

2022年度 統計データ分析コンペティション

審査員奨励賞 [高校生の部]

景気の変動要因と景気の変動による影響の分析

佐々木 俊輔、鈴木 権人
(東京都立大泉高等学校)

景気の変動要因と景気の変動による影響の分析

佐々木俊輔*1・鈴木権人*2

*1: 東京都立大泉高等学校

*2: 東京都立大泉高等学校

1. 研究のテーマと目的

現在経済界では円安が進行しており、令和4年9月5日現在1ドルあたり140.42円と日本経済に大きな打撃を与えている。

一般に景気動向の指標として扱われる数値の一つに、有効求人倍率がある。有効求人倍率は、有効求人数を有効求職者数で割った値であり、この値が高くなるほど景気が良くなるというのが自然な見解である。

本研究では、苦しい状況にあるとされている日本経済の景気向上を目的として、有効求人倍率の向上に作用する要素を調査した。

2. 研究の方法と手順

本研究では、各年の各都道府県における月間有効求人倍率（一般）（計算方法は第3節参照）を、景気を表す指標として用い、これを被説明変数とした分析を行った。表現力が高く、予測精度の高いモデルの構築を目指し回帰型のニューラルネットワークを構築し、そのモデルによる SHAP 値を求めることとした。

2.1 ニューラルネットワークの構造

入力値に対し、線形変換と活性化関数による変換を繰り返し、実測値と予測値との損失を求める。活性化関数は、Mish 関数を用いる。正規化について、データ数が少ないため、Dropout は有効でないと考え行わず、Batch Normalization に関してもバッチの大きさが小さいため行わない。MSE（平均二乗誤差）に L2 ノルムを加算した値が最小となるようにパラメータチューニングを行った。最適化アルゴリズムには、SGD（確率的勾配降下法）、AdaGrad、Adam の3種類を試し、結果を確認しながら決定する。中間層数、中間層ノード数、正規化項の係数、バッチの大きさをハイパーパラメータとし、結果を確認しながら決定する。なお、中間層のノード数はすべて同じとする。

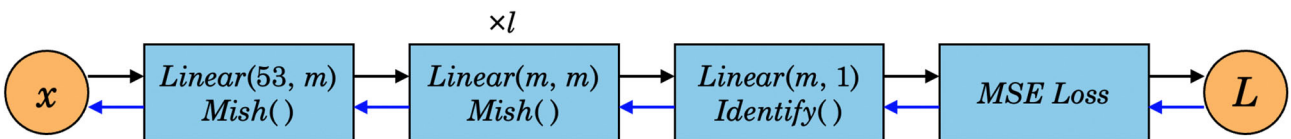


図1 ニューラルネットワークモデルのイメージ

2.1.1 線形変換

構築するニューラルネットワークは全結合層のみで構成される。全結合層では、隣接する2層間の全てのニューロンユニット間において、線形重みパラメータによる線形識別的な変換を行う。下式は、線形変換を表したものである。

$$\sum_{k=1}^m x_k w_k + b = y$$

m : ノード数, b : バイアス, w : 重み, x : 入力値, y : 出力値

2.1.2 Mish 関数

活性化関数には、Mish 関数を用いる。Mish 関数は、ReLU 関数に似た曲線を描き、ReLU 関数と同様に、誤差逆伝播における勾配消失問題を起こしにくいという特長がある。また、完全な連続関数であるため学習結果が滑らかになる点や、入力が負の値の際も微分値が 0 にならない点においても優れている。一方、ReLU 関数と比較し計算コストは上がる。

