

機械学習による景気局面判断の分析

An analysis of business cycles by machine learning

山澤成康

Nariyasu YAMASAWA

要 旨

本研究は、政府の景気局面判断がどのようなメカニズムでなされているのかを機械学習の手法を使って実証的に検討したものだ。

被説明変数を、政府の景気基準日付に基づく景気局面（拡大期=0、後退期=1）、説明変数を景気動向指数（先行、一致、遅行）のすべての構成指標 27 系列（前期差）とした。サンプルを教師データと評価データの 2 つに分けて予測精度を調べるホールドアウトテストでは、手法の違いによりばらつきはあるが、おおむね良好な結果となった。

次に外挿テストを行った。2012 年 3 月の景気の「山」までを教師データとしてそこから先を予測するケースと 2012 年 11 月の景気の「谷」までを教師データとしてそこから先を予測するケースの 2 種類を行い、局面判断ができるかを検討した。決定木を使うと概ね的中することがわかった。

次に決定木を使って、景気局面判断に各指標がどのように使われているかについて調べた。2014 年以前の景気局面判断は簡明で、DI とその差分の 2 つでほぼ判断ができていた。しかし、2014 年の消費税率引き上げ以降は、局面判断の基準が変わっていることがわかった。2018 年以降、公式の局面判断では景気後退のサインが出ていないが、従来の手法では景気後退のサインが頻繁に出ていることが確認できた。

最後に、決定木の予測をランダムフォレスト、勾配ブースティングなどアンサンブル予測も行って、予測が改善するかどうかを検証した。結果は決定木を使った場合と同様だった。

1. はじめに

景気が拡大局面にあるか後退局面にあるかを定める景気基準日付は、有識者で構成される内閣

府の景気動向指数研究会で決定される。景気基準日付は、どのような手順で決定していると分析できるだろうか。本稿では人工知能（AI）に関わる分析技術である機械学習の手法を使ってその手順を探り、最新の経済指標に適用した結果を紹介する。

現在の景気基準日付決定法（と推測される手順）を前提にすれば、2019年4月までは景気拡大期が続いていることになる。つまり、2012年11月を景気の谷とする戦後最長の景気拡大期が続いていると予測される。しかし、2014年の消費税率上昇前後で景気判断の手順が変わっており、以前の判断基準なら景気後退のサインが頻繁に出ていることが分かる。

2. 先行研究

景気判断についての先行研究は数多くある。たとえば、浦沢（2017）は、景気循環の特徴を探り、2000年前後に構造変化があることを明らかにした。

景気分析に機械学習を使ったものとして、山本・松尾（2016）は、景気ウォッチャー調査の言葉を学習し、政府や日銀のレポートを指数化した。

大高・菅（2018）は、予想インフレ率やインフレ率に関連する言葉を抽出して、ランダムフォレストの手法を使っている。

小寺ほか（2018）は、POS データに、機械学習を使い商業販売額の予測を試みている。

本稿では、景気の拡大期、後退期を判断するにあたって景気動向指数構成指数などを利用し、機械学習を行うものだが、こうした試みは初めてである。

3. 決定木を使った景気分析

3.1 決定木とは

本稿では、上述のように応用が進んできた機械学習の手法のうち、木構造を活用した決定木を使う。決定木は手法としては1980年代からある古いものだが、解釈がしやすいため経済分析に向いている。木構造を使った手法の発展形であるランダムフォレストと勾配ブースティングを使って予測も試みた。

決定木は、データを2つのカテゴリーに分けることを目標とする。たとえば、各変数に○か×の属性があり、それをうまく○のグループと×のグループに分けることが目標だ。○と×に分ける基準としては、「変数Xの値が5以上」といったある閾値を境にして二つに分けることとする。

2つにうまく分けられたかどうかの基準に、ジニ不純度係数などを使う。経済学の不平等度で使うジニ係数とは別のものである。

2つに分けたグループの不純度の和と元のグループの不純度の差が最大になるように、閾値を決める。

式で表すと、以下のようになる（金森 2017）。2つの状態を示すラベル付きのデータがあるとする。データ $\mathbf{x}=\{x_1\cdots x_d\} \in \mathbf{R}^d$ に対して、条件に従って複数に分類していく。最初のデータを「 $\mathbf{x}k>C$ を満たすかどうかで L_1, L_2 にわけ、それぞれのラベルも y_1, y_2 を割り当てる。」これを繰り返すことにより、グループの分岐が行われる。

