.1151
1	0	4	441.6686
2	0	0	443.7724
2	0	3	443.8035
3	0	0	445.7606
4	0	0	447.6356
5	0	0	437.0472

表 3.1 より, AICC 基準で最小となるモデルは AR(5) となる. 雇用人員 DI 推移を AR 過程で表現できる可能性が示唆されたことより VAR(p) によるモデル推定を行う. ここでは, 雇用人員 DI, 完全失業率, 業況 DI の 3 変数 VAR モデルから AIC 基準最小となるモデルを選択する. VAR(p) モデルの探索結果は表 3.2 である.

表 3.2: 3 変数 VAR(p) 探索結果

雇用人員 DI	完全失業率	業況 DI	選択次数	AIC
○			5	2.956332
○	○		1	-1.5310665
○		○	5	5.366136
○	○	○	2	1.092843

表 3.2 より, 雇用人員 DI と完全失業率の 2 変数 VAR(1) が AIC 基準で最小となる. 従って本モデルを雇用人員 DI 予測サブモデルとする. 2 変数 VAR(1) の推定結果を以下にまとめる.

表 3.3: 雇用人員 DI サブモデル 2 変数 VAR(1) 推定結果

被説明変数:		
説明変数	雇用人員 DI	完全失業率
雇用人員 DI_lag1	1.0342*** (1.41e-15)	0.017376*** (2.00e-08)
完全失業率_lag1	-1.6460 (0.367)	0.681445*** ($<2e-16$)
定数項	6.3224 (0.425)	1.352122*** (1.98e-08)
観測数	75	
誤差項の独立性:Ljung and Box(lag=4)	13.594 (0.3274)	
誤差項の不均一分散:ARCH-LM(lag=4)	41.904 (0.2301)	

() 内は p 値 Note:*p<0.1; **p<0.05; ***p<0.01

表 3.3 の 2 変数 VAR(1) 推定結果より, 系列相関, 不均一分散の問題に対処した推定値が得られた. さらに VAR 構造の定常性の診断を特性方程式の同伴行列の固有値をもとに行なった. 以下に, 特性方程式の同伴行列の固有値を複素平面上にプロットして示す.

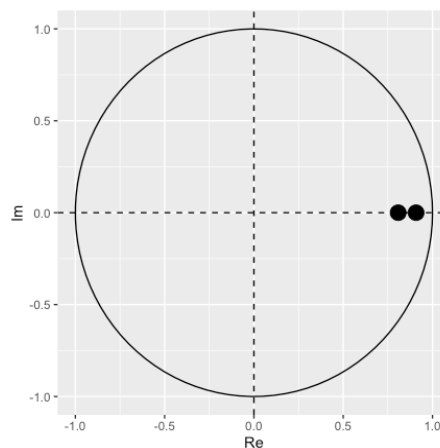


図 3.4: VAR 特性方程式の固有値プロット

図 3.4 より, 特性方程式の同伴行列の固有値が単位円内 (固有値が 1 以下) にあるため, 推定モデル 2 変数 VAR(1) は AR 構造の定常性が満たされている.

上記診断結果より, 雇用人員 DI 予測サブモデルとして雇用人員 DI と完全失業率の 2 変数 VAR(1) を使用する. 雇用人員 DI 予測サブモデルによる, 2021 年第 2 四半期の予測結果

は以下となる。なお、2021年第1四半期は日本銀行が公表している予測値を用い、2021年第2四半期予測の際に使用する。

表 3.4: 雇用人員 DI 予測結果

年月	'20-Q2 実績値	'20-Q3 実績値	'20-Q4 実績値	'21-Q1 公表値	'21-Q2 予測値
雇用人員 DI	-7	-6	-13	-16	-15.673

完全失業率予測モデル推定のデータとして、表 3.4 の雇用人員 DI2021 年第 2 四半期予測値を利用する。

3.3.2 完全失業率サブモデル

完全失業率は、総務省統計局による労働力調査で集計・公表されており翌月に公表される。12月中旬時点で完全失業率の2期先予測を行うには翌年1月に公表される12月分完全失業率を予測する必要がある。そこで、完全失業率予測サブモデルを作成した。推定期間は、2002年12月から2020年11月とする。なお、完全失業率は公表されている季節調整値を使用する。

定常性を満たすARIMA(p,d,q)モデルを、AICC最小となるように探索した。探索結果が表 3.5 である。なお、最大探索範囲は p,d,q 全て 12 までとした。

表 3.5: ARIMA(p,d,q) 探索結果

p	d	q	AICC
0	0	0	576.2055
0	0	1	315.7933
0	0	2	156.676
0	0	3	47.12129
0	0	4	-22.77069
0	0	5	-70.44861
4	0	1	-278.7144

表 3.5 より、AICC 基準で最小となるモデル ARMA(4,1) の推定結果を以下にまとめる。

表 3.6: 完全失業率サブモデル ARMA(4,1) 推定結果

説明変数	被説明変数:
	完全失業率
AR1	1.6486*** (0.0000)
AR2	-0.6125*** (5.17e-04)
AR3	0.1666 (2.25e-01)
AR4	-0.2053*** (4.34e-03)
MA1	-0.8319*** (9.46e-09)
定数項	3.9991*** (2.22e-16)
観測数	217
$Q^*(10)$	2.6669 (0.751)
転換点に関する検定	0.9162 (0.360)

() 内は p 値

Note: *p<0.1; **p<0.05; ***p<0.01

表 3.6 の Ljung and Box のカバン検定 $Q^*(10)$ と転換点に関する検定結果より, どちらも帰無仮説が有意水準 5% で棄却されず誤差項が自己相関をラグ 10 期まで持たず i.i.d. に従うといえる. さらに AR 構造の定常性と MA 構造の反転可能性の診断を特性方程式の同伴行列の固有値をもとに行なった. 以下に, 特性方程式の同伴行列の固有値を複素平面上にプロットして示す.

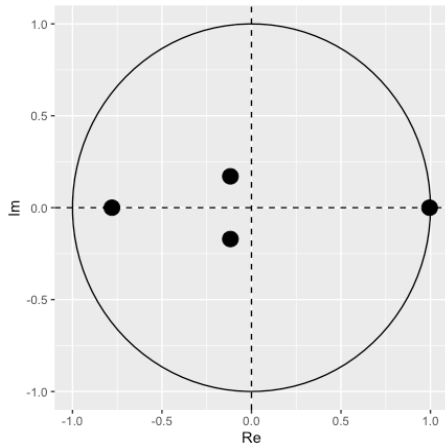


図 3.5: AR 特性方程式の固有値プロット

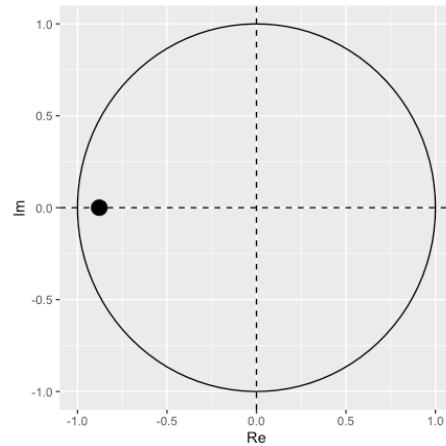


図 3.6: MA 特性方程式の固有値プロット

図 3.5, 図 3.6 より, 特性方程式の同伴行列の固有値が単位円内 (固有値が 1 以下) にあるため, 推定モデル ARMA(4,1) は AR 構造の定常性および MA 構造の反転可能性が満たされている。

上記診断結果より, 内外インフレ率差の予測サブモデルとして ARMA(4,1) を使用する。完全失業率予測サブモデルによる, 2020 年 12 月の予測結果は以下となる。

表 3.7: 完全失業率予測結果

年月	2020/08 実績値	2020/09 実績値	2020/10 実績値	2020/11 実績値	2020/12 予測値
完全失業率	3.0	3.0	3.1	2.9	2.9793

完全失業率予測モデル推定のデータとして, 表 3.7 の完全失業率 2020 年 12 月予測値を利用する。

3.4 推定結果

3.4.1 説明変数の選択結果

完全失業率 1 期・2 期先予測モデルそれぞれの説明変数候補の中から符号条件が合致するモデルの中から AIC 最小基準でモデル選択を行なった。説明変数の候補として使用した変数およびその符号条件, 各モデルが最終的に使用する説明変数を以下に表形式でまとめる。

表 3.8: 説明変数選択結果

対象指標	説明変数候補	時点	符号条件	1期先	2期先
業況	業況 DI(中小企業・全産業)	当期	負		
		前期	負		
		前々期	負		
	業況 DI(中小企業・製造業)	当期	負		
		前期	負		
		前々期	負		
人件費負担	雇用人員 DI(中小企業・全産業)	当期	正	○	○
		前期	正	○	○
		前々期	正	○	○
		3期前	正	○	○
減価償却費負担	生産設備 DI(中小企業・全産業)	当期	正		
		前期	正		
		前々期	正		
	生産設備 DI(中小企業・製造業)	当期	正		
		前期	正		
		前々期	正		
採算性	販売価格 DI-仕入価格 DI(中小企業・全産業)	当期	負		
		前期	負		
		前々期	負		
借入環境	資金繰り DI(中小企業・全産業)	前期	負		
		前々期	負		
	貸出態度 DI(中小企業・全産業)	前期	負		
		前々期	負	○	○
金利面	貸出約定平均金利 (新規・総合・国内銀行)	前期	正		
		前々期	正		
	貸出約定平均金利 (新規・長期・国内銀行)	前期	正		
		前々期	正	○	○
対外競争力	実質実効為替レート	前期	正	○	
		前々期	正		○

表 3.8 より、1期先完全失業率予測モデルは、雇用人員 DI・貸出態度 DI・貸出約定平均金利・実質実効為替レートを含む7変数の重回帰モデル、2期先完全失業率予測モデルも、雇用人員 DI・貸出態度 DI・貸出約定平均金利・実質実効為替レートを含む7変数の重回帰モデルを予測モデルとして採用する。但し、日銀短観の雇用人員 DI・貸出態度 DIは、前期調査時の1四半期先予測値が日本銀行から公表されている。そこで、1四半期先モデルの予測の際は「当期」をこの1四半期先予測値で代用する。

3.4.2 完全失業率1四半期先予測モデル

3.4.1で選択された1期先完全失業率予測用7変数重回帰モデルのパラメータ推定結果を示す。ここで、通常のOLSを行ないDW検定の結果、帰無仮説が有意水準5%で棄却され誤差項が1階の自己相関構造を持つ可能性が示唆された。従って、PW変換によるGLS推

定を行いその L-R プロットを以下に示す。赤点は外れ値 (Y 方向の影響点), 紫点は高い作用点 (X 方向の影響点) を表す。

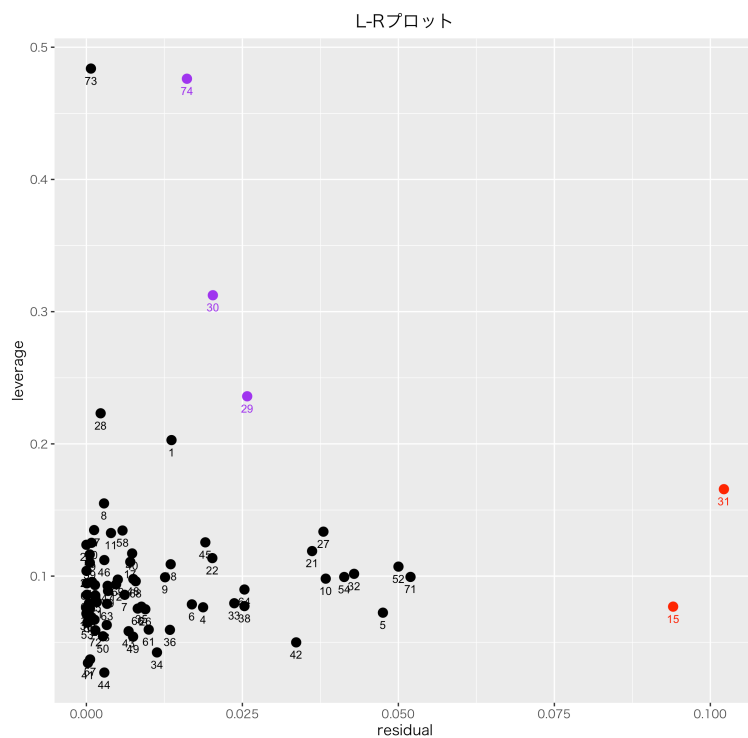


図 3.7: 完全失業率 1 期先モデル L-R プロット

図 3.7 にて、赤点・紫点は式 A.39 基準に従って検出した影響点を指す。影響点を 5 点検出したため PW 変換後に MM 推定を行い頑健なパラメータ推定を行なった。1 期先モデルの GLS・MM 推定結果を表 3.9 にまとめる。

MM 推定結果より、多重共線性、系列相関、不均一分散の問題に対処した推定値が得られたと判断。従って、表 3.9 の結果を完全失業率 1 期先予測モデルとして利用する。また、MM 推定による加重最小二乗推定の各データ点へのウェイトを棒グラフで表現し図 3.8 に示す。図 3.8 から、外れ値 (赤棒線) を中心にウェイトを低下させ影響点による推定量への影響を制御していることがわかる。

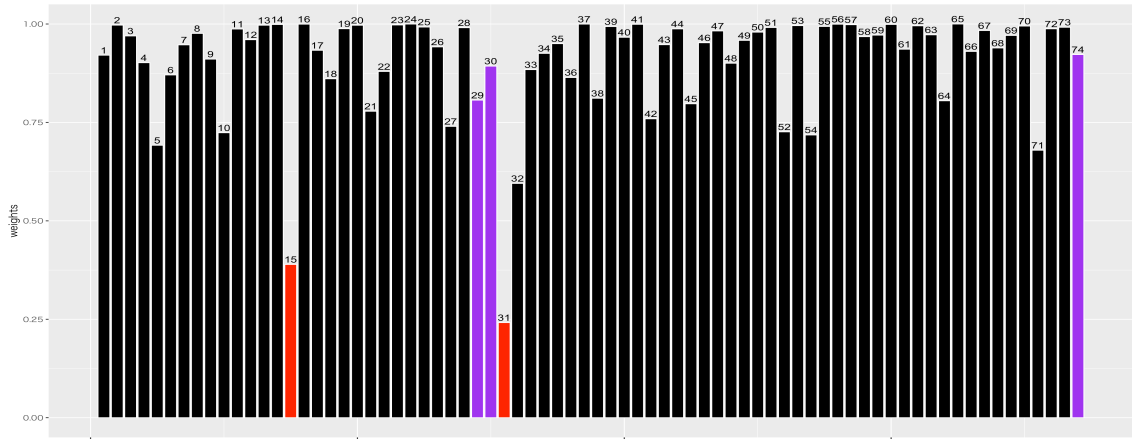


図 3.8: 完全失業率 1 期先予測モデル MM 推定の各ウェイト

表 3.9: 完全失業率 1 期先予測モデル MM 推定結果

説明変数	時点	符号条件	被説明変数:	
			完全失業率	VIF
雇用人員 DI	当期	正	0.0097*** (0.0020)	3.778
雇用人員 DI	前期	正	0.0067** (0.0350)	4.033
雇用人員 DI	前々期	正	0.0064** (0.0360)	3.830
雇用人員 DI	3 期前	正	0.0110*** (0.0021)	4.964
貸出態度 DI	前々期	負	-0.0238*** (0.0012)	7.214
貸出約定平均金利	前々期	正	0.2117* (0.0770)	2.808
実質実効為替レート	前期	正	0.0065* (0.0504)	2.781
定数項	-	-	3.4752*** (3.9e-16)	-
観測数	75	DW 検定	1.8292	
決定係数	0.9858947		(0.1354)	
自由度修正済決定係数	0.9848726	BP 検定	8.4253	
Residual Std. Error	0.08322		(0.2966)	

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

3.4.3 完全失業率 2 四半期先予測モデル

3.4.1 で選択された 2 期先倒産件数予測用 7 変数重回帰モデルのパラメータ推定結果を示す。ここで、通常の OLS を行ない DW 検定の結果、帰無仮説が有意水準 5% で棄却され誤差項が 1 階の自己相関構造を持つ可能性が示唆された。従って、PW 変換による GLS 推定を行いその L-R プロットを以下に示す。赤点は外れ値 (Y 方向の影響点)、紫点は高い作用点 (X 方向の影響点) を表す。

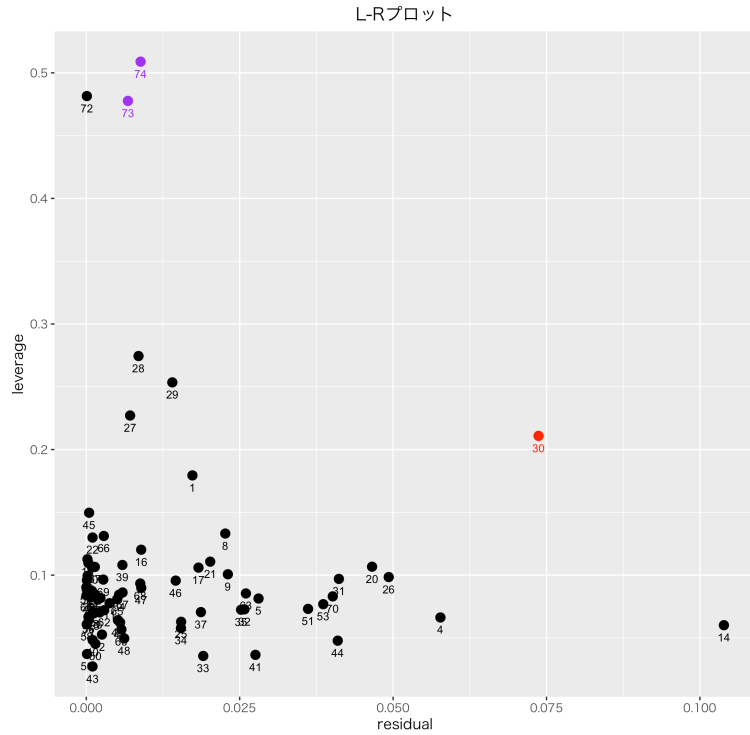


図 3.9: 完全失業率 2 期先モデル L-R プロット

図 3.9 にて、赤点・紫点は式 A.39 基準に従って検出した影響点を指す。影響点を 3 点検出したため PW 変換後に MM 推定を行い頑健なパラメータ推定を行なった。2 期先モデルの GLS・MM 推定結果を表 3.10 にまとめる。

MM 推定結果より、多重共線性、系列相関、不均一分散の問題に対処した推定値が得られたと判断。従って、表 3.10 の結果を完全失業率 2 期先予測モデルとして利用する。また、MM 推定による加重最小二乗推定の各データ点へのウェイトを棒グラフで表現し図 3.10 に示す。図 3.10 から、外れ値 (赤棒線) を中心にウェイトを低下させ影響点による推定量への影響を制御していることがわかる。