$$\text{mish}(x) = x \cdot \tanh(\log(1 + \exp(x)))$$

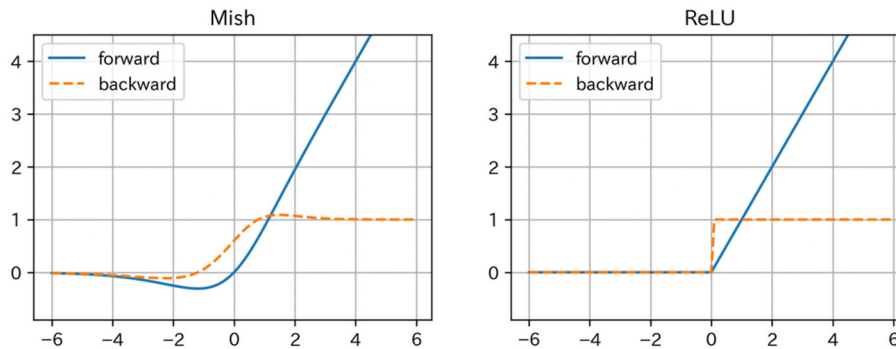


図 2 Mish 関数と ReLU 関数の比較

2.1.3 損失関数

損失関数は、実測値と予測値の平均二乗誤差に L2 正則化項を加算したものとする。回帰式が複雑な形状になる場合にペナルティを設けることにより、回帰式を単純な形状に保つことができ、汎化性能の向上が期待できる。

$$L = \frac{1}{n} \sum_{j=1}^n (t_j - y_j)^2 + \frac{1}{2} \alpha \sum_{k=1}^m w_k^2$$

L : 損失関数, n : データ数, t : 実測値, y : 出力値, α : ハイパーパラメータ, m : 一層のノード数, w : 重み

2.2 SHAP の概要

SHAP (SHapley Additive exPlanations) とは、XAI (eXplainable AI) 技術の 1 つで、モデルにて出力された予測結果に対し、各説明変数がどのような影響を与えたかを定量的に算出する手法である。ニューラルネットワークモデルには DeepExplainer インスタンスを用いる。

3. データセットの概要

3.1 データセットの加工

作成したデータセットのサンプルは 2005 年から 2019 までの 47 都道府県であり、データ数 $N=705$ である。分析に用いる指標とその計算方法及び出典を表 1 に示す。