ある程度の純度になると分岐をやめることにしておき、複雑に分岐し過ぎるのは避ける。予測することを考えると過去のデータをきちんと分けるだけが最善ではないためだ。モデルが過度に学習データに適合すると、サンプル外集合へ対応できなくなり、この現象は過学習と呼ばれる。予測用にはある程度粗い分類にとどめておく必要がある。これを検証するために「複雑度」が計算される。

3.2 景気分析に応用

本研究では、決定木を景気の分析に使った。景気動向指数の一致指数 **DI** がゼロ以上かゼロ未満かで、景気拡大期か景気後退期かの2つのグループに分けるといった方法だ。2つに分けた結果が、景気拡大期と景気後退期にきれいに分けられればよいが、そう簡単ではない。うまく分けられない場合は他の変数を使って2つに分ける。これを繰り返すことで、分類の精度が増していくことになる。本稿では、2つに分ける基準には、ジニ不純度（サンプルがすべて同じグループの場合は1、サンプルがそれぞれ違うグループの場合はゼロになる係数）を使った。

4. データ

使ったデータは、予測の対象となる目的変数としては景気拡大期か後退期かの2値変数を用いた。また、目的変数を説明する変数としては、景気基準日付を決定するに際して重要視していると想定される変数を用いた。具体的には、景気動向指数の一致指数、遅行指数、先行指数について、各指数の構成系列（計29系列）、ディフュージョン・インデックス（以下「**DI**」という。）（計3系列）、累積 **DI**（計3系列）、コンポジット・インデックス（以下「**CI**」という。）（計3系列）、調査月（1系列）である¹。集めたデータの期間は1985年1月から2019年4月までとした。景気拡大期か後退期かの2値変数と調査月以外は、前期差のデータを使った。

景気拡大期か後退期かの分類ラベルは内閣府の景気基準日付を使い、1985年2月から2017年8月までの391か月をサンプルとした。2017年8月を終期としたのは、景気の谷の最新期である2012年11月の後において、2018年12月に開かれた景気動向指数研究会の結論として、「2017年8月以前に景気の山は設定されない」と内閣府社会総合研究所が発表したためだ。

5. モデルの精度について

決定木の分析には CART (Classification And Regression Trees) というアルゴリズムを使い、ソフトウェア R の rpart パッケージを使用した (下川・杉本・後藤 2013 など参照)。

過去のデータを二つに分け、一方を学習用、一方を正解としたホールドアウトテストを行うと、的中率が最大の葉の数を探すと cp が 0.01 で葉の数は 5 個となり、87.7% の的中率だった。

全期間に占める景気拡大期の比率は約 7 割なので、常に景気拡大期と予測しても的中率は 7 割を超える。こうした有利さも考慮して当てはまりをまず、混同行列で評価してみよう。これを使うと、再現率 (本当に景気後退期の場合に、何% 景気後退期と予測できたか) 81.6%、適合率 (景気後退期と予測したうち、どれだけが本当に景気後退期だったか) 72.7% が計算できる。

図表 1 ホールドアウトテスト (手法の違い)

| 手法 | 正答率 (%) |
|---------------------|---------|
| 決定木 | 87.7 |
| k 近傍法 | 86.7 |
| ニューラルネット | 83.0 |
| サポートベクターマシン | 84.1 |
| ナイーブベイズ | 85.6 |
| ランダムフォレスト (変数選択 25) | 88.2 |

図表 2 ホールドアウトテスト (変数の違い)

| 決定木 | 正答率 (%) |
|------------------|---------|
| 時間変数あり、DI, CI あり | 87.7 |
| 時間変数なし、DI, CI あり | 85.1 |
| 変数のみ | 79.5 |

1 各指数の構成系列、DI、累積 DI、CI の内容や動きの詳細は内閣府経済社会総合研究所の景気動向指数のウェブページ (<https://www.esri.cao.go.jp/jp/stat/di/di.html>) 参照。