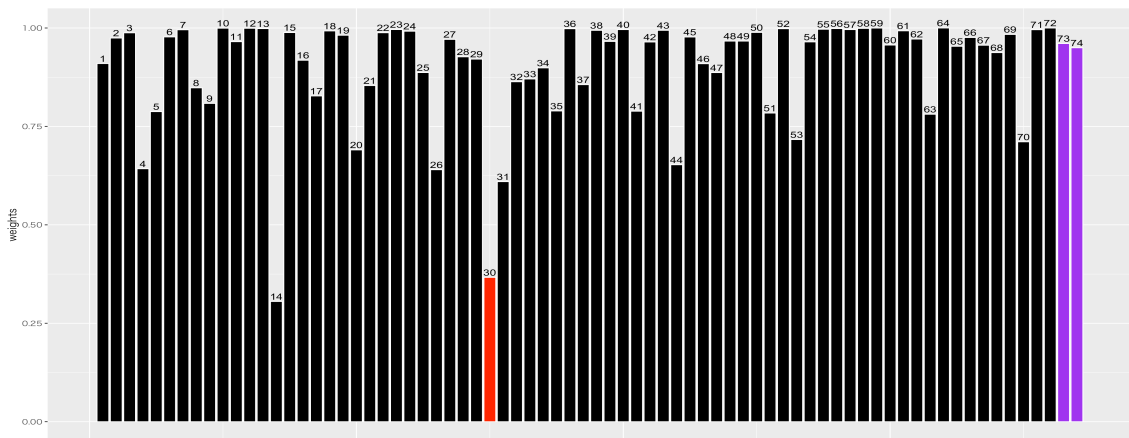


図 3.10: 完全失業率 2 期先予測モデル MM 推定の各ウェイト

表 3.10: 完全失業率 2 期先予測モデル MM 推定結果

説明変数	時点	符号条件	被説明変数:	
			完全失業率	VIF
雇用人員 DI	当期	正	0.0107*** (0.0007)	3.641
雇用人員 DI	前期	正	0.0062* (0.0523)	4.096
雇用人員 DI	前々期	正	0.0051* (0.0907)	3.995
雇用人員 DI	3 期前	正	0.0109*** (0.0026)	4.938
貸出態度 DI	前々期	負	-0.0248*** (0.0012)	7.707
貸出約定平均金利	前々期	正	0.2358** (0.0455)	2.640
実質実効為替レート	前期	正	0.0063* (0.0510)	2.542
定数項	-	-	3.4752*** (3.9e-16)	-
観測数	74	DW 検定	1.8892	
決定係数	0.9842326		(0.1986)	
自由度修正済決定係数	0.98309	BP 検定	7.2874	
Residual Std. Error	0.07912		(0.3996)	

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

3.5 予測結果

3.5.1 完全失業率 1 四半期先予測結果

3.4.2 のモデルを使用して 1 期先 (2021 年第 1 四半期) の完全失業率を予測する。予測値とその 95% 信頼区間および実績値推移のプロットを図 3.11 に示す。なお、縦点線は MM 推定量を算出する際に用いたデータ期間を表す (2002 年第 2 四半期から 2020 年第 4 四半期)。図

3.11 の推移より、実績値が推定対象全期間で 95% 信頼区間に収まっており、完全失業率推移の変化を概ね捉えている。本モデルによる 2021 年第 1 四半期予測値を 1 期先予測として使用する。

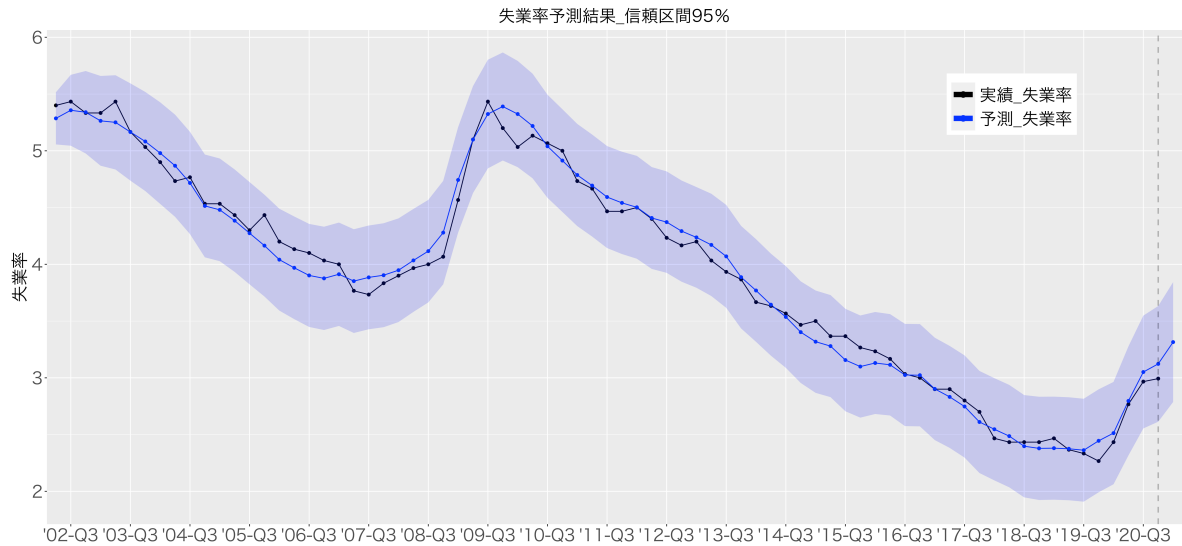


図 3.11: 完全失業率 1 期先モデルの予測結果

3.5.2 完全失業率 2 四半期先予測結果

3.4.3 のモデルを使用して 2 期先 (2021 年第 2 四半期) の完全失業率を予測する。予測値とその 95% 信頼区間および実績値推移のプロットを図 3.12 に示す。なお、縦点線は MM 推定量を算出する際に用いたデータ期間を表す (2002 年第 3 四半期から 2020 年第 4 四半期)。図 3.12 の推移より、実績値が推定対象全期間で 95% 信頼区間に収まっており、完全失業率推移の変化を概ね捉えている。本モデルによる 2021 年第 2 四半期予測値を 2 期先予測として使用する。

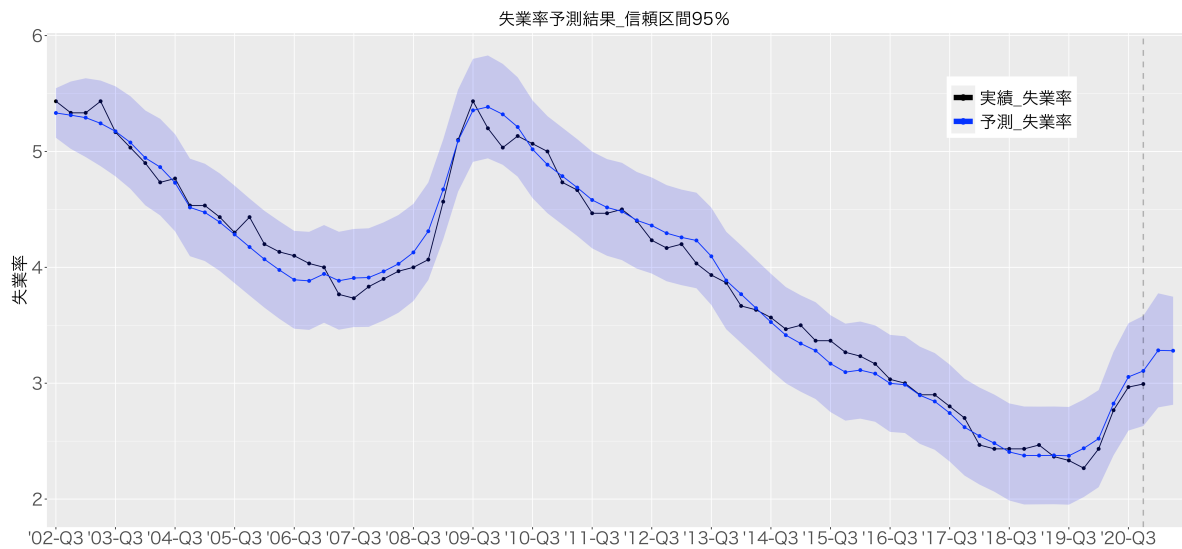


図 3.12: 完全失業率 2 期先モデルの予測結果

3.6 予測の要因分解

1 期先モデルの要因分解図を以下に示す。

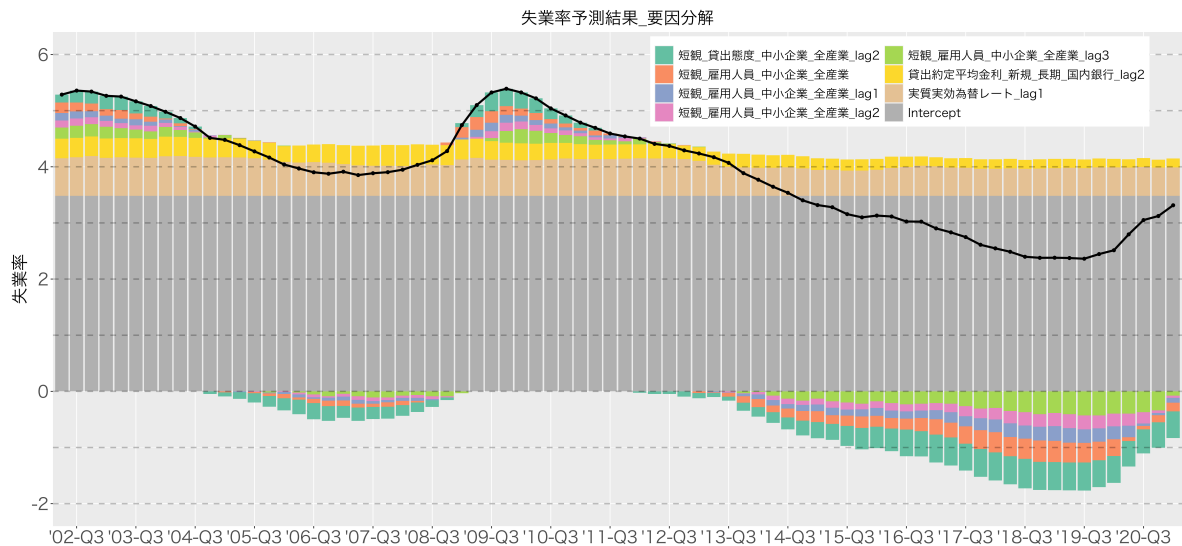


図 3.13: 完全失業率 1 期先モデルの要因分解図

3.7 結果

図 3.13 より、世界金融危機時と比較してコロナ禍の完全失業率が低水準に止まっている主要因は、雇用人員 DI から推測される雇用需要の長期拡大、貸出約定平均金利の低下、賃

出態度 DI から推測される貸出態度の軟化，実質実効為替レートの低下であることが示唆された。また，2014 年第 1 四半期以降の完全失業率減少もこれら 4 要因が説明に寄与することから本予測モデルがマクロな経済動向から完全失業率への影響を定量的に評価する可能性を持つことも示唆される。これについては，評価事項を明確にした上で他要因を考慮する必要がありさらなる研究が必要である。完全失業率の 2 四半期先 (2021 年第 2 四半期) までの予測結果を実績値と併せて表にまとめる。また，実績値と予測値の推移を図 3.14 に示す。

表 3.11: 完全失業率予測結果

	'20-Q1 実績値	'20-Q2 実績値	'20-Q3 実績値	'20-Q4 実績値	'21-Q1 予測値	'21-Q2 予測値
95% 信頼区間上限	-	-	-	-	3.84	3.75
完全失業率	2.43	2.77	2.97	2.99	3.31	3.28
95% 信頼区間下限	-	-	-	-	2.79	2.81

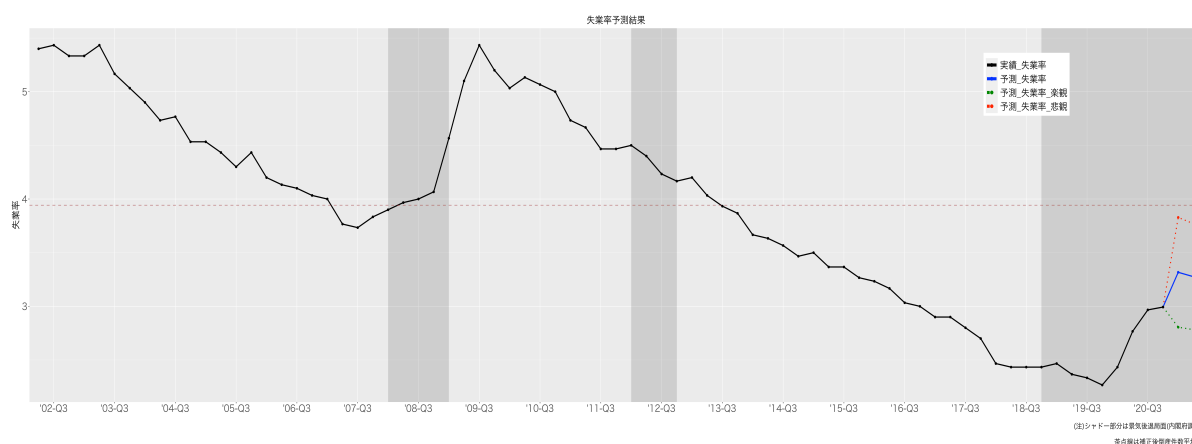


図 3.14: 完全失業率予測推移

図 3.14 より，95% 信頼区間上限の予測であっても推定対象期間平均 (3.94%) を下回る予測となっており，本研究による予測では 2021 年第 2 四半期まで完全失業率は低水準で推移すると予測している。ただし，本モデルの予測に用いたデータは主に 11 月調査のものであるため足許の景況感を含めた予測でないことには注意が必要である。

第4章 就業者数予測

本章では、コロナ禍によるダメージが大きかった業種を分析するために業種別のデータが公開されている就業者数に注目し作成した、コロナ禍での就業者数予測モデルについて述べる。はじめに今回の分析対象とする業種について説明し、モデル作成に用いたデータの詳細を説明、推定結果および第2四半期先までの予測結果、推定量の要因分解による予測に寄与する要因の考察を行う。なお、業種は日本標準産業分類に基づき扱う。

4.1 業種別の就業者推移

日本標準産業分類の業種別にコロナ禍に相当する期間2020年4月9月での就業者数前年同月比較を行う。業種別平均の前年同月比を図4.1に、前年同月数を図4.2に示す。

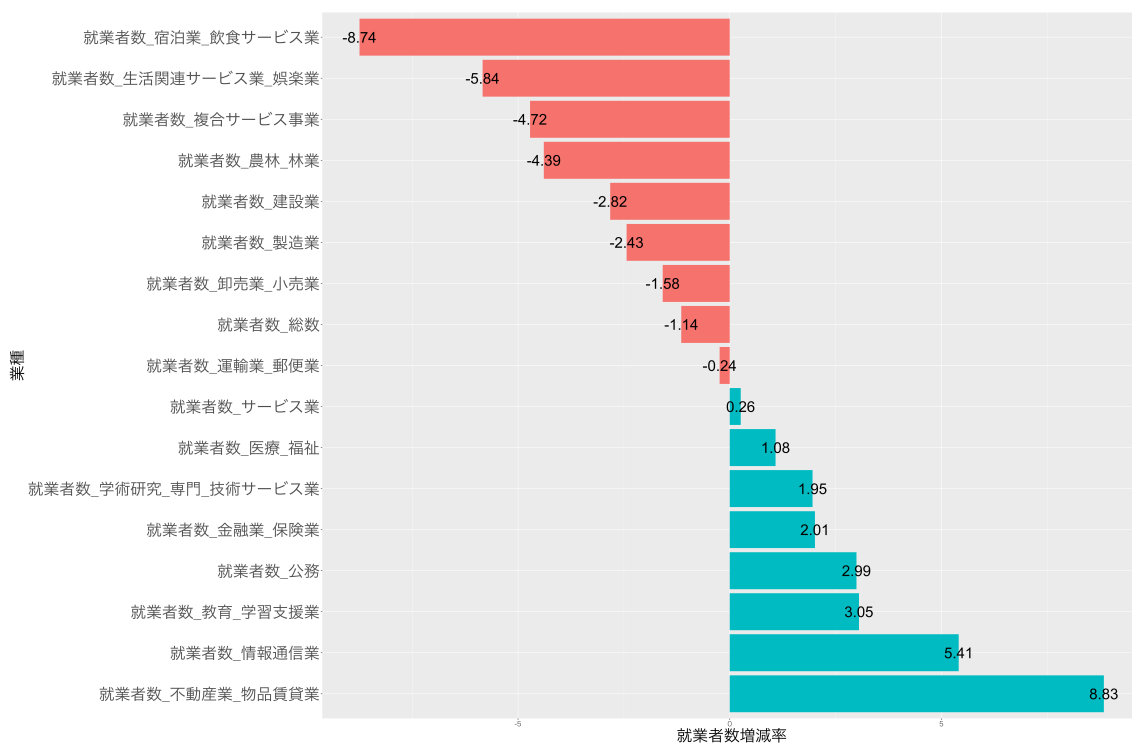


図 4.1: 業種別前年同月比

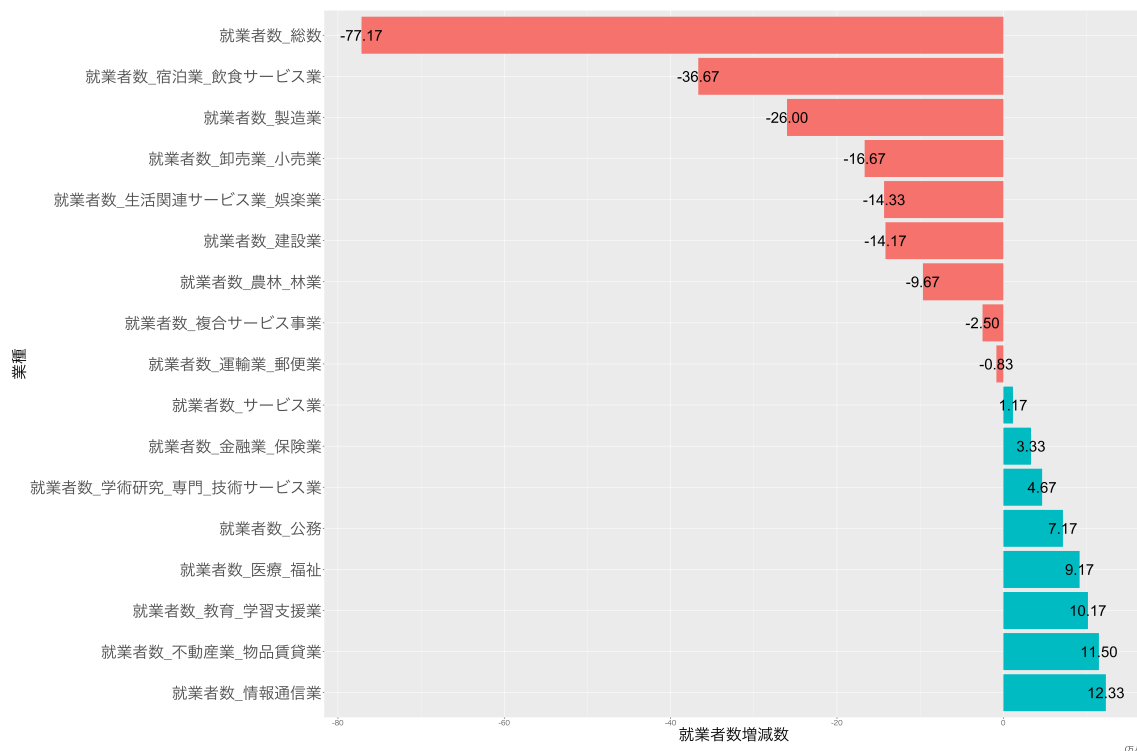


図 4.2: 業種別前年同月数

上記より、本研究では最も就業者減少が大きい飲食・宿泊サービス業を対象を絞って就業者数予測モデルを作成することとした。

4.2 飲食・宿泊サービス業就業者数予測モデル

以下では、飲食・宿泊サービス業就業者数予測モデルについて用いたデータおよび推定方法、考察を述べていく。

4.3 使用データ

本モデルで使用する日銀短観、貸出約定平均金利、実質実効為替レートは、倒産件数予測、完全失業率予測で用いたものと同様のデータを使用する。但し、日銀短観データは一部業種別に公表されており、今回は宿泊・飲食サービス業の短観データを利用する。本節では、新たに使用する就業者数データと住宅宿泊事業者数データについて述べる。