表1 使用するデータ

| 項目名 ^{※1} | 計算方法 | 出典 |
|---------------------------------|---|---|
| 月間有効求人倍率（一般） | 月間有効求人数（一般） / 月間有効求職者数（一般） | |
| 総人口に占める男性の割合 | 総人口（男） / 総人口 | |
| 15歳未満人口割合 | 15歳未満人口 / 総人口 | |
| 15歳～64歳人口割合 | 15歳～64歳人口 / 総人口 | |
| 65歳以上人口割合 | 65歳以上人口 / 総人口 | |
| 合計特殊出生率 | | |
| * 転入者数（日本人移動者） | | |
| * 転出者数（日本人移動者） | | |
| * 結婚件数 | | |
| * 離婚件数 | | |
| * 着工建築物数 | | |
| * 着工新設住宅戸数 | | |
| * 着工新設持家数 | | |
| * 着工新設貸家数 | | |
| * 着工新設分譲住宅数 | | |
| 標準価格（平均価格）（住宅地） | | |
| 標準価格（平均価格）（商業地） | | |
| * 医療施設数 | 一般病院数 + 一般診療所数 | |
| * 保育園等在所児数 | | |
| 保育所等一施設あたりの在所児数 | 保育所等在所児数 / 保育所等数 | |
| 保育所等保育士一人当たりの在所児数 | 保育所等在所児数 / 保育所等保育士数 | |
| * 幼稚園在園者数 | | |
| 幼稚園一園あたりの在園者数 | 幼稚園在園者数 / 幼稚園数 | |
| 幼稚園教員一人当たりの在園者数 | 幼稚園在園者数 / 幼稚園教員数 | SSDSE-B-2022, SSDSE-2019B |
| 小学校一校あたりの児童数 | 小学校児童数 / 小学校数 | |
| 小学校教員一人当たりの児童数 | 小学校児童数 / 小学校教員数 | |
| 中学校一校あたりの生徒数 | 中学校生徒数 / 中学校数 | |
| 中学校教員一人当たりの生徒数 | 中学校生徒数 / 中学校教員数 | |
| 中学校卒業者のうち進学者数の割合 | 中学校卒業者のうち進学者数 / 中学校卒業者数 | |
| 高等学校一校あたりの生徒数 | 高等学校生徒数 / 高等学校数 | |
| 高等学校教員一人当たりの生徒数 | 高等学校生徒数 / 高等学校教員数 | |
| 高等学校卒業者のうち進学者数の割合 | 高等学校卒業者のうち進学者数 / 高等学校卒業者数 | |
| 大学一校あたりの学生数 | 大学学生数 / 大学数 | |
| 大学教員一人当たりの学生数 | 大学学生数 / 大学教員数 | |
| 大学卒業者のうち進学者数の割合 | 大学卒業者のうち進学者数 / 大学卒業者数 | |
| 消費支出（二人以上の世帯） | | |
| 食料費（二人以上の世帯） | | |
| 住居費（二人以上の世帯） | | |
| 光熱・水道費（二人以上の世帯） | | |
| 家具・家事用品費（二人以上の世帯） | | |
| 被服及び履物費（二人以上の世帯） | | |
| 保健医療費（二人以上の世帯） | | |
| 交通・通信費（二人以上の世帯） | | |
| 教育費（二人以上の世帯） | | |
| 教養娯楽費（二人以上の世帯） | | |
| その他の消費支出（二人以上の世帯） | | |
| ドル円為替レート ^{※2} | ドル円為替レート（7月） | 日本銀行 ^[1] |
| 日銀金利 ^{※2} | 無担保コールO/N物レート（7月） | |
| 米FRB金利 ^{※2} | フェデラル・ファンドレート（7月） | Federal Reserve Board ^[2] |
| 日経平均株価 ^{※2} | 日経平均株価（12月終値） | |
| NYダウ平均株価 ^{※2} | NYダウ平均株価（12月終値） | Yahoo! Finance ^[3] |
| * 県内総生産（名目）[2006～2018] | | 県民経済計算（平成18年度 - 平成30年度） （2008SNA, 平成23年基準） ^[4] |
| * 県内総生産（名目）[2005] ^{※3} | 県内総生産（名目）[2006] / （県内総生産（名目）（平成17年基準）[2006] / 県内総生産（名目）（平成17年基準）[2005]） | 県民経済計算（平成13年度 - 平成26年度） （1993SNA, 平成17年基準） ^[4] |
| * 県内総生産（名目）[2019] ^{※3} | 県内総生産（名目）[2018] * （県内総生産（名目）（平成27年基準）[2019] / 県内総生産（名目）（平成27年基準）[2018]） | 県民経済計算（平成23年度 - 令和元年度） （2008SNA, 平成27年基準） ^{※4[4]} |

項目名のうち、先頭にアスタリスク（*）が付されている指標については、値を公平に比較するため総人口で除する（※1）。ドル円為替レート、日銀金利、米FRB金利、日経平均株価およびNYダウ平均株価は、47都道府県で同じ値を使用している（※2）。また、県内総生産の算出の基準が年度により異なる箇所があり、値を公平に比較するため上記の処理を行う（※3）。県民経済計算（平成23年度 - 令和元年度）（2008SNA, 平成27年基準）においてデータが公表されていない16都道府県（2022年8月現在）については、公表されている31都道府県の平均値を用いて計算する（※4）。

3.2 データセットの前処理

N 数 705 のサンプルを、訓練データ (60%)、検証データ (20%)、テストデータ (20%) に分割し、訓練データを標準化 (平均 0, 分散 1 となるように変換) し、検証データ、テストデータは訓練データの平均と標準偏差の値を用いて次式にて変換する.

$$x'_i = \frac{x_i - \bar{x}}{\sigma}$$

\bar{x} : 訓練データの平均, σ : 訓練データの標準偏差, x : 変換前の値, x' : 変換後の値

4. データ分析の結果

4.1 ニューラルネットワークの構造

PyTorch ライブラリを用い、ニューラルネットワークモデルを構築した. 最適化アルゴリズムには精度が最も安定していた AdaGrad を用い、中間層数 16, 中間層ノード数 41, 正則化項の係数 0.002, バッチの大きさ 3 と決定した. 学習率 0.01 で 150 エポック学習を繰り返した.