機械学習による景気局面判断の分析

また、偶然による的中率²も考慮した指標の一つに「カッパ係数」がある。

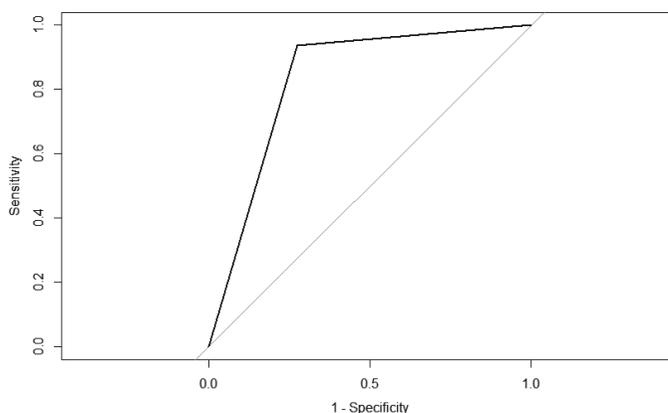
$$\text{カッパ係数} = \frac{\text{的中率} - \text{偶然による的中率}}{1 - \text{偶然による的中率}}$$

ゼロから1までの値をとり、大きいほど当てはまりが良い。カッパ係数を計算すると0.69であり、当てはまり具合としては「良好」と判断される水準である。

次に、ROC 曲線を調べてみた。縦軸に真陽性率、横軸に偽陽性率をとったものだ。曲線が左上に引っ張られるほど、良い予測といえる。予測値をもとに、ROC 曲線より下の面積が全体を1とした時どの程度を占めるかを表したものがAUC (Area Under the Curve) である。AUCは0.8315で良好だ。

図表3 混同行列

| | | 実績 | | |
|----|----|----|-----|-----|
| | | 後退 | 拡大 | 合計 |
| 予測 | 後退 | 40 | 15 | 55 |
| | 拡大 | 9 | 131 | 140 |
| | 合計 | 49 | 146 | 195 |



図表4 ROC 曲線

2 偶然による的中率は以下のように計算できる。実績値が景気後退期で予測値も景気後退期になる確率： $0.282 \times 0.251 = 0.071$ 、実績値が景気拡大期で予測値も景気後退期になる確率： $0.718 \times 0.749 = 0.538$ 、偶然による的中率： $0.071 + 0.538 = 0.609$

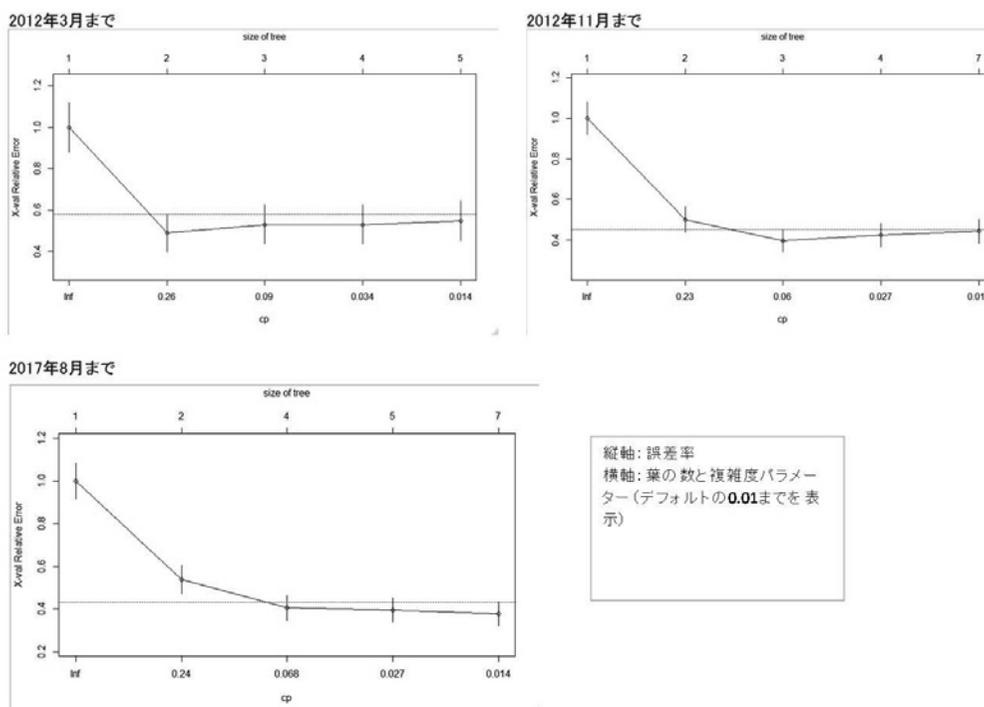
6. 外挿テストと予測

次に、現実に応じた予測精度をみるために、外挿テストを行った。三種類のケースを想定した。

① 1985年2月から2012年3月までのデータを用いるもの② 1985年から2012年11月までのデータを用いるもの③ 1985年2月から2017年8月までを用いるものでその後を予測するもの——である。2017年8月までは実績値があるので、そこまでの期間は外挿テスト、2017年9月から2019年4月は実績値がないので予測ということになる。予測は観測時点から将来に向けてするものだとすれば、足元の状況を予測するという意味でナウキャストिंगといってもよい。