4.3.1 就業者数データ

就業者数とは、従業者と休業者を合わせた数である。本研究で使用する業種別就業者数データは、総務省統計局が実施している労働力調査に基づくものである [14]。なお、業種は日本標準産業分類に基づいている。詳細は総務省統計局 HP「労働力調査用語の解説」 [16] を参照されたい。

飲食・宿泊サービス業就業者数予測モデルでは、このデータを予測対象としてモデルを構築する。以下に四半期ごと 2008 年第 1 四半期 (2008-Q1) から 2020 年第 3 四半期 (2020-Q3) 完全失業率推移を示す。

これは総務省統計局が公表する労働力調査の月次データを四半期ごとに足し合わせ作成した。なお、シャドー部分は内閣府が発表する景気基準日付 [6] を元にした景気後退局面である。2018 年第 4 四半期 (2018-Q4) 以降の景気後退局面は暫定であることに注意されたい。

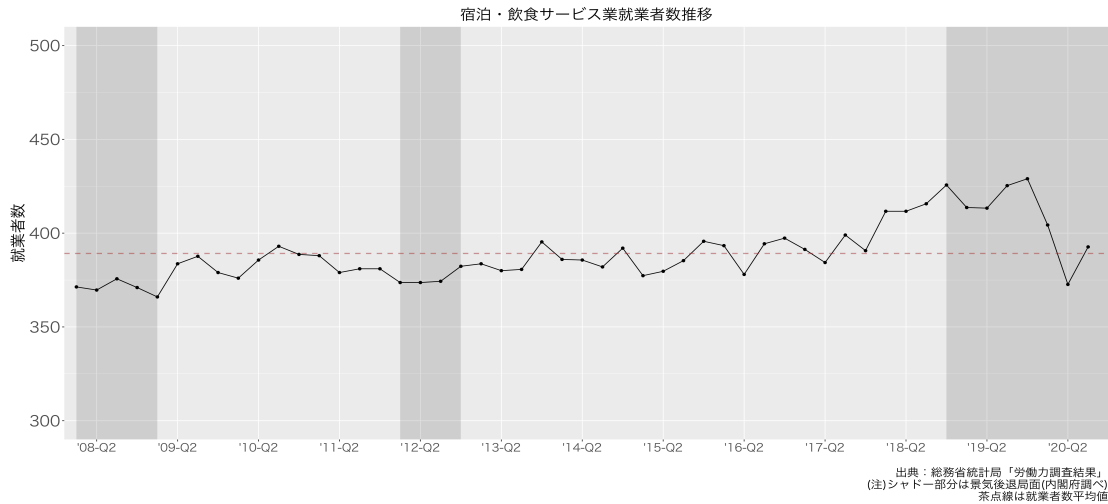


図 4.3: 飲食・宿泊サービス業就業者数の推移 (四半期)

飲食・宿泊サービス業就業者数推移の特徴として、2017 年第 1 四半期まで緩やかに上昇し 2018 年第 1 四半期からさらなる増加、2020 年の前半は落ち込んだが足許は推定期間平均まで回復の傾向が見られる。

4.3.2 住宅宿泊事業者数データ

民泊サービスを営む事業者が各都道府県知事に対して届出を出した数のデータ [9]。飲食・宿泊サービス業就業者数の 2018 年第 1 四半期からの増加を訪日外国人増加に伴う民泊サービスの普及に伴うものと推測し、本データを説明変数として使用した。住宅宿泊事業者数は、毎月官公庁が運営する民泊ポータルサイトにて公表されておりこのデータの 3・6・9・12 月を四半期ごとのデータとして集計し利用した。なお、住宅宿泊事業法施行日 (2018 年 6 月 15 日) 以前はデータが存在せず 0 となっている。詳細は、官公庁「民泊ポータルサイト」を参照されたい [9]。

4.3.3 民泊解禁効果データ

2018 年第 1 四半期から第 4 四半期は民泊解禁による特需が存在すると仮定し、重回帰モデルの推定を行う上で仮の効果量を表すデータを作成した。こちらは、2018 年第 1 四半期から第 4 四半期までを、「1」、「0.75」、「0.5」、「0.25」と置きそれ以外の期間は「0」としている。

4.3.4 季節ダミーデータ

就業者数の季節性を考慮し、Q1 ダミー、Q2 ダミー、Q3 ダミーを用いる。これらは該当する四半期なら「1」、それ以外「0」とおいたダミー変数である。

4.4 推定方法

2.1 節と 3.1 節、4.3 節で述べたデータを用いて飲食・宿泊サービス業就業者数を予測するための重回帰モデルを構築する。飲食・宿泊サービス業就業者数予測モデルは、1 四半期先飲食・宿泊サービス業就業者数予測モデルと 2 四半期先飲食・宿泊サービス業就業者数予測モデルの 2 種類を作成した。

4.4.1 飲食・宿泊サービス業就業者数予測モデル構築の問題点

飲食・宿泊サービス業就業者数予測モデル構築に当たり、2.2.1 小節、3.2.1 小節と同様の問題が 2 点存在する。

問題点 1：説明変数の時点制約

問題点 2：公表の速報性

従って、飲食・宿泊サービス業就業者数予測モデルにおいても説明変数の適切な時点データの使用とサブモデル作成による速報性の担保を行う。

ここで飲食・宿泊サービス業就業者数予測の全体像を俯瞰するために、モデル構築の概要を図示する。

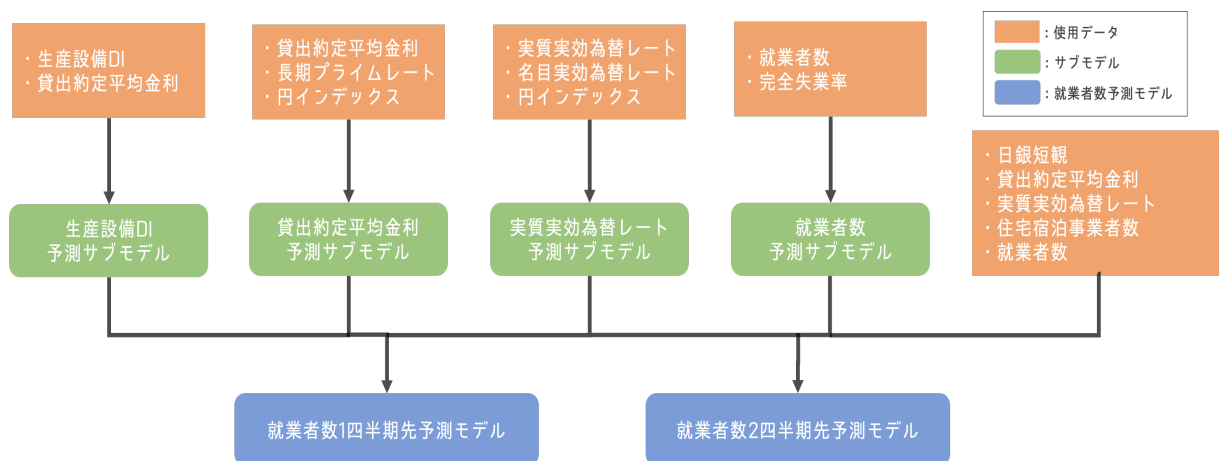


図 4.4: 飲食・宿泊サービス業就業者数予測モデル構築概要

4.5 サブモデルおよびシナリオ分析

本節では、飲食・宿泊サービス業就業者数予測公表の速報性を確保するために作成した生産設備 DI、飲食・宿泊サービス業就業者数のサブモデルを説明する。なお、本モデルで

使用する貸出約定平均金利と実質実効為替レートのサブモデルは倒産件数予測モデル時に説明したものと同様のものを使用するため説明を省略する。また、住宅宿泊事業者数はシナリオを仮定し、データ補完を行なったためその説明を本節で行う。

4.5.1 生産設備 DI サブモデル

生産設備 DI は、日本銀行短観データのため四半期ごとに公表される。飲食・宿泊サービス業就業者数 2 期先モデルでは生産設備 DI の当期を使用する。そのため 12 月中旬時点での 2 期先予測のためには 2021 年第 2 四半期の値が必要となる。そこで生産設備 DI サブモデルを作成した。推定期間は、飲食・宿泊サービス業就業者数モデルの推定対象期間に合わせた 2004 年第 1 四半期から 2020 年第 4 四半期とした。生産設備 DI の四半期時系列推移を図 4.5 に示す。

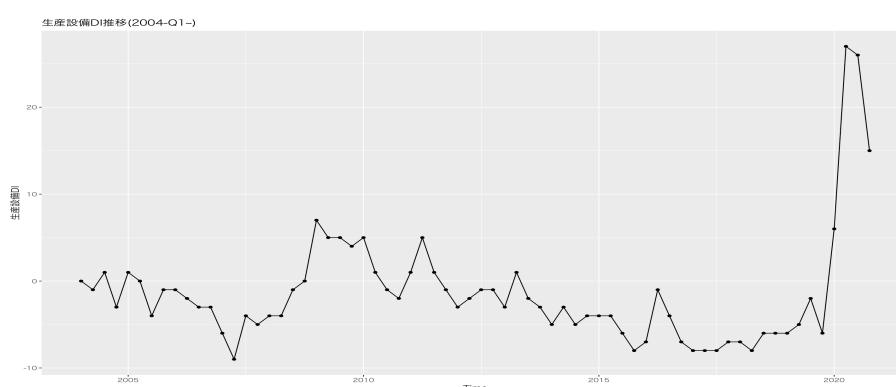


図 4.5: 生産設備 DI の四半期推移

生産設備 DI データに対して、ARIMA(p,d,q) モデルで AIC 基準で最小となるモデル ARMA(1,1) が選択された。しかし、ARMA(1,1) は誤差項の不均一分散の仮定が検定結果より満たされておらず予測モデルとして使用するのに適さないと判断し、VAR(p) によるモデル推定を行う。ここでは、生産設備 DI、業況 DI、雇用人員 DI、貸出約定平均金利の 4 変数 VAR モデルから AIC 基準最小となるモデルを選択する。VAR(p) モデルの探索結果は表 4.1 である。

表 4.1: 4 変数 VAR(p) 探索結果

生産設備 DI	業況 DI	雇用人員 DI	貸出約定平均金利	選択次数	AIC
○				2	2.735063
○	○			2	6.428616
○		○		3	5.724777
○			○	2	-3.23567790
○	○	○		3	9.369340
○	○		○	2	0.4242934
○		○	○	3	-0.31801855
○	○	○	○	3	3.308857

表 4.1 より、生産設備 DI と貸出約定平均金利の 2 変数 VAR(2) が AIC 基準で最小となる。従って本モデルを生産設備 DI 予測サブモデルとする。2 変数 VAR(2) の推定結果を以下に

まとめる.

表 4.2: 生産設備 DI サブモデル 2 変数 VAR(2) 推定結果

説明変数	被説明変数:	
	生産設備 DI	貸出約定平均金利
生産設備 DI_lag1	1.0817*** (6.59e-12)	-0.000413 (0.789)
貸出約定平均金利_lag1	-4.6526 (0.6576)	1.174689*** (2.27e-13)
生産設備 DI_lag2	-0.3650** (0.0209)	0.001693 (0.364)
貸出約定平均金利_lag2	4.5687 (0.6631)	-0.191450 (0.133)
定数項	-0.3292 (0.8563)	0.013097 (0.550)
観測数	68	
誤差項の独立性:Ljung and Box(lag=12)	24.16 (0.977)	
誤差項の不均一分散:ARCH-LM(lag=12)	0.287 (0.287)	

()内は p 値 Note:*p<0.1; **p<0.05; ***p<0.01

表 4.2 の 2 変数 VAR(2) 推定結果より, 系列相関, 不均一分散の問題に対処した推定値が得られた. さらに VAR 構造の定常性の診断を特性方程式の同伴行列の固有値をもとに行なった. 以下に, 特性方程式の同伴行列の固有値を複素平面上にプロットして示す.

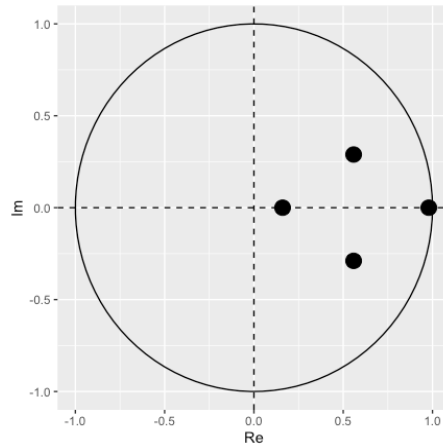


図 4.6: VAR 特性方程式の固有値プロット

図 4.6 より，特性方程式の同伴行列の固有値が単位円内 (固有値が 1 以下) にあるため，推定モデル 2 変量 VAR(2) は AR 構造の定常性が満たされている。

上記診断結果より，生産設備 DI 予測サブモデルとして生産設備 DI と貸出約定平均金利の 2 変量 VAR(2) を使用する。生産設備 DI 予測サブモデルによる，2021 年第 2 四半期の予測結果は以下となる。なお，2021 年第 1 四半期は日本銀行が公表している予測値を用い，2021 年第 2 四半期予測の際に使用する。

表 4.3: 生産設備 DI 予測結果

年月	'20-Q2 実績値	'20-Q3 実績値	'20-Q4 実績値	'21-Q1 公表値	'21-Q2 予測値
生産設備 DI	27	26	15	15	10.14

飲食・宿泊サービス業就業者数予測モデル推定のデータとして，表 4.3 の生産設備 DI2021 年第 2 四半期予測値を利用する。

4.5.2 飲食・宿泊サービス業就業者数サブモデル

飲食・宿泊サービス業就業者数は，総務省統計局による労働力調査で集計・公表されており翌月に公表される。12 月中旬時点で飲食・宿泊サービス業就業者数の 2 期先予測を行うには翌年 1 月に公表される 12 月分飲食・宿泊サービス業就業者数を予測する必要がある。そこで，飲食・宿泊サービス業就業者数予測サブモデルを作成した。推定期間は，2008 年 1 月から 2020 年 11 月とする。

定常性を満たす ARIMA(p,d,q) モデルを，AICC 最小となるように探索した。探索結果が表 4.4 である。なお，最大探索範囲は p,d,q 全て 12 までとした。

表 4.4: ARIMA(p,d,q) 探索結果

p	d	q	AICC
0	0	0	1279.32
0	0	1	1146.243
0	0	2	1126.997
0	0	3	1088.675
0	0	4	1087.748
0	0	5	1087.857
1	0	0	1068.495
1	0	1	1070.191
1	0	2	1057.188
1	0	3	1057.402
1	0	4	1057.962
2	0	0	1070.424
2	0	1	1063.767
2	0	2	1055.861
2	0	3	1057.293
3	0	0	1063.241
3	0	1	1059.898
3	0	2	1057.06
4	0	0	1065.268
4	0	1	1061.126
5	0	0	1054.46

表 4.4 より、AICC 基準で最小となるモデルは AR(5) となる。飲食・宿泊サービス業就業者数推移を AR 過程で表現できる可能性が示唆されたことより VAR(p) によるモデル推定を行う。ここでは、飲食・宿泊サービス業就業者数、完全失業率の 2 変数 VAR モデルから AIC 基準最小となるモデルを選択する。VAR(p) モデルの探索結果は表 4.5 である。

表 4.5: 2 変数 VAR(p) 探索結果

就業者数	完全失業率	選択次数	AIC
○		5	4.004284
○	○	5	-0.07034429

表 4.5 より、飲食・宿泊サービス業就業者数と完全失業率の 2 変数 VAR(5) が AIC 基準で最小となる。従って本モデルを飲食・宿泊サービス業就業者数予測サブモデルとする。2 変数 VAR(5) の推定結果を以下にまとめる。

表 4.6: 飲食・宿泊サービス業就業者数予測サブモデル 2 変数 VAR(5) 推定結果

説明変数	被説明変数:	
	就業者数	完全失業率
就業者数_lag1	0.79537*** ($<2e-16$)	4.809e-05 (0.97222)
完全失業率_lag1	-2.74829 (0.58003)	8.135e-01*** ($<2e-16$)
就業者数_lag2	-0.23642** (0.02678)	-3.422e-03* (0.05457)
完全失業率_lag2	0.71628 (0.91183)	1.748e-01 (0.10747)
就業者数_lag3	0.24839** (0.02140)	3.908e-03** (0.03014)
完全失業率_lag3	-0.38318 (0.95135)	1.825e-01* (0.08378)
就業者数_lag4	-0.20889* (0.05586)	-1.959e-03 (0.28110)
完全失業率_lag4	1.93425 (0.75610)	2.343e-02 (0.82189)
就業者数_lag5	0.26592*** (0.00216)	-2.683e-04 (0.85058)
完全失業率_lag5	-0.78018 (0.87164)	-2.241e-01*** (0.00616)
定数項	57.58808** (0.03708)	7.650e-01* (0.09653)
観測数	154	
誤差項の独立性:Ljung and Box(lag=12)	30.99 (0.317)	
誤差項の不均一分散:ARCH-LM(lag=12)	114.74 (0.311)	