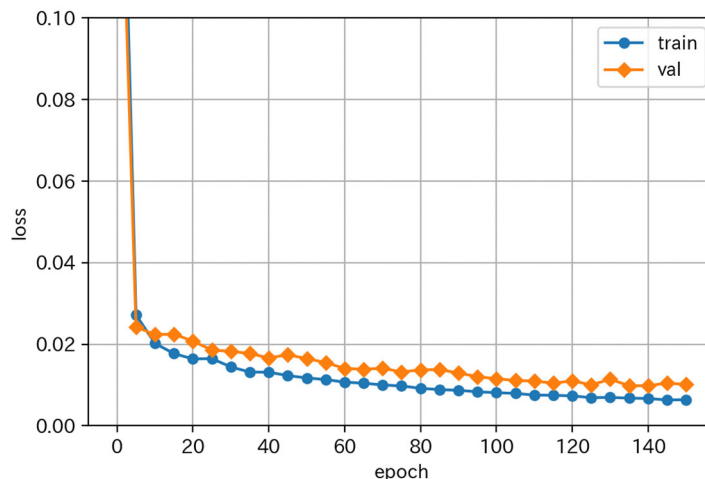


図3 損失の収束

図3はエポックごとの損失を表したものであり、最終的な損失は、訓練 (train) データが 0.00642, 検証 (validation) データが 0.0101 であった. 全体を通して訓練データの損失の方が小さいが、損失は両者ともにとっても小さく、問題になる規模での過学習は起きていないと判断する. 訓練及び検証に使用していないテストデータの損失が 0.0104, テストデータによる予測値と実測値による決定係数は 0.939 であり、予測精度の高いモデルであると言える.