決定木は過学習を防ぐため、ある程度分岐点を少なくする必要がある。この分析では、横軸に複雑さのパラメーター (cp)、縦軸に交差確認法による誤差率をとり、誤差が最小になる複雑さのパラメーターを決めた。複雑さのパラメーターによって葉の数が決まってくる。

ハイパーパラメーターである決定木の葉の数（最終的な分類数）は、交差確認法による誤差が最も小さくなる場所で選んだ。



図表5 複雑度と葉の数

機械学習による景気局面判断の分析

推計結果は表の通り。ケース1は、2012年3月の景気の高まりまでのデータを使って、2012年4月以降の予測ができるかどうかをテストしたものだ。2012年4月は外したものの、ほぼ景気後退期を予測できていることがわかる。その後については、消費税率が5%から8%に上昇した2014年に、景気後退期のサインが何度か出ている。2014年5月、6月、8月、12月、2015年3月である。連続して景気後退期になっていないことから、この部分を景気後退期と認定することは難しいが、景気後退期のサインが何度もでていたことがわかる。

ケース2は、2012年4月から2012年11月までの景気後退期まで学習期間を延長したもののだが、ケース1と結果は同じだった。

2018年9月以降は実績値のない純粋な予測期間である。ケース1とケース2では、2018年中に6回景気後退のサインが出ている。一方、2017年8月までを学習期間とすると、景気後退期のサインは大きく減る。これらの解釈については、次節に譲る。