() 内は p 値 Note:*p<0.1; **p<0.05; ***p<0.01

表 4.6 の 2 変数 VAR(5) 推定結果より，系列相関，不均一分散の問題に対処した推定値が得られた。さらに VAR 構造の定常性の診断を特性方程式の同伴行列の固有値をもとに行なった。以下に，特性方程式の同伴行列の固有値を複素平面上にプロットして示す。

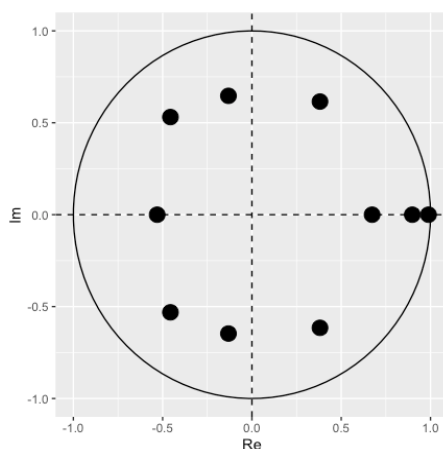


図 4.7: VAR 特性方程式の固有値プロット

図 4.7 より，特性方程式の同伴行列の固有値が単位円内 (固有値が 1 以下) にあるため，推定モデル 2 変数 VAR(5) は AR 構造の定常性が満たされている。

上記診断結果より，飲食・宿泊サービス業就業者数予測サブモデルとして飲食・宿泊サービス業就業者数と完全失業率の 2 変数 VAR(5) を使用する。飲食・宿泊サービス業就業者数予測サブモデルによる，2020 年 12 月の予測結果は以下となる。

表 4.7: 飲食・宿泊サービス業就業者数予測結果

年月	2020/08 実績値	2020/09 実績値	2020/10 実績値	2020/11 実績値	2020/12 予測値
完全失業率	391	403	410	391	386

飲食・宿泊サービス業就業者数予測モデル推定のデータとして，表 4.12 の飲食・宿泊サービス業就業者数 2020 年 12 月予測値を利用する。

4.5.3 住宅宿泊就業者数シナリオ予測

住宅宿泊事業者数は，毎月公表される。飲食・宿泊サービス業就業者数予測モデルでは住宅宿泊事業者数の当期を使用する。そのため 12 月中旬時点で 2 期先予測のためには 2021 年第 1，第 2 四半期の値が必要となる。本稿では，住宅宿泊事業者数のデータ数および変動から予測モデル構築が困難と判断しシナリオを想定し予測値を算出することとした。シナリオ予測の方法としては，足許の減少速度 (毎月 133 件の減少) が継続すると予測した。以下に，予測値を図表で示す。

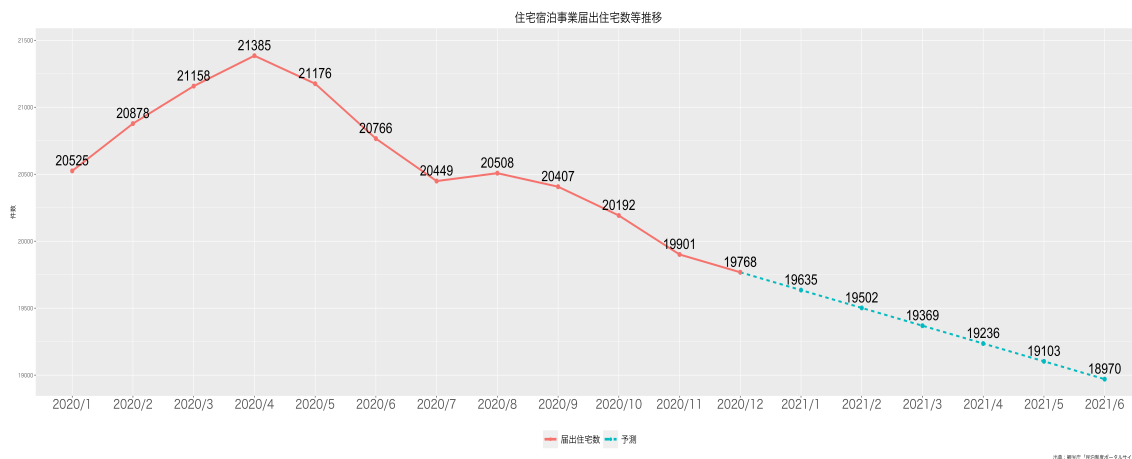


図 4.8: 住宅宿泊事業者数シナリオ予測

表 4.8: 住宅宿泊事業者数シナリオ予測結果

年月	'20-Q2 実績値	'20-Q3 実績値	'20-Q4 実績値	'21-Q1 予測値	'21-Q2 予測値
生産設備 DI	20,766	20,407	19,768	19,369	18,970

飲食・宿泊サービス業就業者数予測モデルのデータとして、表 4.8 の住宅宿泊事業者数数 2021 年第 1, 第 2 四半期予測値を利用する。

4.6 推定結果

4.6.1 説明変数の選択結果

飲食・宿泊サービス業就業者数 1 期・2 期先予測モデルそれぞれの説明変数候補の中から符号条件が合致するモデルの中から AIC 最小基準でモデル選択を行なった。説明変数の候補として使用した変数およびその符号条件、各モデルが最終的に使用する説明変数を以下に表形式でまとめる。

表 4.9: 説明変数選択結果

対象指標	説明変数候補	時点	符号条件	1期先	2期先
業況	業況 DI(全企業・宿泊飲食)	当期	正		
人件費負担	雇用人員 DI(全企業・宿泊飲食)	当期	負		
設備費負担	生産設備 DI(全企業・宿泊飲食)	当期	負	○	○
採算性	販売価格 DI-仕入価格 DI(全企業・宿泊飲食)	当期	正		
		前期	正		
借入環境	資金繰り DI(全企業・宿泊飲食)	前期	正		
		前々期	正		
	貸出態度 DI(全企業・宿泊飲食)	前期	正		
		前々期	正		
金利面	貸出約定平均金利(新規・長期・国内銀行)	前期	負	○	
		前々期	負		○
対外競争力	実質実効為替レート	前期	正	○	
		前々期	正		○
特需要因	住宅宿泊事業者数	当期	正	○	○
	民泊解禁効果	-	-	○	○
季節調整	Q1 ダミー	-	-	○	○
	Q2 ダミー	-	-	○	○
	Q3 ダミー	-	-		

表 4.9 より、1 期先飲食・宿泊サービス業就業者数予測モデルは、生産設備 DI・貸出約定平均金利・実質実効為替レート・住宅宿泊事業者数・民泊解禁効果・季節ダミーを含む 7 変数の重回帰モデル、2 期先飲食・宿泊サービス業就業者数予測モデルも、生産設備 DI・貸出約定平均金利・実質実効為替レート・住宅宿泊事業者数・民泊解禁効果・季節ダミーを含む 7 変数の重回帰モデルを予測モデルとして採用する。但し、日銀短観の生産設備 DI は、前期調査時の 1 四半期先予測値が日本銀行から公表されている。そこで、1 四半期先モデルの予測の際は「当期」をこの 1 四半期先予測値で代用する。

4.6.2 飲食・宿泊サービス業就業者数 1 四半期先予測モデル

4.9 で選択された 1 期先飲食・宿泊サービス業就業者数予測用 7 変数重回帰モデルのパラメータ推定結果を示す。ここで、通常の OLS を行ない DW 検定の結果、帰無仮説が有意水準 5% で棄却され誤差項が 1 階の自己相関構造を持つ可能性が示唆された。従って、PW 変換による GLS 推定を行いその L-R プロットを以下に示す。赤点は外れ値 (Y 方向の影響点)、紫点は高い作用点 (X 方向の影響点) を表す。

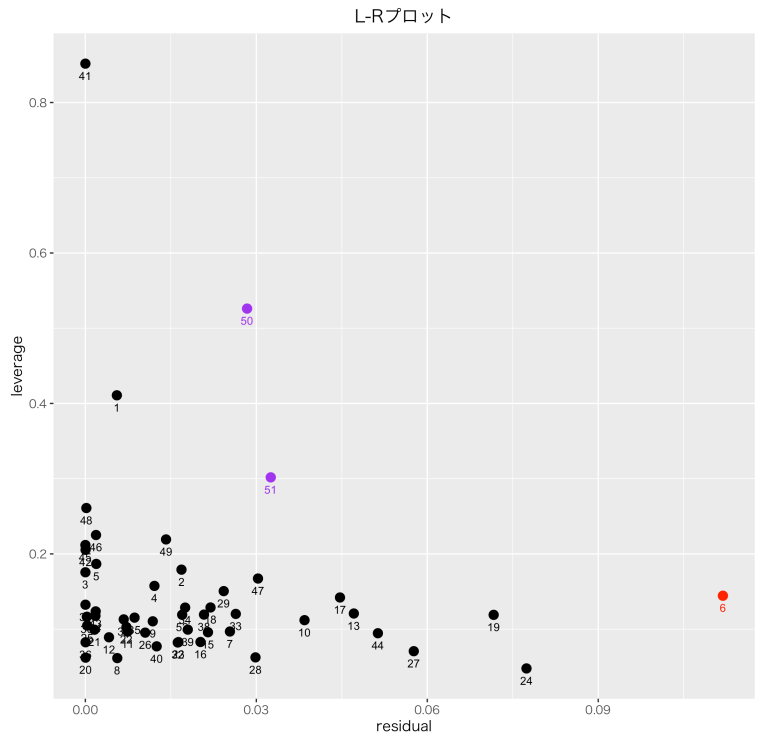


図 4.9: 飲食・宿泊サービス業就業者数 1 期先モデル L-R プロット

図 4.9 にて、赤点・紫点は式 A.39 基準に従って検出した影響点を指す。影響点を 3 点検出したため PW 変換後に MM 推定を行い頑健なパラメータ推定を行なった。1 期先モデルの GLS・MM 推定結果を表 4.10 にまとめる。

MM 推定結果より、多重共線性、系列相関、不均一分散の問題に対処した推定値が得られたと判断。従って、表 4.10 の結果を飲食・宿泊サービス業就業者数 1 期先予測モデルとして利用する。また、MM 推定による加重最小二乗推定の各データ点へのウェイトを棒グラフで表現し図 4.10 に示す。図 4.10 から、外れ値(赤棒線)を中心にウェイトを低下させ影響点による推定量への影響を制御していることがわかる。

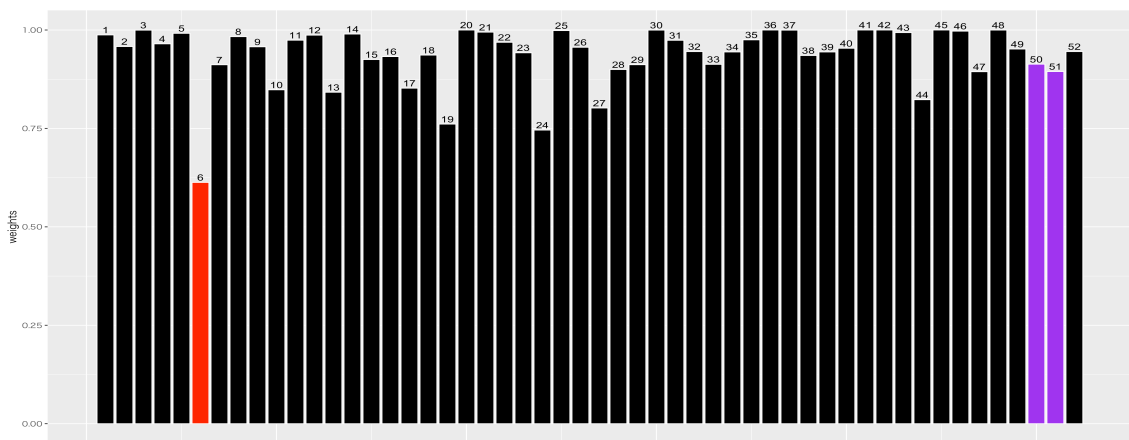


図 4.10: 飲食・宿泊サービス業就業者数 1 期先予測モデル MM 推定の各ウェイト

表 4.10: 飲食・宿泊サービス業就業者数 1 期先予測モデル MM 推定結果

説明変数	時点	符号条件	被説明変数:	
			就業者数	VIF
生産設備 DI	当期	負	-1.2371*** (5.3e-07)	1.475
貸出約定平均金利	前期	負	-13.6432** (0.039)	2.470
実質実効為替レート	前期	正	0.2372 (0.111)	2.215
住宅宿泊事業者数	当期	正	0.0016*** (1.3e-06)	1.782
民泊解禁効果	-	-	24.6394*** (3.5e-04)	1.097
季節 Q1 ダミー	-	-	-5.5683*** (0.009)	1.327
季節 Q2 ダミー	-	-	-6.0549*** (0.005)	1.310
定数項	-	-	378.7088*** (4.0e-31)	-
観測数	52	DW 検定	2.0785	
決定係数	0.8559962		(0.4169)	
自由度修正済決定係数	0.8437406	BP 検定	4.5744	
Residual Std. Error	6.356		(0.7117)	

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

4.6.3 飲食・宿泊サービス業就業者数 2 四半期先予測モデル

4.9 で選択された 2 期先飲食・宿泊サービス業就業者数予測用 7 変数重回帰モデルのパラメータ推定結果を示す。ここで、通常の OLS を行ない DW 検定の結果、帰無仮説が有意水準 5% で棄却され誤差項が 1 階の自己相関構造を持つ可能性が示唆された。従って、PW 変換による GLS 推定を行いその L-R プロットを以下に示す。赤点は外れ値 (Y 方向の影響点)、紫点は高い作用点 (X 方向の影響点) を表す。

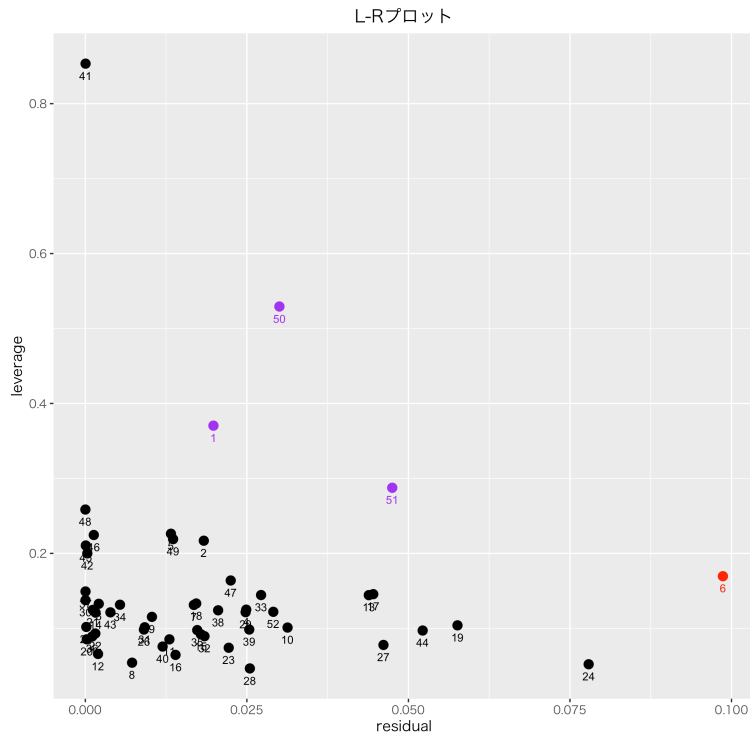


図 4.11: 飲食・宿泊サービス業就業者数 2 期先モデル L-R プロット

図 4.11 にて、赤点・紫点は式 A.39 基準に従って検出した影響点を指す。影響点を 4 点検出したため PW 変換後に MM 推定を行い頑健なパラメータ推定を行なった。2 期先モデルの GLS・MM 推定結果を表 4.11 にまとめる。

MM 推定結果より、多重共線性、系列相関、不均一分散の問題に対処した推定値が得られたと判断。従って、表 4.11 の結果を飲食・宿泊サービス業就業者数 2 期先予測モデルとして利用する。また、MM 推定による加重最小二乗推定の各データ点へのウェイトを棒グラフで表現し図 4.12 に示す。図 4.12 から、外れ値(赤棒線)を中心にウェイトを低下させ影響点による推定量への影響を制御していることがわかる。

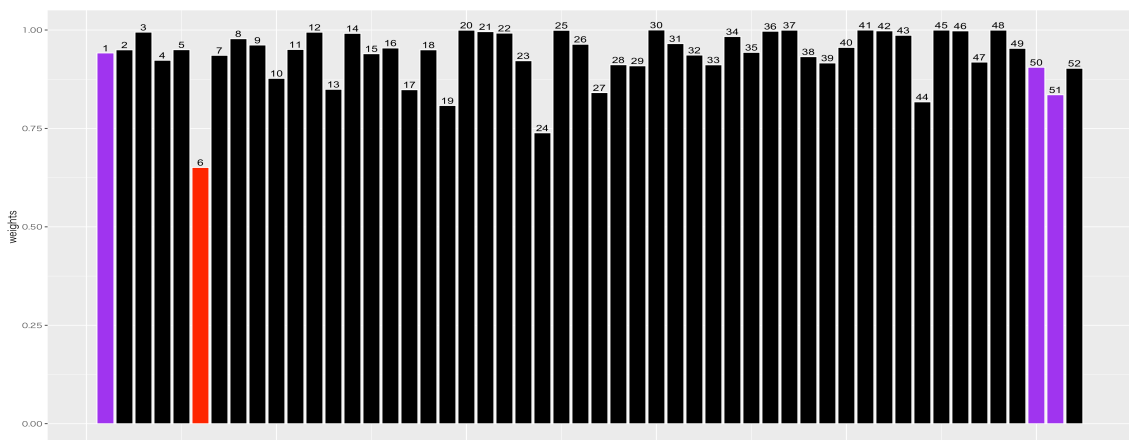


図 4.12: 飲食・宿泊サービス業就業者数 2 期先予測モデル MM 推定の各ウェイト

表 4.11: 飲食・宿泊サービス業就業者数 2 期先予測モデル MM 推定結果

説明変数	時点	符号条件	被説明変数:	
			就業者数	VIF
生産設備 DI	当期	負	-1.2557*** (7.6e-07)	1.819
貸出約定平均金利	前々期	負	-8.7344 (0.148)	2.329
実質実効為替レート	前々期	正	0.1901 (0.179)	1.991
住宅宿泊事業者数	当期	正	0.0017*** (1.4e-06)	1.446
民泊解禁効果	-	-	25.4477*** (4.0e-04)	1.104
季節 Q1 ダミー	-	-	-5.4411** (0.013)	1.325
季節 Q2 ダミー	-	-	-5.8061*** (0.009)	1.313
定数項	-	-	377.1072*** (4.6e-30)	-
観測数	52	DW 検定	2.0488	
決定係数	0.8383779		(0.3744)	
自由度修正済決定係数	0.8126653	BP 検定	5.1825	
Residual Std. Error	6.488		(0.6377)	