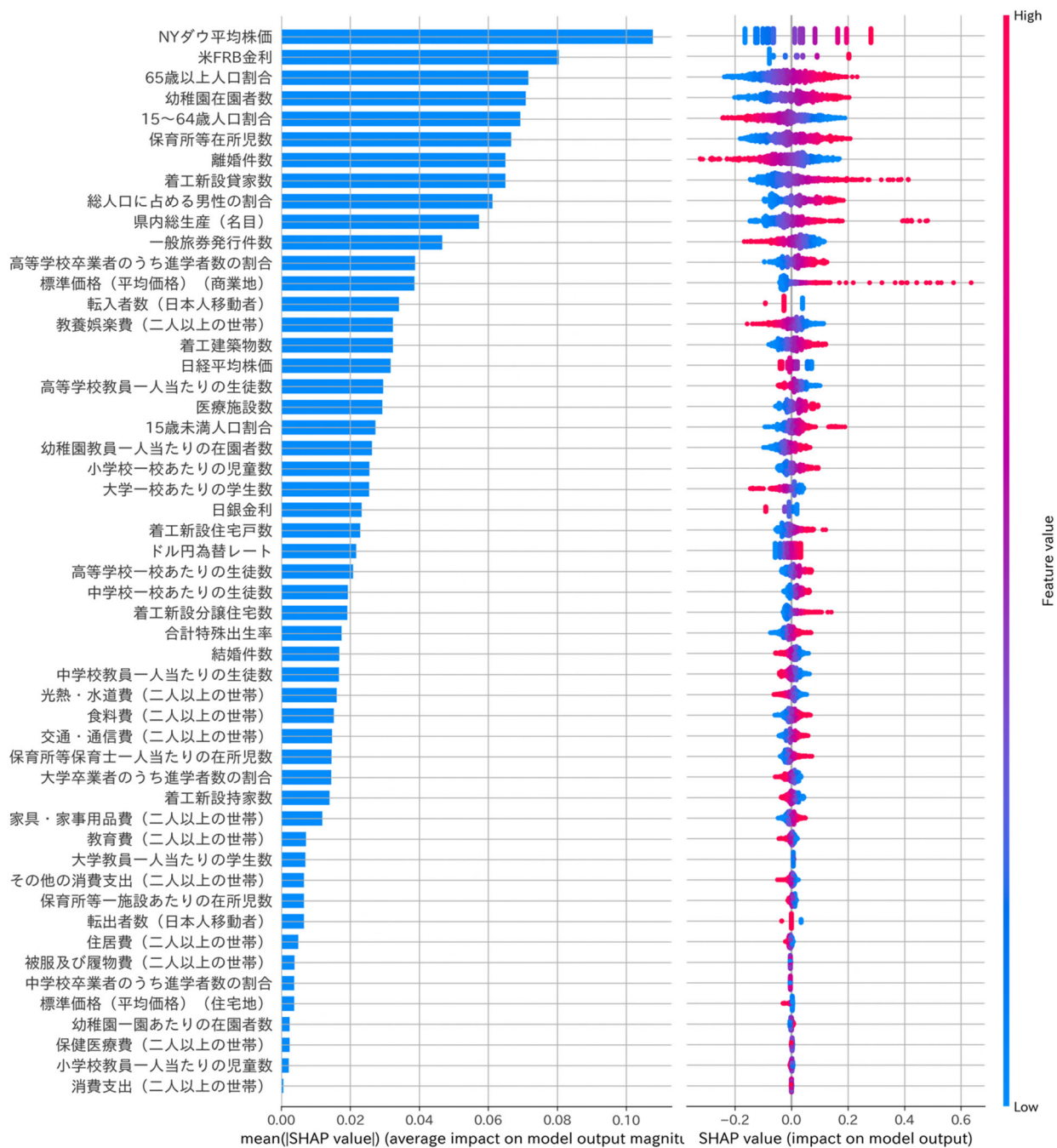


図4 SHAP 値の可視化

図4左は、各説明変数の貢献度を表しており、棒グラフの長さは各説明変数の重要度と解釈できる。同図右は、各説明変数と予測値との相関関係を可視化したものである。図4右のグラフの横軸は、各説明変数が予測値の正負のどちらの方向にどれだけ寄与しているかを、色は各説明変数の値の高低を示している。

5. 結果の解釈

5.1 生活環境について

保健医療費（二人以上の世帯）、被服及び履物費（二人以上の世帯）、住居費（二人以上の世帯）の貢献度は認められず、教育費（二人以上の世帯）、家具・家事用品費（二人以上の世帯）、交通・通信費（二人以上の世帯）の貢献度は

帯), 食料費 (二人以上の世帯), 光熱・水道費 (二人以上の世帯) との相関関係も小さい。これらの消費支出は, 生活をする上で必要不可欠な費用であるため, 景気の良し悪しによる影響をあまり受けないと考えられる。一方でこれらの消費支出系の中では, 食料費の貢献度が高く正の相関関係を示している。これは, 好景気であると外食等のサービス消費が増加するためだと考えられる。

5.2 教育について

教育面では, 全体的に他の分野に比べて貢献度が低い, この分野の中では幼稚園在園者数や保育所等在所児数の貢献度が上位となっている。景気が良くなるとお金の流れが活発になり, 子育てにかけられる費用も増加するためだと考えられる。

教育費と有効求人倍率 (一般) との間の明瞭な相関関係は認められず, 教育にかけられる費用は景気にあまり左右されないと読み取ることができる。教育は景気の動向による影響を受けにくいことが示された。

5.3 社会・経済について

全説明変数の中でも, NY ダウ平均株価や米 FRB 金利は強い正の相関があり, 日本の景気はアメリカ経済の影響を強く受けていることが示された。米国は, 日本の2番目の輸入・輸出相手国であり, 米国での輸出入においても日本は主要な貿易相手国である (外務省^[5])。日本と米国は, 貿易の面での強力な結びつきがあることが分かる。