図表6 学習データと外挿期間、予測期間

図表7 外挿、予測結果

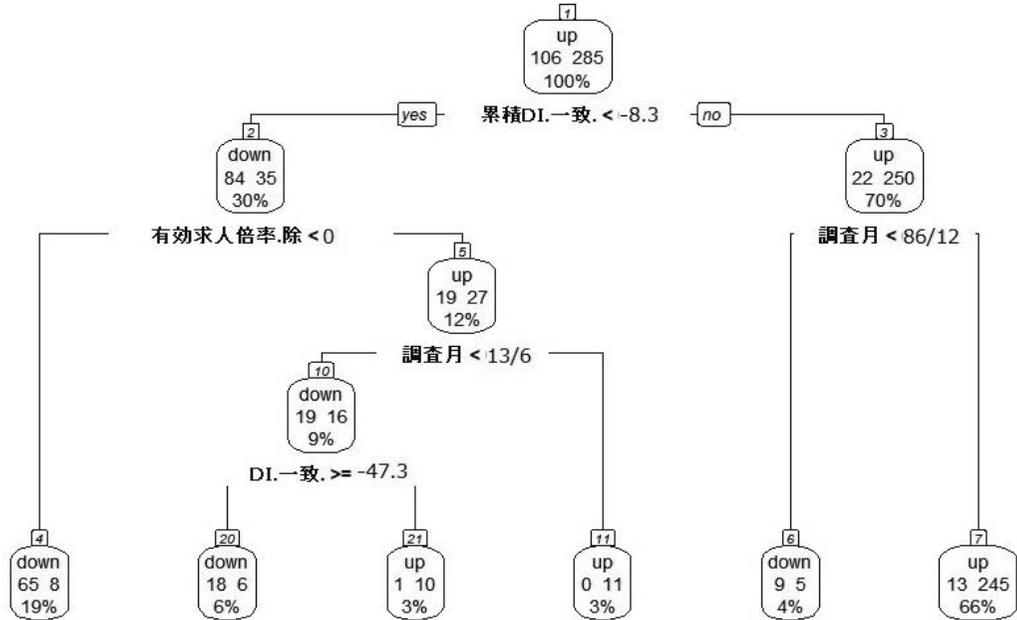
| 日付 | データの種 類 | 実績 | 決定木 | | |
|---------|------------|----|----------------------|---------------------------|----------------------|
| | | | ケース1:2012 年4月以降推計 | ケース2:2012 年12月以降推 計 | ケース3:2017 年9月以降推計 |
| 2012/04 | 学習データ | 1 | 0 | | |
| 2012/05 | | 1 | 1 | | |
| 2012/06 | | 1 | 1 | | |
| 2012/07 | | 1 | 1 | | |
| 2012/08 | | 1 | 1 | | |
| 2012/09 | | 1 | 1 | | |
| 2012/10 | | 1 | 1 | | |
| 2012/11 | | 1 | 1 | | |
| 2012/12 | | 0 | 0 | 0 | |
| 2013/01 | | 0 | 0 | 0 | 0 |
| 2013/02 | | 0 | 0 | 0 | 0 |
| 2013/03 | | 0 | 0 | 0 | 0 |
| 2013/04 | 0 | 0 | 0 | 0 | |
| 2013/05 | 0 | 0 | 0 | 0 | |
| 2013/06 | 0 | 0 | 0 | 0 | |
| 2013/07 | 0 | 0 | 0 | 0 | |
| 2013/08 | 0 | 0 | 0 | 0 | |
| 2013/09 | 0 | 0 | 0 | 0 | |
| 2013/10 | 0 | 0 | 0 | 0 | |
| 2013/11 | 0 | 0 | 0 | 0 | |
| 2013/12 | 0 | 0 | 0 | 0 | |
| 2014/01 | 外挿期間 | 0 | 0 | 0 | 0 |
| 2014/02 | | 0 | 0 | 0 | 0 |
| 2014/03 | | 0 | 0 | 0 | 0 |
| 2014/04 | | 0 | 0 | 0 | 0 |
| 2014/05 | | 0 | 1 | 1 | 1 |
| 2014/06 | | 0 | 1 | 1 | 1 |
| 2014/07 | | 0 | 0 | 0 | 0 |
| 2014/08 | | 0 | 1 | 1 | 1 |
| 2014/09 | | 0 | 0 | 0 | 0 |
| 2014/10 | | 0 | 0 | 0 | 0 |
| 2014/11 | | 0 | 0 | 0 | 0 |
| 2014/12 | | 0 | 1 | 1 | 1 |
| 2015/01 | 外挿期間 | 0 | 0 | 0 | 0 |
| 2015/02 | | 0 | 0 | 0 | 0 |
| 2015/03 | | 0 | 1 | 1 | 1 |
| 2015/04 | | 0 | 0 | 0 | 0 |
| 2015/05 | | 0 | 0 | 0 | 0 |
| 2015/06 | | 0 | 0 | 0 | 0 |
| 2015/07 | | 0 | 0 | 0 | 0 |
| 2015/08 | | 0 | 0 | 0 | 0 |
| 2015/09 | | 0 | 0 | 0 | 0 |
| 2015/10 | | 0 | 0 | 0 | 0 |
| 2015/11 | | 0 | 0 | 0 | 0 |
| 2015/12 | | 0 | 1 | 1 | 1 |
| 2016/01 | 外挿期間 | 0 | 0 | 0 | 0 |
| 2016/02 | | 0 | 1 | 1 | 1 |
| 2016/03 | | 0 | 0 | 0 | 0 |
| 2016/04 | | 0 | 1 | 1 | 1 |
| 2016/05 | | 0 | 1 | 1 | 1 |
| 2016/06 | | 0 | 0 | 0 | 0 |
| 2016/07 | | 0 | 0 | 0 | 0 |
| 2016/08 | | 0 | 0 | 0 | 0 |
| 2016/09 | | 0 | 0 | 0 | 0 |
| 2016/10 | | 0 | 0 | 0 | 0 |
| 2016/11 | | 0 | 0 | 0 | 0 |
| 2016/12 | | 0 | 0 | 0 | 0 |
| 2017/01 | 外挿期間 | 0 | 0 | 0 | 0 |
| 2017/02 | | 0 | 0 | 0 | 0 |
| 2017/03 | | 0 | 0 | 0 | 0 |
| 2017/04 | | 0 | 0 | 0 | 0 |
| 2017/05 | | 0 | 0 | 0 | 0 |
| 2017/06 | | 0 | 0 | 0 | 0 |
| 2017/07 | | 0 | 0 | 0 | 0 |
| 2017/08 | | 0 | 0 | 0 | 0 |
| 2017/09 | | 0 | 0 | 0 | 0 |
| 2017/10 | | 0 | 0 | 0 | 0 |
| 2017/11 | | 0 | 0 | 0 | 0 |
| 2017/12 | | 0 | 0 | 0 | 0 |
| 2018/01 | 予測期間 | 0 | 0 | 0 | 0 |
| 2018/02 | | 0 | 0 | 0 | 0 |
| 2018/03 | | 1 | 1 | 1 | 1 |
| 2018/04 | | 0 | 0 | 0 | 0 |
| 2018/05 | | 0 | 0 | 0 | 0 |
| 2018/06 | | 1 | 1 | 1 | 1 |
| 2018/07 | | 1 | 1 | 1 | 1 |
| 2018/08 | | 1 | 1 | 1 | 1 |
| 2018/09 | | 1 | 1 | 1 | 1 |
| 2018/10 | | 0 | 0 | 0 | 0 |
| 2018/11 | | 0 | 0 | 0 | 0 |
| 2018/12 | | 1 | 1 | 1 | 1 |
| 2019/01 | | | 1 | 1 | 0 |
| 2019/02 | | | 0 | 0 | 0 |
| 2019/03 | | | 1 | 1 | 0 |
| 2019/04 | | | 0 | 0 | 0 |