() 内は p 値 Note: *p<0.1; **p<0.05; ***p<0.01

4.7 予測結果

4.7.1 飲食・宿泊サービス業就業者数 1 四半期先予測結果

4.6.2 のモデルを使用して 1 期先 (2021 年第 1 四半期) の飲食・宿泊サービス業就業者数を予測する。予測値とその 95% 信頼区間および実績値推移のプロットを図 4.13 に示す。なお、縦点線は MM 推定量を算出する際に用いたデータ期間を表す (2008 年第 1 四半期から 2020

年第4四半期)。図4.13の推移より、実績値が推定対象全期間で95%信頼区間に収まっており、飲食・宿泊サービス業就業者数推移の変化を概ね捉えている。本モデルによる2021年第1四半期予測値を1期先予測として使用する。

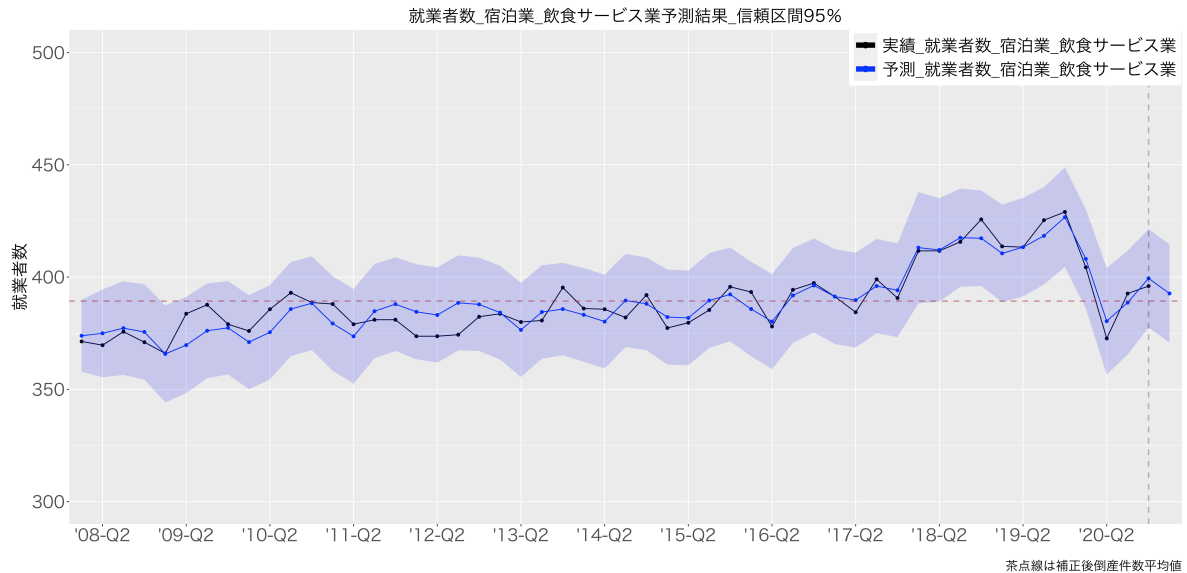


図 4.13: 飲食・宿泊サービス業就業者数 1 期先モデルの予測結果

4.7.2 飲食・宿泊サービス業就業者数 2 四半期先予測結果

4.6.3 のモデルを使用して 2 期先 (2021 年第 2 四半期) の飲食・宿泊サービス業就業者数を予測する。予測値とその 95% 信頼区間および実績値推移のプロットを図 4.14 に示す。なお、縦点線は MM 推定量を算出する際に用いたデータ期間を表す (2008 年第 1 四半期から 2020 年第 4 四半期)。図 4.14 の推移より、実績値が推定対象全期間で 95% 信頼区間に収まっており、飲食・宿泊サービス業就業者数推移の変化を概ね捉えている。本モデルによる 2021 年第 2 四半期予測値を 2 期先予測として使用する。

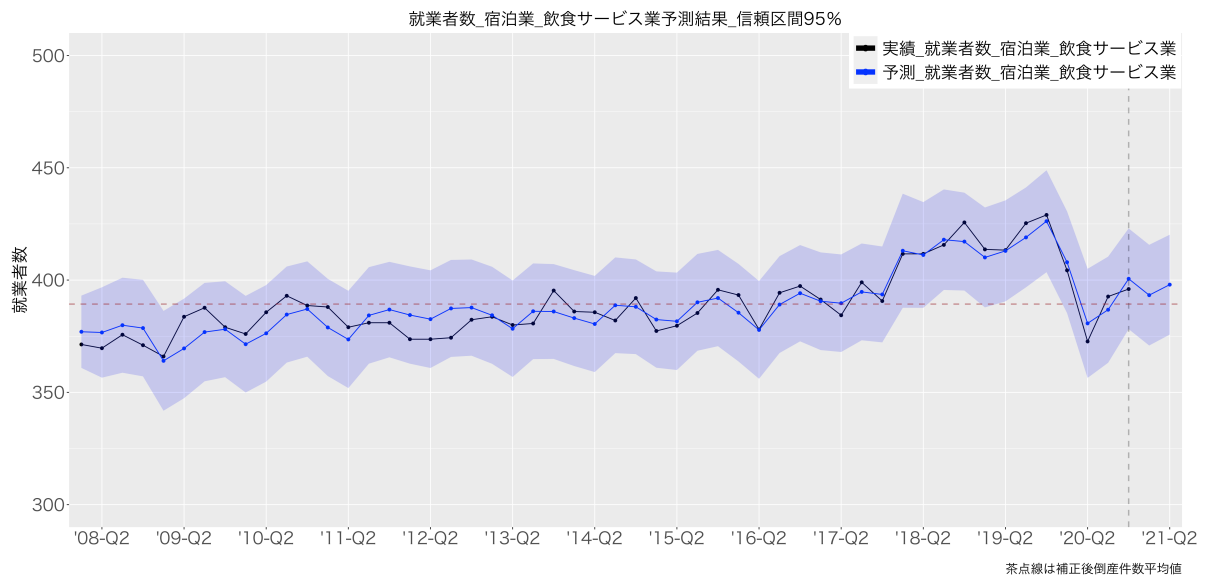


図 4.14: 飲食・宿泊サービス業就業者数 2 期先モデルの予測結果

4.8 予測の要因分解

1 期先モデルの要因分解図を以下に示す。

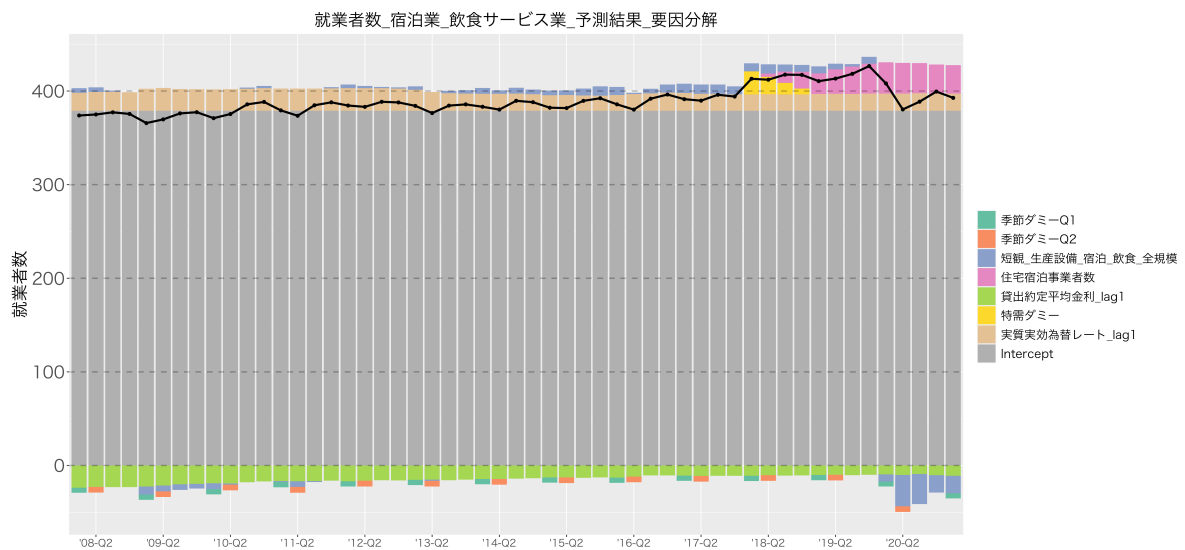


図 4.15: 飲食・宿泊サービス業就業者数 1 期先モデルの要因分解図

4.9 結果

図 4.15 より、2017 年第 4 四半期までの右肩上がり傾向は生産設備 DI から推測される需要の長期拡大と経営環境の改善、2018 年第 1 四半期から第 4 四半期までの一層の増加は広義

の民泊解禁効果，2020年第1，第2四半期の急な落ち込みは生産設備DIから推測される需要の急激な減少と固定費負担の増大が示唆された。これについては，評価事項を明確にした上で他要因を考慮する必要がありさらなる研究が必要である。飲食・宿泊サービス業就業者数の2四半期先(2021年第2四半期)までの予測結果を実績値と併せて表にまとめる。なお，予測値は原系列と季節調整値両方を載せる。また，実績値と予測値の推移を図4.16に示す。

表 4.12: 飲食・宿泊サービス業就業者数予測結果

	'20-Q1 実績値	'20-Q2 実績値	'20-Q3 実績値	'20-Q4 実績値	'21-Q1 予測値	'21-Q2 予測値
95% 信頼区間上限	-	-	-	-	415	420
就業者数 (季節調整)	410	379	393	396	398	403
就業者数 (原系列)	404	373	393	396	382	397
95% 信頼区間下限 (季節調整)	-	-	-	-	375	382
95% 信頼区間下限 (原系列)	-	-	-	-	370	376

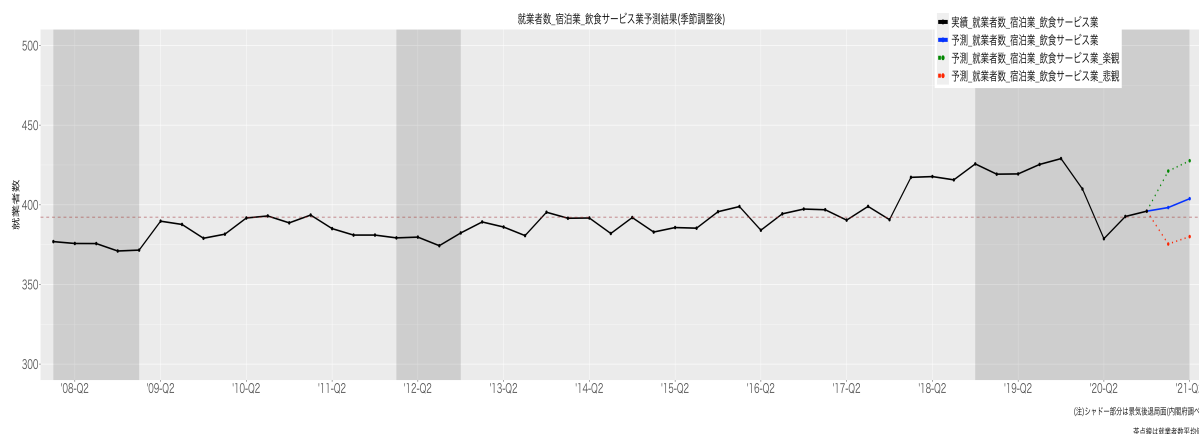


図 4.16: 飲食・宿泊サービス業就業者数予測推移

図 4.16 より，飲食・宿泊サービス業就業者数は 2020 年第 3 四半期に回復後，推定対象期間平均 (392.2 万人) を若干上回る水準で推移する予測となっている。ただし，95% 信頼区間下限値では 2021 年第 1 四半期予測は，コロナ禍での最悪水準 (2020 年第 2 四半期:379 万人) を若干下回る水準となっている。なお，本モデルの予測に用いたデータは主に 11 月調査のものであるため足許の景況感を含めた予測でないことには注意が必要である。

第5章 考察と今後の課題

本研究では、経済問題の内、特に深刻な影響が懸念される倒産・失業問題に注目し、EBPM過程に資する倒産・失業関連指標予測モデルの開発を行った。

まず、倒産件数予測モデルの2四半期先予測結果は、95%信頼区間上限(悲観シナリオ)の予測であっても推定対象期間平均を下回る予測となっており2020年12月時点の景況感では倒産に深刻な影響は見られなかった。予測に寄与する主要因は、貸出約定平均金利の低下、貸出態度の軟化、為替レートの低下が起因しており、これら結果は今後、倒産・失業問題が深刻化した時のEBPM過程の実行に有用となり得る結果が示唆された。

次に倒産件数に表れていない失業問題の深刻さを捉える目的で、完全失業率予測モデルを作成した。同モデルの2四半期先予測結果は、95%信頼区間上限(悲観シナリオ)の予測であっても推定対象期間平均を下回る予測となっており2020年12月時点の景況感では倒産件数と同様に深刻な影響は見られなかった。予測に寄与する主要因は、雇用需要の長期拡大、貸出約定平均金利の低下、貸出態度の軟化、実質実効為替レートの低下が起因している。今後、景況感の悪化に伴う失業問題が深刻化した際、これらの要因はEBPM過程の実行に有用であることが示唆される。

最後に、倒産件数・失業率両指標には顕在化していない倒産・失業の深刻さを特定すべく業種別の就業者数を対象として分析を行った。まず、業種別に前年同期比・数との比較を行い最も減少が大きかった飲食・宿泊サービス業の就業者数を本研究での分析対象とした。飲食・宿泊サービス業の就業者数を対象として同様の予測モデルを作成した。同モデルの2四半期先予測結果は、推定対象期間平均を若干上回る予測となっており2020年12月時点の景況感では飲食・宿泊サービス業の就業者数に深刻な影響は見られなかった。ただし、95%信頼区間下限値(悲観シナリオ)での2021年第1四半期予測はコロナ禍での最悪水準(2020年第2四半期:379万人)を若干下回る水準となっているため警戒は必要である。予測に寄与する主要因は、日銀短観の生産設備DIから推測される需要の長期拡大と経営環境改善が考えられる。今後、景況感の悪化に伴い悲観シナリオ相当の失業問題が深刻化した際、これらの要因はEBPM過程の実行に有用であることが示唆される。

上記の各予測モデルの結果より、コロナ禍での倒産・失業問題の深刻さは倒産件数、完全失業率、飲食・宿泊サービス業の就業者数には表れていないことが明らかとなり、EBPM過程の第1段階・第2段階に資する予測モデルが開発できた。また、重回帰分析の予測の要因分解から倒産・失業問題の深刻さ回避を下支えしている要因を特定することができEBPM過程の第3段階・第4段階の実行を可能とする予測モデルともいえる。従って、本研究の目的でEBPM過程に資する倒産失業関連指標予測モデルの開発は達成できたといえる。

今後の課題として、まず第1に、賃金を対象とした同様の予測モデル開発を想定している。本研究では、倒産・失業問題の深刻さが倒産件数や失業率、就業者数には表れていないことが分かった。景況感が悪化した際、企業は、倒産や解雇を行う以前に従業員の賃金をカットすると想定される。従って、倒産や失業に表れていない倒産・失業問題への深刻さは従業員賃金に波及していると予想されるため、賃金を対象とした同様の分析が必要であると考えられる。

第2に、第2次緊急事態宣言が発出された現在において、第1次緊急事態宣言発出時との

類似性から宿泊・飲食サービス業就業者数予測を試みることを検討している。宿泊・飲食サービス業就業者数の予測サブモデルとして採用した3変数VARモデルにおいて第1次緊急事態宣言発出時の残差情報に基づきインパルス応答分析を試みる。

謝辞

主指導教員の梅津高朗准教授には、まず、担当教員を引き受けてくださったことに深く感謝いたします。研究のアドバイスはもちろんのこと、組合せ最適化書籍の輪読や就活時の成果物作成のための助言までくださり真摯に対応いただきました。ありがとうございました。

経済学部楠田浩二教授には、本研究に共同で取り組んでいただき非常に感謝しております。楠田先生に出会っていなければ、本修論の執筆は叶いませんでした。経済や統計解析の分野に明るくない私に初歩的な箇所から懇切丁寧に手取り足取り指導いただきました。楠田先生との本共同研究は毎日がとても刺激的で多くの学びや発見があり、私の人生の中でとても貴重な機会をいただいたと感じております。ありがとうございました。

データサイエンス学部笛田薫教授には、滋賀大学と(株)帝国データバンクによる共同研究センター(以下、DEMLセンター)での活動を通じた本研究への助言のみならず、大学院ゼミへの参加や課外活動の投資コンテストにて指導教員を引き受けてくださり大変お世話になりました。感謝いたします。

DEMLセンター長兼データサイエンス学部杉本知之教授には、DEMLセンターを通じた本研究結果のプレスリリース手配や修士研究の進捗に配慮いただき大変お世話になりました。感謝の意を表明します。

DEMLセンター主任研究員兼データサイエンス教育研究センター特任助教大里隆也氏には、DEMLセンター活動のコロナ対策プロジェクトに私を配属していただき本研究開始のキッカケをいただきました。また、DEMLセンターの活動だけでなく、(株)帝国データバンク内でのアルバイトの機会や進路相談に乗っていただき非常に2年間お世話になりました。感謝いたします。

(株)帝国データバンク社員の安江直芳氏、吉村鴻紀氏、岡部壮一郎氏、日野祐美氏には、本研究を進めるにあたりDEMLセンターのサーバデータの管理やTDBデータに関する様々な知見を共有いただき大変お世話になりました。特に、吉村鴻紀氏には(株)帝国データバンク大阪支社にてアルバイトの受け入れやTDB倒産件数データの収集に多大なる協力をいただきました。ここに感謝いたします。

最後に、DEMLセンターに所属する皆様並びに運営にご協力いただいた大学職員の皆様に深い感謝を申し上げます。

付録A 補論

本章では、予測モデルの構築に用いた統計解析手法の理論的背景についてまとめる。A.1節では、重回帰分析の基礎概念および推定量が想定している仮定(不均一分散や系列相関)およびその検定手法を述べる。A.2節では、最小二乗推定量の系列相関への対応として一般化最小二乗法による対処法を紹介する。A.3節で、影響点に対応した推定法として頑健回帰推定の概要を述べる。

A.1 重回帰分析

A.1.1 回帰分析の基礎概念

重回帰モデルの定式化

説明変数が k 個の重回帰モデル

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \quad (\text{A.1})$$

を考える。ここで、 Y_t は t 期の被説明変数、 X_{jt} は j 番目 t 期の説明変数、 u_t は t 期の誤差項。 $t = 1, \dots, n$ 期で式 (A.1) が成り立つならば、