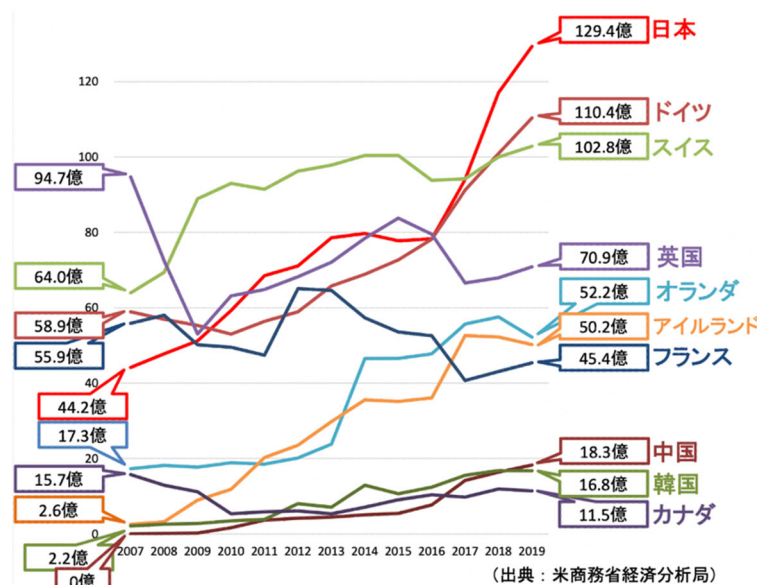


図5 在米各国企業のR&D投資額の推移 (外務省^[5]より)

また日本は研究・開発 (R&D) への投資が年々増加しており, 2019年は世界最大の対米投資国となっている (図5)。この投資額は今後も増加し, 日米間の経済的な結びつきはより一層強固なものになると見込まれ, 日本の景気は今後もアメリカ経済に強く影響され続けると考えられる。

有効求人倍率 (一般) は 65 歳以上人口割合と正の相関関係にあり, それとは対象に 15~64 歳人口割合とは負の関係である。求職者数は生産年齢人口と密接に関わり, これは有効求人倍率 (一般) の定義により直接的に影響する。高齢者人口割合が変化すると, 相対的に生産年齢人口割合は逆に変化する。生産年齢人口割合との相関と高齢者人口割合との相関の正負が不一致であるのはこのためだと考えられる。

総人口に占める男性の割合の貢献度は高く, 正の相関が見られた。これは, 男性割合が高い地域は景気が良

いことを意味し、明らかな男女格差が認められる。女性の社会進出を促進させることで、景気にも良い影響を与えると考えられる。

また図4より、一般旅券発行件数は有効求人倍率（一般）と負の相関があることがわかる。世界的な金融危機であった2008年のリーマンショック直後に、1ドル=110円台だった円相場が90円台まで急騰した。日本でも記録的な不景気となり、円高になったため海外旅行に行く人が増えた。これと同様の要因であると考えられる。

6. 本研究について

本研究では、有効求人倍率（一般）を目的変数とした精度の高い予測モデルにより、景気の動向に関連する指標を調査した。予測精度の高いモデルによって、考察を進めたところに特徴がある。

注意すべき点を挙げる。SHAPにより求められる各説明変数の貢献度は、パラメータのわずかな値の違いによって変動しうる。そのため貢献度の細かい順位については慎重に言及する必要がある。また有効求人倍率（一般）は、あくまでも就職の状況を表すものであるため、どの分野の事象においても景気動向の指標として常に用いることができるのかという点については議論が必要である。

技術的な問題点を挙げる。SHAPにより説明できることは各説明変数と予測結果の関連性であり、事象の発生順序については読み取ることができない。そのため、有効求人倍率（一般）を増加させるための有効な方法であるかどうかについては追加の検討が必要である。また、本研究には一年ごとのデータを使用しているため、年を超えて長期的に影響する指標については説明不可能である。長期変動・季節変動の説明には、データセットの工夫や、時系列データの分析に適する Prophet などの統計的手法を用いることなどが必要であると考えられる。

出典

- [1] 日本銀行（2022）。「日本銀行時系列データ検索サイト」。
<https://www.stat-search.boj.or.jp>. 2022年9月8日.
- [2] Federal Reserve Board（2022）。「Federal Reserve Board - Home」。
<https://www.federalreserve.gov>. 2022年9月8日.
- [3] Yahoo!（2022）。「Yahoo! Finance」。<https://finance.yahoo.com>. 2022年9月8日.
- [4] 内閣府（2022）。「県民経済計算」。
https://www.esri.cao.go.jp/jp/sna/sonota/kenmin/kenmin_top.html. 2022年9月8日.
- [5] 外務省（2022）。「米国経済と日米経済関係」。
<https://www.mofa.go.jp/mofaj/files/000240495.pdf>. 2022年9月8日.