7. 景気判断のメカニズム

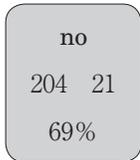
次に、景気判断のメカニズムをみるため、実績値のすべて（1985年2月から2017年11月まで）を使って景気判断のメカニズムを探った。図表8は、決定木分析の結果を図示したものである。決定木の見方は、以下の通り。分岐（ノード）にはそれぞれ番号が振ってある。図表8を上から順にみていく。なお、四角形の下に書かれている条件式に当てはまる場合（「yes」である場合）には左（景気後退期）に、当てはまらない（「no」である場合）には右（景気拡大期）に分類される。サンプル期間を1985年2月から2017年8月までとした決定木分析によると、景気動向指数の一致指数の累積DIの前期差である一致指数DIの数値が重要であることがわかる。一致指数DIがマイナス8.3未満の場合は景気後退期、マイナス8.3以上の場合は景気拡大期に分類される。さらに、景気拡大期と判断したサンプル（図表8の右半分の272サンプル）のうち、1986年12月より前は景気後退期となる。

また、景気後退期に分類されたサンプル（図表8の左半分の119サンプル）のうち、一致指数の構成系列の一つである有効求人倍率（除学卒）の前期差がゼロ未満の場合は景気後退期に分類され、ゼロ以上の場合は景気拡大期に分類される。さらに、景気拡大期に分類されたサンプルのうち、2013年6月以降のものはすべて景気拡大期となる。それより前の場合には、一致指数DIの前期差がマイナス47.3より小さい場合は景気拡大期、マイナス47.3以上の場合は景気後退期に分類される。

最後に述べた一致指数DIの前期差に関する条件の解釈は難しい。一致指数DIが大きく落ち込んだ時はむしろ景気拡大期となっている。これは、景気拡大期に起こる急激な落ち込みはその月特有のショックが原因の場合が多く、翌月反動で急上昇することがよくあるためであると解釈できる。2004年から2005年にかけては景気拡大期だが、一時的に急減して翌月急増している（図表9）。2004年10月、2005年2月、2005年7月などである。また、東日本大震災の発生した2011年3月もこの例に当てはまる。



(注) 分析期間は1985年2月から2017年8月までの391ヵ月。「累積DI.一致」は景気動向指数の一致指数の累積DI。「有効求人倍率.除」は一致指数の構成系列の一つである有効求人倍率(除学卒)。調査月は景気動向指数の調査月。「DI.一致」は景気動向指数の一致指数のDI。変数は調査月を除いて前期差。選択枝の分岐は条件式に対して左がyes、右がno。四角の中のupは景気拡大期、downは景気後退期。四角の中の数字は左が実際のデータにおける景気後退期の月数、右が景気拡大期の月数。%表示は全サンプルに占める割合。



- 1 行目：分類した項目 (yes か no か)
- 2 行目：は左側が正しく分類した数、右側が誤った数
- 3 行目：全サンプルのうち何%を分類したか。(204+21) / (228+98)