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (\text{A.2})$$

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{2n} & \cdots & X_{kn} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

と定式化される [40]。重回帰分析では、式 (A.2) の未知パラメータ $\boldsymbol{\beta}$ を推定することで被説明変数と説明変数間の関係性を評価する。

重回帰モデルの諸仮定

式 (A.2) の重回帰モデルは、以下の諸仮定が成り立つものとする。

(1) $E(\mathbf{u}) = \mathbf{0}$

(2) $E(\mathbf{u}\mathbf{u}') = \sigma^2 \mathbf{I}$

\mathbf{I} は $n \times n$ の単位行列

(3) u_t は正規分布に従う

(4) \mathbf{X} は所与

$$\text{rank}(\mathbf{X}) = k < n$$

$$\lim_{n \rightarrow \infty} \frac{\mathbf{X}'\mathbf{X}}{n} = \mathbf{Q} \neq \mathbf{0}, \mathbf{Q} \text{ は非特異行列}$$

最小二乗法 (Ordinary least square)

β の推定量を $\hat{\beta}$ とする。この時、 $\hat{\beta}$ による被説明変数 y の推定値 \hat{y} は

$$\hat{y} = X\hat{\beta} \quad (\text{A.3})$$

と表現できる。 y と \hat{y} の差を e とおくと、 e の残差平方和は

$$e'e = (y - \hat{y})'(y - \hat{y}) = (y - X\hat{\beta})'(y - X\hat{\beta}) \quad (\text{A.4})$$

となる。式 (A.4) の残差平方和を最小とする $\hat{\beta}$ を求める推定法を最小二乗法 (Ordinary least square 以下, OLS) と呼ぶ。OLS による推定値は以下の方程式の解として求まる。

$$\frac{\partial e'e}{\partial \hat{\beta}} = 0 \quad (\text{A.5})$$

式 (A.5) を $\hat{\beta}$ について解くと

$$\hat{\beta} = (X'X)^{-1}X'y \quad (\text{A.6})$$

式 (A.6) のように $\hat{\beta}$ が求まる。この OLS による β の推定量 $\hat{\beta}$ を最小二乗推定量 (Ordinary least square's estimator 以下, OLSE) と呼ぶ。

最小二乗推定量の性質

$\hat{\beta}$ は、最良線形不偏推定量 (best linear unbiased estimator 以下, BLUE) である。BLUE とは、線形な不偏推定量の中で最も分散が小さい推定量である。重回帰モデルの諸仮定 (3) 以外が全て成り立つ場合、OLSE は BLUE となる。

さらに、重回帰モデルの諸仮定 (3) が成り立つ場合、OLSE は不偏推定量の中で最も分散が小さい推定量となる。この推定量を有効推定量と呼ぶ。

??~??では、OLSE の性質もしくは推定値そのものに影響を与えるいくつかの問題を取り挙げる。[38][39] [41] 【仮定と BP テストの概要について書く】

A.1.2 系列相関

同一の時系列データにおいて、異時点間の相関を系列相関 (serial correlation) と呼ぶ [42]。重回帰モデルの誤差項が系列相関を持つ場合、OLSE は BLUE ではなくなり、見せかけの回帰の危険が高まることが知られている。本小節では、重回帰分析における誤差項が一階の自己回帰過程 AR(1) に従うと仮定した場合の検定法を述べる。

1 階の自己回帰過程 AR(1)

重回帰分析における誤差項が AR(1) に従うとは

$$y = X\beta + u \quad (\text{A.7})$$

において、誤差項 u_t が

$$\begin{aligned} u_t &= \rho u_{t-1} + \epsilon_t \\ |\rho| &< 1 \\ \epsilon_t &\sim N(0, \sigma_\epsilon^2) \end{aligned} \quad (\text{A.8})$$

と 1 期前の u_t と確率的誤差項 ϵ_t により表現できることを指す。

誤差項が系列相関を持つ場合の OLSE

誤差項が AR(1) に従う時、通常の OLS を適用してパラメータ推定した場合の問題および特徴をまとめる。

(1) $E[\hat{\beta}] = \beta$ は成り立つ

誤差項が系列相関を持つ場合でも、OLSE は不偏性を満たす

(2) $\hat{\beta}$ は β の BLUE ではない

線形不偏推定量の中で OLSE は最小分散を満たさない

通常の OLSE の分散 $\sigma_{\hat{\beta}}^2$ は

$$\sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \quad (\text{A.9})$$

を想定しているが、系列相関を誤差項を持つ場合の分散 $\sigma_{\hat{\beta}}^{+2}$ は

$$\sigma_{\hat{\beta}}^{+2} = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} + \frac{\sum_{i \neq j} \sum (X_i - \bar{X})(X_j - \bar{X})\sigma_{ij}}{\{\sum (X_i - \bar{X})^2\}^2} \quad (\text{A.10})$$

式 (A.10) の右辺第 2 項が正の相関を持つ時、系列相関なし、均一分散を満たす OLSE の不偏推定量と比較して分散が大きくなるため BLUE ではない推定量が求まる問題が発生する。誤差項が AR(1) に従う場合に BLUE となる推定量を求める方法として A.2 節で一般化最小二乗法を挙げる。

ダービン・ワトソン検定

誤差項 u_t が AR(1) に従うと仮定した時、

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0 (\text{あるいは } \rho < 0 \text{ あるいは } \rho = 0)$$

のように帰無仮説と対立仮説を設定する検定をダービンワトソン検定 (以下、DW 検定) と呼ぶ。DW 検定には、ダービン・ワトソン比 (DW)

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (\text{A.11})$$

が用いられる。ここで e_t は OLS の残差を表す。

Durbin and Watson[23] は DW の分布が、以下の $a + b d_u$ で近似できることを示している。ここで d_u は DW の上限分布の臨界点を表す。従って、 a と b は

$$\begin{aligned} E(DW) &= a + bE(d_u) \\ \text{Var}(DW) &= b^2 \text{Var}(d_u) \end{aligned} \quad (\text{A.12})$$

によって決定される値である。すなわち

$$\begin{aligned} a &= E(DW) - \sqrt{\frac{\text{Var}(DW)}{\text{Var}(d_u)}} E(d_u) \\ b &= \sqrt{\frac{\text{Var}(DW)}{\text{Var}(d_u)}} \end{aligned} \quad (\text{A.13})$$

である。従って a と b を求め、

$$d^* = a + bd_u \quad (\text{A.14})$$

を計算して

$$DW < d^* \Rightarrow H_0: \rho = 0 \text{ 棄却}, \rho > 0 \quad (\text{A.15})$$

と判断するのが DW 検定の 1 つの方法である。

A.1.3 多重共線性

多重共線性とは説明変数間の高い相関のことである。多重共線性が発生している時に、最小 2 乗法で β を推定すると β の分散が大きくなり、推定精度は悪くなる。

ここで以下の重回帰モデルを例に考える。

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \quad (\text{A.16})$$

式 (A.16) のパラメータ β_j の OLSE を $\hat{\beta}_j$ とすれば、 $\hat{\beta}_j$ の分散は、

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_1^2} \left(\frac{1}{1 - r_{12}^2} \right) \quad (\text{A.17})$$

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_2^2} \left(\frac{1}{1 - r_{12}^2} \right) \quad (\text{A.18})$$

と表すことができる。ここで $\sum x_j^2 = \sum_{t=1}^n (X_{jt} - \bar{X}_j)^2$ 、 r_{12} は X_1 と X_2 の単相関係数。

X_1 と X_2 が完全に無相関であれば $r_{12} = 0$ であるから式 (A.17) は最小になる。一方、 X_1 と X_2 の相関が高くなるにしたがって r_{12}^2 は 1 に近づくため $\hat{\beta}_j$ の分散は大きくなる。

本小節では、多重共線性の尺度として分散拡大要因を説明し対処法として、リッジ回帰を挙げる。

分散拡大要因 (VIF)

式 (A.17) において、分散が大きくなる要因

$$\frac{1}{1 - r_{12}^2} \quad (\text{A.19})$$

は分散拡大要因 Variance inflation factor (VIF) と呼ばれ

$$\text{VIF}_1 = \text{VIF}_2 = \frac{1}{1 - r_{12}^2} \quad (\text{A.20})$$

のように表される。また、 $1 - r_{12}^2$ は許容度 tolerance と呼ばれている。

一般に、重回帰モデルを以下のように表す。

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i \quad (\text{A.21})$$

このとき、説明変数の $k \times k$ の単相関行列を R とすると、 VIF_j は R^{-1} の (j, j) 要素で与えられるから

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \quad (\text{A.22})$$

と表現できる。ここで R_j^2 は X_j の $X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_k$ への回帰を取った時の決定係数である。

$R_j^2 \geq 0.9$ のとき、 $\text{VIF}_j \geq 10$ となるため、 VIF_j が 10 以上のとき多重共線性によって $\hat{\beta}_j$ の分散が大きくなり、推定精度が損なわれる恐れが十分あるとされている。

リッジ回帰

多重共線性の対処法の1つとして、リッジ回帰を挙げる。但し蓑谷 [38] で述べられるようにリッジ回帰は多重共線性の問題を必ず解決するものではないことを強調しておく。

リッジ回帰による推定量 (リッジ推定量) $\hat{\beta}_R$ は、次式で与えられる。

$$\hat{\beta}_R = (\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \quad (\text{A.23})$$

ここで、 $c > 0$ は任意定数、 \mathbf{I} は単位行列である。リッジ推定量の期待値は

$$E[\hat{\beta}_R] = (\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\beta \neq \beta \quad (\text{A.24})$$

であるから、リッジ推定量は不偏性を満たさない。すなわち、リッジ推定量は偏り (bias) を持ち、そのバイアス ($B(\hat{\beta}_R)$) は

$$B(\hat{\beta}_R) = -c(\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-1} \quad (\text{A.25})$$

となる。

このように $\hat{\beta}_R$ は β の不偏推定量ではないが、 $\text{Var}(\beta) - \text{Var}(\hat{\beta}_R)$ は任意の $c > 0$ に対し正值定符号であることが知られている。

つまり、 $\hat{\beta}_R$ の要素 $\hat{\beta}_{Rj}$ でいえば、 $\hat{\beta}_{Rj}$ は偏りがあるが、 $\text{Var}(\hat{\beta}_{Rj}) < \text{Var}(\hat{\beta}_j)$ である。従って、 $\hat{\beta}_{Rj}$ の偏りの2乗が $\text{Var}(\hat{\beta}_{Rj}) < \text{Var}(\hat{\beta}_j)$ より小さければ $\text{MSE}(\hat{\beta}_{Rj}) < \text{MSE}(\hat{\beta}_j)$ となる。

つまり、リッジ回帰は不偏性を満たさないが、MSE の観点で OLSE よりも良い推定量を求めることができる方法である。

A.1.4 影響点

影響点とは、OLS の推定値に影響を与える観測値のことを指す。より具体的には、被説明変数の外れ値と説明変数の高い作用点をまとめて影響点と呼ぶ。ここでは、影響点の説明とその検出方法として Cook's Distance と L-R プロットを述べる。

影響点を発見した場合の対処法については、A.3 で詳しく述べる。

影響関数

影響関数 (influence function) とは、推定量が観測値に対してどのように反応するかを調べる概念である。

z_1, z_2, \dots, z_n を分布関数 F からの無作為標本とし、あるパラメータ $T(F)$ の推定量を

$$T_n = T_n(z_1, z_2, \dots, z_n) \quad (\text{A.26})$$

とする。そして F_n を n 個の z_1, z_2, \dots, z_n にもとづく経験的分布関数とする。また、 T_n を

$$T_n = T_n(z_1, z_2, \dots, z_n) = T(F_n) \quad (\text{A.27})$$

と表し

$$T(F_n) \xrightarrow{p} T(F) \quad (\text{A.28})$$

推定量 $T(F_n)$ は確率収束すると定義する。この時、極めて大きな無作為標本に1個の観測点 z が追加される場合、 T がどのような影響を受けるかは以下の影響曲線 (influence curve) IC によって示される。

$$IC(z; F, T) = \lim_{\epsilon \rightarrow 0} \frac{T[(1 - \epsilon)F + \epsilon\delta_z] - T(F)}{\epsilon} \quad (\text{A.29})$$

ここで δ_z は z で 1 をとり、それ以外では 0 の値をとる分布関数である。分布関数 $F_\epsilon = (1 - \epsilon)F + \epsilon\delta_z$ は、 ϵ が十分小さく 0 に近い時ほとんど F に等しいが、点 z で追加的なウェイト $\epsilon\delta_z$ を持つ。

影響関数の例：期待値

影響関数の具体例を期待値を取り上げ説明する。母平均 μ 、標本平均 \bar{z} は、

$$T(F) = \int z dF(z) = E(Z) = \mu \quad (\text{A.30})$$

$$T(F_n) = \int z dF_n(z) = \bar{z} \quad (\text{A.31})$$

と表現できる。 $dF_\epsilon = (1 - \epsilon)dF + \epsilon d\delta_z$ に注意すると

$$\begin{aligned} T[(1 - \epsilon)F + \epsilon\delta_z] &= \int (1 - \epsilon)z dF + \int \epsilon z d\delta_z \\ &= (1 - \epsilon)\mu + \epsilon z \end{aligned} \quad (\text{A.32})$$

と式展開される。従って、式 (A.29) に式 (A.32) の結果を代入すると期待値の影響関数は

$$IC(z; F, T) = \lim_{\epsilon \rightarrow 0} \frac{(1 - \epsilon)\mu + \epsilon z - \mu}{\epsilon} = z - \mu \quad (\text{A.33})$$

となる。式 (A.33) より、 z が μ から離れれば離れるほど z の影響は大きくなり、限界のない変化を期待値の推定量に与える。

OLSE の影響関数

同様に線形回帰モデルにおける β の OLSE $\hat{\beta}$ の影響関数を考える。 \mathbf{x}' を $1 \times k$ のベクトル、 Y を 1×1 のスカラーとし、分布関数 F のもとで

$$E_F(\mathbf{x}\mathbf{x}') = \Sigma_{XX}(F) \quad (\text{A.34})$$

$$E_F(\mathbf{x}Y) = \Sigma_{XY}(F) \quad (\text{A.35})$$

とする。 $T(F_n) = \hat{\beta}(F_n)$ とすれば

$$T(F) = \beta(F) = \Sigma_{XX}^{-1}(F)\Sigma_{XY}(F) \quad (\text{A.36})$$

で与えられるから、 β の影響関数は次式となる。

$$IC[\mathbf{x}', Y; \beta(F)] = \{Y - \mathbf{x}'\beta(F)\}\Sigma_{XX}^{-1}(F)\mathbf{x} \quad (\text{A.37})$$

式 (A.37) の右辺 $\{Y - \mathbf{x}'\beta(F)\}$ 部分は、 Y 方向の外れ値による OLSE への影響は限界がないことを表している。また、右辺 $\Sigma_{XX}^{-1}(F)\mathbf{x}$ 部分は、 X 方向の高い作用点による OLSE への影響は限界がないことを表している。

従って、OLSE は X 方向、 Y 方向両方の影響点に限界なく影響を受けるため、揺らぎの少ない推定値を得るには影響点の発見とその対処が必要となる。

クックの距離

クックの距離 (Cook's Distance) とは, i 番目の観測値の $\hat{\beta}$ への影響力を示す指標であり, 以下のように表される.

$$\begin{aligned} D_i &= \frac{[\hat{\beta} - \hat{\beta}(i)]'(\mathbf{X}'\mathbf{X})[\hat{\beta} - \hat{\beta}(i)]/(k+1)}{\sum e^2/(n-k-1)} \\ &= \frac{[\hat{\beta} - \hat{\beta}(i)]'(\mathbf{X}'\mathbf{X})[\hat{\beta} - \hat{\beta}(i)]}{(k+1)s^2} \end{aligned} \quad (\text{A.38})$$

式 (A.38) にて, $\hat{\beta}(i)$ は i 番目の観測値を抜いて推定した OLSE, n は観測数, k は定数項以外の説明変数の数, s^2 は誤差分散の不偏推定量を表す. X 方向の誤差 (高い作用点) と Y 方向の誤差 (外れ値) それぞれが大きいほどクックの距離は, 大きくなる特徴がある. そのため, 影響点の度合いを測る尺度として使用できる.

蓑谷 [38] によると, 以下の切断点を用いてクックの距離による影響点を検出するのが良いと述べられている. 本研究においても, 以下の切断点を基準に影響点検出を行う.

$$D_i > \frac{F_{0.05}(1, n-k-2)}{(n-k-2 + F_{0.05}(1, n-k-2))} \quad (\text{A.39})$$

ここで, $F_{0.05}(1, n-k-2)$ は, 自由度 $(1, n-k-2)$ に従う F 分布の有意水準 5% 臨界値である.

L-R プロット

説明変数空間において, x_i が中心からどれだけ離れているかは

$$h_{ii} = x_i'(\mathbf{X}'\mathbf{X})^{-1}x_i \quad (\text{A.40})$$

によって表される. さらに

$$a_{ii}^2 = \frac{e_i^2}{\sum_{i=1}^n e_i^2} \quad (\text{A.41})$$

とすると, X 方向の高い作用点を h_{ii} , Y 方向の外れ値を a_{ii}^2 により表すことができる. 横軸に a_{ii}^2 , 縦軸に h_{ii} をプロットしたものを L-R プロットと呼ぶ. L-R プロットにより, 影響点か外れ値なのか高い作用点なのかあるいは両方かを視覚的に判断ができる.