図表8 景気基準日付の決定プロセスを示す決定木

8. 2012年以前の判断は明快

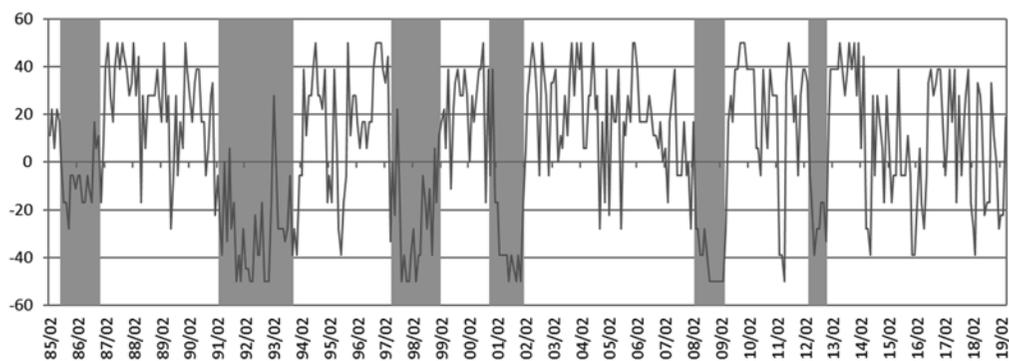
サンプル期間を変えて推計すると、消費税率上昇に伴う景気変動があった2014年以前の景気判断の手順はもっとシンプルだった。たとえば、前回の景気の谷である2012年11月までのデータを基にした決定木は2つの分岐点しかない（図表9）。それでも的中率は89.2%だ。

景気判断は一致指数DIの動きだけでほぼ判断でき、マイナス8.3以上の場合は景気拡大期、それ未満なら景気後退期だ。

決定木が複雑になっている背景には、2014年の消費税率上昇後の景気判断があると考えられる。この後に景気後退期があったと考えるエコノミストは多い。この部分を景気拡大期として判断しなかったため、条件が複雑になった可能性がある。

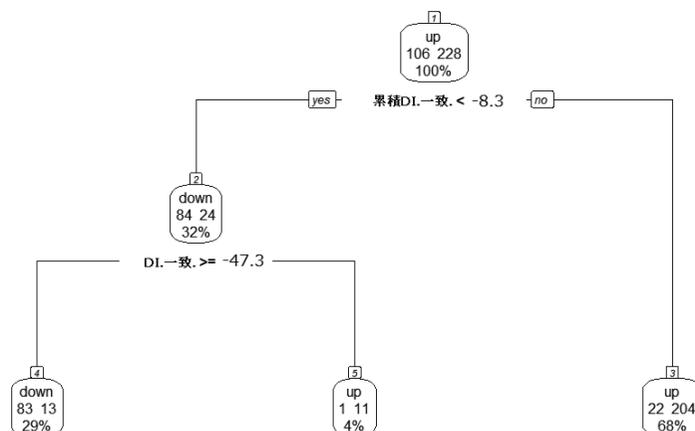
9. アンサンブル予測との比較

決定木の予測に加え、決定木を発展させたアンサンブル予測も行った。小高ほか（2018）に従って、ランダムフォレストと勾配ブースティングを使って予測した。ランダムフォレストは、木を一本だけ作るのではなく、サンプルや条件を変えて多数の木を作り、多数決で予測値を決める。勾配ブースティングは、次の学習モデルを前回の学習モデルの残差に適合させようとする手法である。ランダムフォレストは、独立した木を多数作るのに対し、勾配ブースティングは学習結果を活かしながら木を作成していくところに特徴がある。