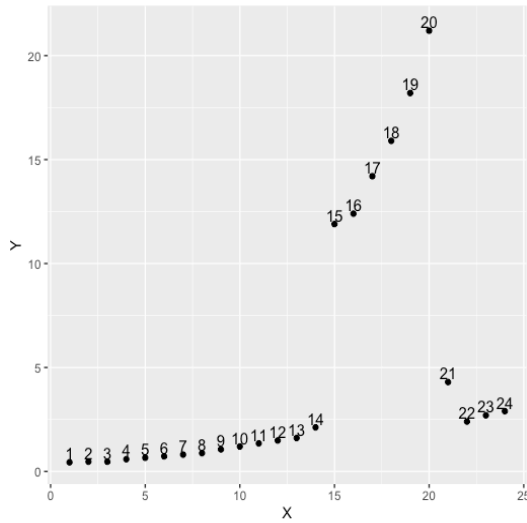


図 A.1: サンプルデータの散布図

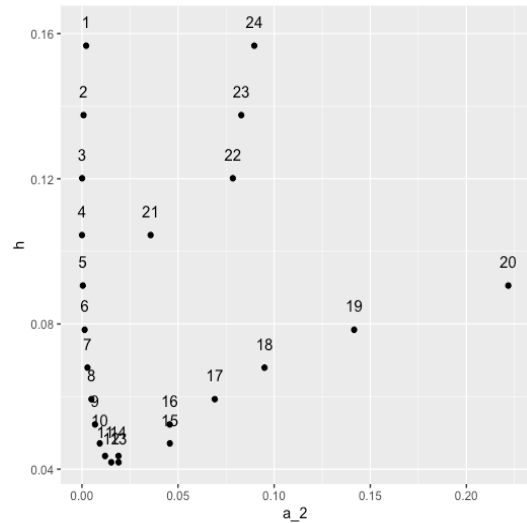


図 A.2: L-R プロットの例

図 A.1 は、24 の観測値を持つサンプルデータの散布図である．このデータを元に L-R プロットを描画したものが図 A.2 である．上記の例より、20 番目の観測値が L-R プロットの横軸に離れた場所に描画されており外れ値であることが見て取れる．また、観測値 1 や 24 は縦軸に離れた場所に描画されており高い作用点を示している．

A.2 一般化最小二乗法 (GLS)

本節では、系列相関への対処法として Prais-Winsten 変換 (PW 変換) による一般化最小二乗法を述べる．この手法を用いることで重回帰モデルの誤差項が AR(1) に従う場合でも BLUE となる推定量を求められる．

A.2.1 GLS によるパラメータ推定

Ω を正定値対称行列とすると、

$$\Omega^{-1} = P'P \quad (\text{A.42})$$

と非特異行列 P に分解できる．この P を重回帰モデル

$$y = X\beta + u \quad (\text{A.43})$$

の両辺に左から掛けると

$$Py = PX\beta + Pu \quad (\text{A.44})$$

となる．式 (A.44) の誤差項の期待値と分散を求めると

$$\begin{aligned} E(Pu) &= 0 \\ \text{Var}(Pu) &= \sigma^2 I \end{aligned} \quad (\text{A.45})$$

式 (A.45) の結果より、 Pu は期待値 0、自己相関なし、均一分散という線形回帰モデルの誤差項の仮定を満たしている．従って、 Py を被説明変数、 PX を説明変数として通常の OLS を適用して得られる β の推定量を $\tilde{\beta}$ とすれば、 $\tilde{\beta}$ は BLUE となる．

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \quad (\text{A.46})$$

は一般化最小二乗推定量 (GLSE) と呼ばれる.

u が AR(1) の時, 共分散行列は

$$E(\mathbf{u}\mathbf{u}') = \sigma_u^2 \mathbf{\Omega} = \frac{\sigma_\epsilon^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} \quad (\text{A.47})$$

である. この時,

$$\mathbf{\Omega}^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & \cdots & 0 & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix} \quad (\text{A.48})$$

となり, \mathbf{P} を

$$\mathbf{P} = \begin{bmatrix} \sqrt{1 - \rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix} \quad (\text{A.49})$$

とすれば,

$$\mathbf{P}'\mathbf{P} = (1 - \rho^2)\mathbf{\Omega}^{-1} \quad (\text{A.50})$$

の関係がある. 従って, β の GLSE は

$$\tilde{\beta} = [(\mathbf{P}\mathbf{X})'\mathbf{P}\mathbf{X}]^{-1}(\mathbf{P}\mathbf{X})'\mathbf{P}\mathbf{y} \quad (\text{A.51})$$

と, $\mathbf{P}\mathbf{y}$ の $\mathbf{P}\mathbf{X}$ への回帰を行うことによって求めることができる. この \mathbf{P} を用いる \mathbf{y} から $\mathbf{P}\mathbf{y}$, \mathbf{X} から $\mathbf{P}\mathbf{X}$ への変換は, Prais-Winsten 変換 (PW 変換) と呼ばれる.

A.2.2 実行可能な GLS

$\mathbf{\Omega}$ が未知の時には, $\mathbf{\Omega}$ の推定量 $\hat{\mathbf{\Omega}}$ を用いて GLSE を求める. この時

$$\tilde{\beta}_* = (\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{y} \quad (\text{A.52})$$

は実行可能な一般化最小二乗推定量 (FGLSE) と呼ばれる. FGLSE の代表的な方法は 2SPW (2 Step Prais-Winsten) とである. 2SPW は, 以下のステップで OLS が 2 度適用されて ρ を推定する.

第 1 ステップ: ρ の推定

通常の OLS を適用し, 残差 e から ρ を推定. ρ の推定方法として代表的なものは以下の通りがある.

(i) 最小 2 乗原理

誤差項 u_t が AR(1) に従っている時、残差にも同様の

$$e_t = \hat{\rho}e_{t-1} + \hat{\epsilon}_t \quad (\text{A.53})$$

という関係ががあると考えられる。この式の $\hat{\rho}$ を最小 2 乗法で推定する。

(ii) Yule-Walker 方程式

u_t が AR(1) に従っている時、Yule-Walker 方程式は

$$\gamma_1 = \rho\gamma_0 \quad (\text{A.54})$$

である。ここで、

$$\gamma_j = E(u_t u_{t-j}) \quad (\text{A.55})$$

γ_j の推定量は

$$\hat{\gamma}_j = \frac{1}{n} \sum_{t=j+1}^n u_t u_{t-j}, j = 0, 1 \quad (\text{A.56})$$

によって与えられるので

$$\hat{\rho} = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \quad (\text{A.57})$$

で推定する。

(iii) ダービン・ワトソン統計量

n が十分大きければ

$$DW \simeq 2(1 - \hat{\rho}) \quad (\text{A.58})$$

の関係より

$$\hat{\rho} = 1 - \frac{DW}{2} \quad (\text{A.59})$$

で推定する。

第 2 ステップ：PW 変換による GLS

第 1 ステップで求めた ρ の推定値 $\hat{\rho}$ を用いて \mathbf{y} , \mathbf{X} を PW 変換して定数項なしの線形回帰を行う。この結果、得られる β の推定量は実行可能な一般化最小二乗推定量 (Feasible GLSE) と呼ばれる。

さらに、2SPW の Feasible GLSE による残差から新たな ρ の推定値を計算し、 ρ が収束するまで繰り返す方法を繰り返し Prais-Winsten (ITERPW) と呼ぶ。

A.3 頑健回帰推定

A.1.4 小節の OLSE の影響関数から、重回帰モデルのパラメータ β の推定値は影響点から限界ない影響を受ける。そこで、影響点を考慮した β の推定法として頑健回帰推定が存在する。本節では、頑健回帰推定法として M 推定および MM 推定を紹介する。

A.3.1 M 推定

OLSE における M 推定

重回帰モデルを

$$Y_i = \mathbf{x}'_i \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n \quad (\text{A.60})$$

とする。ここで $1 \times k$ ベクトル

$$\mathbf{x}'_i = \left(1 \quad X_{2i} \quad \dots \quad X_{ki} \right) \quad (\text{A.61})$$

は所与、確率誤差項 ϵ_i は $E[\epsilon_i] = 0$ と仮定する。ここで $\boldsymbol{\beta}$ の M 推定量 $\hat{\boldsymbol{\beta}}_M$ は

$$\hat{\boldsymbol{\beta}}_M = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho(Y_i - \mathbf{x}'_i \boldsymbol{\beta}) \quad (\text{A.62})$$

の解と表現される。OLSE $\hat{\boldsymbol{\beta}}$ は

$$\rho(u) = \frac{u^2}{2} \quad (\text{A.63})$$

の M 推定量である。すなわち、

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho(Y_i - \mathbf{x}'_i \boldsymbol{\beta}) = \arg \min_{\boldsymbol{\beta}} \frac{1}{2} \sum_{i=1}^n (Y_i - \mathbf{x}'_i \boldsymbol{\beta})^2 \quad (\text{A.64})$$

が OLS の最小にすべき損失関数である。

$\rho(u)$ は微分可能で、0 まわりで対照的な凸関数の時、 $\boldsymbol{\beta}$ の M 推定量を $\hat{\boldsymbol{\beta}}_M$ は

$$\sum_{i=1}^n \rho'(Y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}_M) \mathbf{x}_i = \sum_{i=1}^n \Psi(e_i) \mathbf{x}_i = \mathbf{0} \quad (\text{A.65})$$

の解として得られる。ここで $\Psi(e_i) = \rho'(e_i)$, $e_i = Y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}_M$ である。

OLSE の場合

$$\Psi(e_i) = \rho'(e_i) = e_i \quad (\text{A.66})$$

であるから、OLSE を求める式は

$$\sum_{i=1}^n e_i \mathbf{x}_i = \mathbf{0} \quad (\text{A.67})$$

となる。ウェイト関数 $w(u)$ を

$$w(u) = \frac{\Psi(u)}{u} \quad (\text{A.68})$$

と定義すれば、式 (A.65) は

$$\sum_{i=1}^n w_i(e_i) e_i \mathbf{x}_i = \sum_{i=1}^n w_i e_i \mathbf{x}_i = \mathbf{0} \quad (\text{A.69})$$

と表すことができる。 $w_i = w_i(e_i)$ であり、ウェイトは残差 e_i に依存する。

加重最小二乗推定量

式 (A.69) は、加重最小二乗推定量 (weighted least-squares estimator) WLSE の解を与える。ここで WLSE $\hat{\beta}_W$ は

$$\hat{\beta}_W = \arg \min_{\beta} \sum_{i=1}^n w_i e_i^2 = \arg \min_{\beta} \sum_{i=1}^n w_i (Y_i - \mathbf{x}'_i \hat{\beta}_W)^2 \quad (\text{A.70})$$

と定義される。従って、

$$\frac{d \sum w_i e_i^2}{d \hat{\beta}_W} = \sum_{i=1}^n w_i e_i \mathbf{x}_i = 0 \quad (\text{A.71})$$

の解でなければならない。式 (A.69) は

$$\sum_{i=1}^n \mathbf{x}_i w_i \mathbf{x}'_i \hat{\beta}_M = \sum_{i=1}^n \mathbf{x}_i w_i Y_i \quad (\text{A.72})$$

と表現できる。行列で表すと、式 (A.72) は次のような加重最小二乗法の正規方程式を与える。

$$\mathbf{X}' \mathbf{W} \mathbf{X} \hat{\beta}_M = \mathbf{X}' \mathbf{W} \mathbf{y} \quad (\text{A.73})$$

ここで、 \mathbf{X} は $n \times k$ 、 \mathbf{y} は $n \times 1$ 、 $\hat{\beta}_M$ は $k \times 1$ 、 \mathbf{W} は

$$\mathbf{W} = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix} = \text{diag}\{w_i\} \quad (\text{A.74})$$

によって与えられる $n \times n$ の対角行列である。式 (A.73) より

$$\hat{\beta}_M = (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} \mathbf{y} \quad (\text{A.75})$$

が得られる。すなわち、 Ψ 関数を用いる β の M 推定量はウェイト w_i が式 (A.68) によって与えられた WLSE として求めることができる。

OLS の場合、 $\Psi(e_i) = e_i$ であるから

$$w_i = \frac{\Psi(e_i)}{e_i} = 1, \quad i = 1, \dots, n \quad (\text{A.76})$$

である。すなわち加重回帰の観点から見ると、OLS はすべての残差に等ウェイト 1 を与える。

上記より、頑健回帰推定は、加重回帰で絶対値の大きな残差に対してはウェイトを小さくする推定法と解釈できる。

しかし、残差 e_i の大きさは被説明変数の単位に依存する。そのため、 e_i の水準によって外れ値かどうか判断することはできない。そこで、 e_i を誤差項の標準偏差 σ で割り、標準化された残差 e_i/σ を考えると、 σ は定数であるから、式 (A.65) は

$$\sum_{i=1}^n \Psi\left(\frac{e_i}{\sigma}\right) \mathbf{x}_i = 0 \quad (\text{A.77})$$

と同等である。 $u_i = e_i/\sigma$ とおけば、式 (A.77) は

$$\sum_{i=1}^n \Psi(u_i) \mathbf{x}_i = 0 \quad (\text{A.78})$$

に等しく、式 (A.84) は

$$w_i = \frac{\Psi(u_i)}{u_i} \quad (\text{A.79})$$

とすれば

$$\sum_{i=1}^n w_i u_i \mathbf{x}_i = 0 \quad (\text{A.80})$$

と同等である.

Huber の Ψ

前小節で、M 推定は Ψ 関数によるウェイトを用いた加重最小二乗推定で求まることを示した. ここで、 Ψ 関数の例として Huber の Ψ 関数を取り上げる. Huber の Ψ 関数は次式で与えられる.

$$\Psi(u_i) = \begin{cases} u_i & |u_i| \leq H \\ H & u_i > H \\ -H & u_i < -H \end{cases} \quad (\text{A.81})$$

Huber の Ψ 関数による損失関数 ρ は

$$\rho(u_i) = \begin{cases} \frac{u_i^2}{2} & |u_i| \leq H \\ H|u_i| - \frac{H^2}{2} & |u_i| > H \end{cases} \quad (\text{A.82})$$

Huber の Ψ 関数によるウェイトでは

$$w_i = \begin{cases} 1 & |u_i| \leq H \\ \frac{H}{|u_i|} & |u_i| > H \end{cases} \quad (\text{A.83})$$

Huber の $\Psi(u_i)$ 関数と、OLS の $\Psi(u_i) = u_i$ 関数をグラフで示す.

図 A.3 は x 軸に標準化残差, y 軸に Ψ 関数を示す. 図 A.4 は x 軸に標準化残差, y 軸にウェイトを示す. この例では $H = 1.345$ と設定.

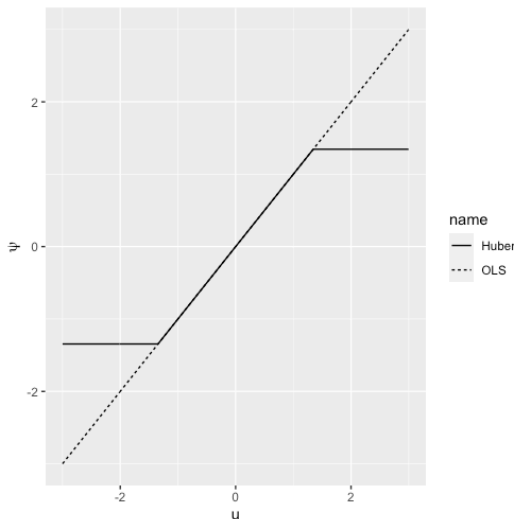


図 A.3: Ψ 関数のプロット例

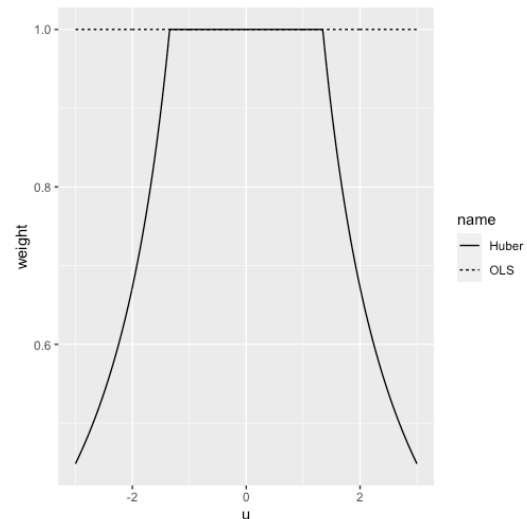


図 A.4: ウェイトのプロット例

図 A.4 より、Huber の Ψ 関数によるウェイトは標準化残差がある一定値 (この例では、1.345) を超えた場合、減衰していくことがわかる. 従って、残差の外れ値となるデータに対して小さなウェイトを与えて加重最小二乗推定を行える.

M 推定量の影響関数

M 推定量の影響関数は、式 (A.84) となることが Hampel et al.[25] により与えられている。

$$IC[\mathbf{x}', Y; \beta_M(F)] = \Psi[Y - \mathbf{x}'\beta_M(F)]\mathbf{B}^{-1}\mathbf{x} \quad (\text{A.84})$$

ここで、

$$\mathbf{B} = \int \Psi'[Y - \mathbf{x}'\beta_M(F)]\mathbf{x}\mathbf{x}'dF(\mathbf{x}, Y) \quad (\text{A.85})$$

式 (A.84) より、M 推定量の影響関数は β の OLS 推定量 $\hat{\beta}(F)$ の影響関数と比較して次の特徴をもつことがわかる。

1. $\hat{\beta}(F)$ は残差 $Y - \mathbf{x}'\hat{\beta}(F)$ に対して限界をもたないが、 $\hat{\beta}_M(F)$ は残差に対して限界をもつ。
2. OLS のとき $\Psi'(u) = 1$ となり、 $\mathbf{B} = \Sigma_{XX}(F)$ となるが M 推定においては $\Psi'(u) = 0$ となる Ψ 関数がほとんどである。
3. もし \mathbf{x} に限界がなければ、 $\mathbf{B}^{-1}\mathbf{x}$ は限界を持たず、 Ψ 関数が大きな残差に限界を与えても $\hat{\beta}_M$ の IC は限界をもたない。つまり X 方向の誤差からの影響に M 推定量は頑健でない。

上記の特徴の内、3 番目の特徴は M 推定量が X 方向の影響点 (高い作用点) に対しては頑健でないことを表す。この対策として、A.3.6 小節で MM 推定を挙げる。

A.3.2 M 推定量の不偏性と漸近的特性

本小節では、M 推定量が持つ性質として不偏性と漸近的特性を述べる。まず、 Ψ 関数が奇関数 ($\Psi(-u) = -\Psi(u)$) かつ分布 F が中心 T の周りで対称ならば

$$\sum_{i=1}^n \Psi\left(\frac{X_i - T_n}{\sigma}\right) = 0 \quad (\text{A.86})$$

の解である M 推定量 T_n は T の不偏推定量である [26]。例として、Huber の Ψ 関数を用いる M 推定量は不偏性をもつ。真の分布が F であるときのパラメータ $T(F)$ の M 推定量 T_n は

$$\sqrt{n}[T_n - T(F)] \xrightarrow{d} N[0, V(T, F)] \quad (\text{A.87})$$

$$V(T, F) = \int IC(\mathbf{x}; T, F)^2 dF(\mathbf{x}) \quad (\text{A.88})$$

と漸近的に正規分布に従い、その分散分散 $V(T, F)$ は影響関数を用いて式 (A.88) のように表される [25][27]。回帰モデルの M 推定量 $\hat{\beta}_M$ も漸近的に正規分布する。

$$\sqrt{n}[\hat{\beta}_M - \beta] \xrightarrow{d} N[0, v\Sigma_{\mathbf{X}\mathbf{X}}^{-1}] \quad (\text{A.89})$$

$$v = \sigma^2 \frac{E[\Psi^2(u)]}{\{E[\Psi'(u)]\}^2}$$

$$\Sigma_{\mathbf{X}\mathbf{X}} = E(\mathbf{x}\mathbf{x}')$$

である [31][36]。従って、 n が十分大きい時、近似的に

$$\hat{\beta}_M \xrightarrow{d} N[\beta, \hat{v}(\mathbf{X}'\mathbf{X})^{-1}] \quad (\text{A.90})$$

が成り立つ。

e_i を基準化して $\hat{u}_i = e_i/\hat{\sigma}$ を求める場合や \hat{v} を求める際、局外パラメータ $\hat{\sigma}$ が頑健回帰推定の結果に大きく影響する。従って、 σ の推定量 $\hat{\sigma}$ をいかにして求めるかが頑健推定の大きな問題である。これについては A.3.5 小節で述べる。

例として、Huber の Ψ 関数による M 推定量の漸近的分散は式 (A.91) となる。

$$V(T, F) = \frac{E(\Psi^2)}{[E(\Psi')]^2} = \frac{-2H\phi(H) + 1 - 2\Phi(-H) + 2H^2\Phi(-H)}{[1 - 2\Phi(-H)]^2} \quad (\text{A.91})$$

標準正規分布の分散は 1 であるから、真の確率分布が正規分布の時、Huber の Ψ 関数による M 推定量 T_n の漸近的有効性 $AE(T_n)$ は式 (A.92) で与えられる。

$$AE(T_n) = \frac{[1 - 2\Phi(-H)]^2}{-2H\phi(H) + 1 - 2\Phi(-H) + 2H^2\Phi(-H)} \quad (\text{A.92})$$

ここで、Huber の Ψ 関数における $H=1.345$ は、 $AE(T_n) = 0.95$ (95% の漸近的有効性) を与える。

A.3.3 崩壊点

推定量の頑健性を測る際に用いられる指標として、崩壊点 BDP(breakdown point) がある。 n 個の標本点を

$$\mathbf{Z} = \{(\mathbf{x}'_1, Y_1), \dots, (\mathbf{x}'_n, Y_n)\} \quad (\text{A.93})$$

とし、 \mathbf{Z} から得られる推定量 β を $\hat{\beta}(\mathbf{Z})$ とする。

この時、 n 個の観測点の内、 m 個 ($1 \leq m \leq n$) の観測点を任意の値に取り替えることで得られる標本を \mathbf{Z}' とする。この m 個の汚染 (contamination) によって推定量がどれくらい変化するかは

$$\|\hat{\beta}(\mathbf{Z}') - \hat{\beta}(\mathbf{Z})\| \quad (\text{A.94})$$

によって表せる。

この外れ値 (あるいは高い作用点) により生じる最大の大きさを $bias(m; \hat{\beta}(\mathbf{Z}))$ と書くと

$$bias(m; \hat{\beta}(\mathbf{Z})) = \sup_{\mathbf{Z}'} \|\hat{\beta}(\mathbf{Z}') - \hat{\beta}(\mathbf{Z})\| \quad (\text{A.95})$$

となる。上限 (supremum) は全ての可能な \mathbf{Z}' に対するものである。この $bias(m; \hat{\beta}(\mathbf{Z}))$ が無限の大きさになる場合、 m 個の外れ値によって $\hat{\beta}$ は推定値として無意味な値へと変化する。つまり、推定値は崩壊 (breakdown) する。

従って、標本 \mathbf{Z} における推定量 $\hat{\beta}$ の有限標本崩壊点は

$$\epsilon_n^*(\hat{\beta}(\mathbf{Z})) = \min\left\{\frac{m}{n}; bias(m; \hat{\beta}(\mathbf{Z})) \rightarrow \infty\right\} \quad (\text{A.96})$$

と定義される。つまり、推定値をどのような値にもすることができる影響点の最小の割合 m/n を崩壊点と呼ぶ。

OLS は 1 個の外れ値によって推定値を無意味な値とするため崩壊点は $1/n$ である。 $n \rightarrow \infty$ のとき 0 となる。このことを OLS の漸近的崩壊点は 0% であるという。

期待し得る崩壊点の最前の値は 50% とされている。なぜならば、崩壊点 50% というのはデータの外れ値 (もしくは作用点) 部分とその他の部分を区別不可能にする比率を意味するからである。

A.3.4 崩壊点と調整定数

崩壊点が何%になるかは、損失関数 ρ と調整定数の値に依存する。 ρ が2つの条件

(R1) ρ は対称、連続微分可能であり $\rho(0) = 0$ である。

(R2) ρ は $[0, c]$ で単調増加、 $[c, \infty]$ で一定となる $c > 0$ が存在する。

を満たし、正規分布のもとでの ρ の期待値を $E_{\Phi}(\rho)$ とすると

$$\frac{E_{\Phi}(\rho)}{\rho(c)} = \lambda \quad (\text{A.97})$$

を満たすように調整定数 c を選ぶことで、漸近的崩壊点を $100 \times \lambda\%$ とすることができる [33].

A.3.5 σ の推定

頑健回帰推定において、 σ の推定値によって残差は基準化され、この基準化残差と Ψ 関数によりウェイトが決まり、ウェイトを用いた加重最小2乗法によって回帰係数が推定される。

従って、頑健回帰推定において σ をいかにして推定するかは極めて重要である。 σ を OLS の残差 e_i を用いて

$$s = \left(\frac{\sum e^2}{n - k} \right)^{\frac{1}{2}} \quad (\text{A.98})$$

により推定する場合、追加された観測点 z からの s^2 の無限標本における影響関数は

$$IC(z; F, T) = (z - \mu)^2 - \sigma \quad (\text{A.99})$$

となり、 s^2 は z から限界のない大きな影響を受ける推定量であることが分かる。従って、 s 自身が外れ値から大きな影響を受けるため、この s を使用して基準化するのは望ましくない。

σ の M 推定

z_1, \dots, z_n を $\text{cdf}F(z)$ からの無作為標本、位置 (location) パラメータ $T(F)$ の推定量を T_n 、尺度 (scale) パラメータ $s(F)$ の推定量を s_n とする。 T_n および s_n が次の2本の方程式を満たす時、 T_n, s_n は同時 M 推定量と呼ばれる。

$$\begin{aligned} \sum_{i=1}^n \Psi \left(\frac{z_i - T_n}{cs_n} \right) \\ \sum_{i=1}^n \chi \left(\frac{z_i - T_n}{cs_n} \right) \end{aligned} \quad (\text{A.100})$$

この時、

$$\sqrt{n}[T_n - T(F)] \xrightarrow{d} N(0, A(F, T)) \quad (\text{A.101})$$

が成り立つ。式 (A.101) の漸近的分散 $A(F, T)$ は次式で与えられる [26].

$$A(F, T) = \frac{c^2 s^2(F) E\{\Psi^2[\frac{z-T(F)}{cs(F)}]\}}{E\{\Psi'[\frac{z-T(F)}{cs(F)}]\}^2} \quad (\text{A.102})$$

この漸近的分散を回帰モデルの $\epsilon \sim iid(0, \sigma^2)$ に適用する. $T(F)$ を $T_n = M = \text{median}(e_i)$ で推定, $s(F)$ を $s_n = MAD$ で推定し

$$u_i = \frac{e_i - M}{cMAD} \quad (\text{A.103})$$

とおく. ここで, $MAD = \text{median}|e_i - M|$ である. そして,

$$E[\Psi'(u)] = \frac{1}{n} \sum_{i=1}^n \Psi'(u_i) \quad (\text{A.104})$$

$$E[\Psi^2(u)] = \frac{1}{n} \sum_{i=1}^n \Psi^2(u_i) \quad (\text{A.105})$$

を用いると, 式 (A.102) の推定量は

$$s_T^2 = \frac{n(cMAD)^2 \sum_{i=1}^n \Psi^2(u_i)}{[\sum_{i=1}^n \Psi'(u_i)]^2} \quad (\text{A.106})$$

従って, σ の M 推定量は

$$s_T = \frac{\sqrt{n}(cMAD)[\sum_{i=1}^n \Psi^2(u_i)]^{\frac{1}{2}}}{|\sum_{i=1}^n \Psi'(u_i)|} \quad (\text{A.107})$$

となる.

A.3.6 MM 推定

M 推定は, 影響関数から分かるように Y 方向 (外れ値) に頑健だが, X 方向 (高い作用点) には頑健でないことを A.3.1 小節で述べた. X 方向にも頑健な推定方法として, 有界影響推定が提案されている. しかし, 有界影響推定の問題点として漸近的有効性が低い点が挙げられる. そこで, 高い BDP と同時に, 誤差項が正規分布するとき高い漸近的有効性をもつ推定法として MM 推定量を説明する.

MM 推定では, 損失関数 ρ が A.3.4 小節で示した条件 (R1) と (R2) を満たしていることが仮定される. MM 推定量の推定アルゴリズムは次の 3 段階からなる.

第 1 段階: 残差の頑健推定

LMS や LTS 等の S 推定と呼ばれる有界影響推定法で BDP50% を担保する調整定数を用いて推定を行い残差 e_i を求める.

第 2 段階: 誤差項の M 推定

第 1 段階の残差 e_i を用いて, 前節で述べた σ の M 推定法を用いて σ の推定値 $\hat{\sigma}_M$ を求める. $\rho^* = 100,000$ とする.

第3段階：回帰パラメータの頑健推定

第2段階で得られた $\hat{\sigma}_M$ の値を固定し、漸近的有効性 95% となる Ψ 関数の調整定数を用いて $\min \rho$ となる $\hat{\beta}_{MM}$ を求める。

ここで、Tukey の Ψ 関数を例に MM 推定の第3段階を説明する。まず、 $r_i \leftarrow e_i$ において

$$\hat{u}_i = \frac{r_i}{\hat{\sigma}_M} \quad (\text{A.108})$$

と基準化した残差から、ウェイト $w(\hat{u}_i)$ を

$$w(\hat{u}_i) = \begin{cases} \left[1 - \left(\frac{\hat{u}_i}{B}\right)^2\right]^2 & |\hat{u}_i| \leq 4.691 \\ 0 & |\hat{u}_i| > 4.691 \end{cases} \quad (\text{A.109})$$

により求める。

ここで $B = 4.691$ は、Tukey の Ψ 関数で漸近的有効性 95% を達成する調整定数。そして、加重最小 2 乗推定により最初の β の推定値

$$\tilde{\beta} = (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} \mathbf{y} \quad (\text{A.110})$$

を得る。次に残差

$$e_i = Y_i - (\tilde{\beta}_1 + \tilde{\beta}_2 \mathbf{X}_{2i}), \quad i = 1, \dots, n \quad (\text{A.111})$$

を求め、

$$v_i = \frac{e_i}{\hat{\sigma}_M B} \quad (\text{A.112})$$

と基準化し、Tukey の ρ 関数の値

$$\rho(v_i) = \begin{cases} \frac{B^2}{6} (3v_i^2 - 3v_i^4 + v_i^6) & |v_i| \leq 1 \\ \frac{B^2}{6} & |v_i| > 1 \end{cases} \quad (\text{A.113})$$

を計算する。

- $\rho(v_i) < \rho^* \implies r_i \leftarrow e_i, \rho^* \leftarrow \rho(v_i)$ と置き換え、第3段階の計算ステップへ戻る
- 全ての i で $\rho(v_i) \geq \rho^* \implies$ ストップ。 $e_i, w_i, \tilde{\beta}, \rho^*$ が収束結果

以上の繰り返し再加重最小 2 乗によって得られる $\hat{\beta}_{MM}$ が β の MM 推定値である。

この MM 推定量は、第1段階の BDP50% の性質を継承する [35]。さらに、第3段階にて漸近的有効性 95% を確保して推定を行う。

従って、MM 推定は高い崩壊点と漸近的有効性を持つ頑健回帰推定法といえる。

A.4 季節調整法

A.4.1 X-12-ARIMA

本研究では、時系列データの季節調整に X-12-ARIMA を用いる。X-12-ARIMA は、米国センサス局において開発され regARIMA モデルを原系列に適用し、その将来の予測値から平滑化による季節調整を行うプログラムである。X-12-ARIMA の詳細に関しては国友 [37] を参照されたい。X-12-ARIMA を季節調整法に採用した理由として、労働力調査をはじめとする日本政府主要統計で利用されている点が挙げられる。例として本研究で用いる完全失業率は ARIMA(012)(011) モデルに基づき X-12-ARIMA プログラムによる季節調整が行われている (2020 年 1 月結果公表時点)[15]。

参考文献

- [1] Effective exchange rate indices. <https://www.bis.org/statistics/eer.htm>. (アクセス日: 12/21/2020).
- [2] 「外国為替市況」の解説. <https://www.boj.or.jp/statistics/outline/exp/exrateyen.html>. (アクセス日: 12/22/2020).
- [3] 「実効為替レート(名目・実質)」の解説. <https://www.boj.or.jp/statistics/outline/exp/exrate02.html>. (アクセス日: 12/20/2020).
- [4] 「貸出約定平均金利の推移」の解説. <https://www.boj.or.jp/statistics/outline/exp/exyaku.html>. (アクセス日: 12/22/2020).
- [5] 「短観(全国企業短期経済観測調査)」の解説. <https://www.boj.or.jp/statistics/outline/exp/tk/data/extk04.pdf>. (アクセス日: 12/20/2020).
- [6] 景気基準日付. <https://www.esri.cao.go.jp/jp/stat/di/hiduke.html>. (アクセス日: 12/18/2020).
- [7] 国債金利情報. https://www.mof.go.jp/jgbs/reference/interest_rate/. (アクセス日: 12/22/2020).
- [8] 実質実効為替レートについて. https://www.boj.or.jp/research/wps_rev/rev2011/data/rev11j01.pdf. (アクセス日: 12/21/2020).
- [9] 住宅宿泊事業法の施行状況. https://www.mlit.go.jp/kankocho/minpaku/business/host/construction_situation.html. (アクセス日: 01/19/2021).
- [10] 全国企業倒産集計-2007年度報・2008年3月報-. <https://www.tdb.co.jp/report/tosan/syukei/pdf/07nendo.pdf>. (アクセス日: 12/20/2020).
- [11] 長・短期プライムレート(主要行)の推移. <https://www.boj.or.jp/statistics/dl/loan/prime/prime.html>. (アクセス日: 12/22/2020).
- [12] 倒産の定義|株式会社 帝国データバンク [tdb]. <https://www.tdb.co.jp/tosan/teigi.html>. (アクセス日: 12/18/2020).
- [13] 倒産集計|株式会社 帝国データバンク [tdb]. <https://www.tdb.co.jp/tosan/syukei/>. (アクセス日: 12/18/2020).
- [14] 統計局ホームページ/労働力調査(基本集計) 2020年(令和2年)11月分結果. <https://www.stat.go.jp/data/roudou/sokuhou/tsuki/index.html>. (アクセス日: 01/06/2021).
- [15] 統計局ホームページ/労働力調査/季節調整値の算出方法. <http://www.stat.go.jp/data/roudou/kisetsu/index.html>. (アクセス日: 01/10/2021).

- [16] 統計局ホームページ/労働力調査/用語の解説. <https://www.stat.go.jp/data/roudou/definit.html>. (アクセス日: 01/06/2021).
- [17] E.I. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, Vol. 23, No. 4, pp. 589–609, 1968.
- [18] W.H. Beaver. Financial ratios as predictors of failure. *Journal of Accounting Research*, Vol. 4, pp. 71–111, 1966.
- [19] T. Chakraborty, A.K. Chakraborty, M. Biswas, S. Banerjee, and S. Bhattacharya. Unemployment Rate Forecasting: A Hybrid Approach. *Computational Economics*, 2020.
- [20] M.Y. Chen. Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers Mathematics with Applications*, Vol. 62, No. 12, pp. 4514 – 4524, 2011.
- [21] C.L. Chua, G.C. Lim, and S. Tsiaplias. A latent variable approach to forecasting the unemployment rate. *Journal of Forecasting*, Vol. 31, No. 3, pp. 229–244, 2012.
- [22] O. Claveria. Forecasting the unemployment rate using the degree of agreement in consumer unemployment expectations. *Journal for Labour Market Research*, Vol. 53, No. 1, p. 3, 2019.
- [23] J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression.III. *Biometrika*, Vol. 58, No. 1, pp. 1–19, 04 1971.
- [24] Y. Fondeur and F. Karamé. Can google data help predict french youth unemployment? *Economic Modelling*, Vol. 30, pp. 117 – 125, 2013.
- [25] F.R. Hampel, E. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust statistics: the approach based on influence functions*. 1986.
- [26] D.C. Hoaglin, F. Mosteller, and J.W. Tukey. *Understanding Robust and Exploratory Data Analysis*. Wiley-Interscience, 1 edition, 2000.
- [27] P.J. Huber and E.M. Ronchetti. *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley, 2009.
- [28] C. Katris. Prediction of Unemployment Rates with Time Series and Machine Learning Techniques. *Computational Economics*, Vol. 55, No. 2, pp. 673–706, 2020.
- [29] E.K. Laitinen and T. Laitinen. Bankruptcy prediction: Application of the taylor’s expansion in logistic regression. *International Review of Financial Analysis*, Vol. 9, No. 4, pp. 327 – 349, 2000.
- [30] B. Maas. Short-term forecasting of the us unemployment rate. *Journal of Forecasting*, Vol. 39, No. 3, pp. 394–411, 2020.
- [31] R.A. Maronna, D.R. Martin, and V.J. Yohai. *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics. Wiley, 2006.

- [32] A.L. Montgomery, V. Zarnowitz, R.S. Tsay, and G.C. Tiao. Forecasting the u.s. unemployment rate. *Journal of the American Statistical Association*, Vol. 93, No. 442, pp. 478–493, 1998.
- [33] P. Rousseeuw and V. Yohai. Robust regression by means of s-estimators. In Jürgen Franke, Wolfgang Härdle, and Douglas Martin, editors, *Robust and Nonlinear Time Series Analysis*, pp. 256–272, New York, NY, 1984. Springer US.
- [34] P. Smith. Google’s midas touch: Predicting uk unemployment with internet search data. *Journal of Forecasting*, Vol. 35, No. 3, pp. 263–284, 2016.
- [35] V.J. Yohai. High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.*, Vol. 15, No. 2, pp. 642–656, 06 1987.
- [36] V.J. Yohai and R.A. Maronna. Asymptotic behavior of m -estimators for the linear model. *Ann. Statist.*, Vol. 7, No. 2, pp. 258–268, 03 1979.
- [37] 国友直人. 季節調整法. Discussion paper series. CIRJE-J, jan 2006.
- [38] 蓑谷千鳳彦. 計量経済学の新しい展開. 多賀出版, 単行本, 1 1992.
- [39] 蓑谷千鳳彦. 計量経済学 (スタンダード経済学シリーズ). 東洋経済新報社, 単行本, 4 1997.
- [40] 蓑谷千鳳彦. 計量経済学大全. 東洋経済新報社, 単行本, 2 2007.
- [41] 蓑谷千鳳彦. 頑健回帰推定 (統計ライブラリー). 朝倉書店, 単行本, 2 2016.
- [42] 沖本竜義. 経済・ファイナンスデータの計量時系列分析 (統計ライブラリー). 朝倉書店, 単行本, 2 2010.