（出所）内閣府「景気動向指数」

図表9 景気動向指数一致 DI



(注) データ期間は1985年2月から2012年11月までの334ヵ月。その他は図表1の注参照。

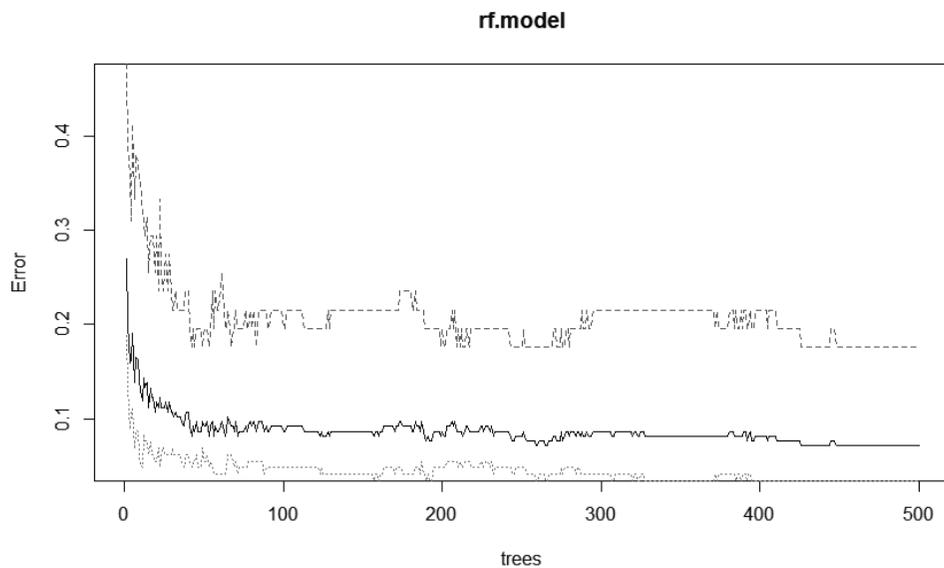
図表10 2012年11月までのデータに絞った場合の決定木

9.1 ランダムフォレストのハイパーパラメーター

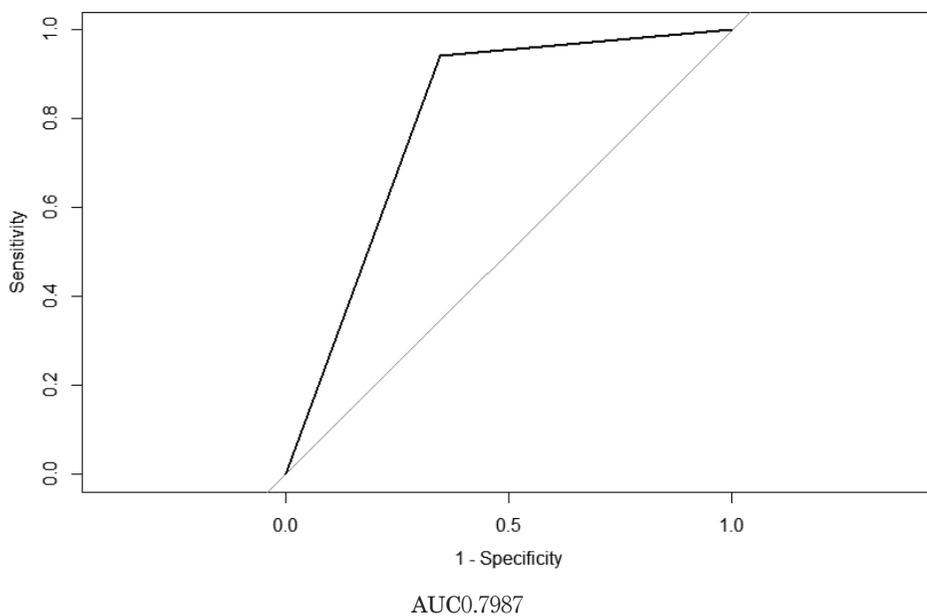
ランダムフォレストのハイパーパラメーターは、①作る木の本数②説明変数を何個使うか——である。図表11は縦軸は誤判別率、横軸は作る木の数を表したものだ。真ん中の曲線は予測値が実績値と異なった比率、上側の曲線は偽陰性比率（景気後退期を景気拡大期と予測する比率）、下側の曲線は偽陽性比率（景気拡大期を景気後退期と予測する比率）である（金 2017）。

randomForest パッケージでホールドアウトテストに関して誤差率を計算すると、100本程度木を作ることで誤差率が安定することがわかる。それ以降はほぼ横ばいなので、この分析では500回木を作ることにする。外挿や予測でも同様の傾向があるので、すべて500回木を作った。ホールドアウトテストのROC曲線を見ると、左上方に引っ張られており、AUCは0.7987と良好だった。

説明変数の数の選択には tuneRF 関数を使った。説明変数の数を変化させ、OOB (Out-Of-Bag) 誤差が最小になる点で選んだ。ホールドアウトテストは12変数、2012年3月までは3変数、2012年11月までは12変数、2017年8月までは6変数とした。ただ、予測値は6変数、12変数、36変数のいずれの場合も変わらない。

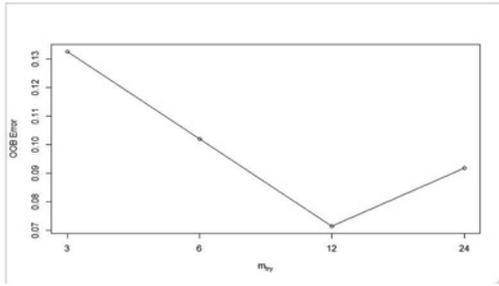


図表 11 木の本数と誤差率（ホールドアウトテスト）

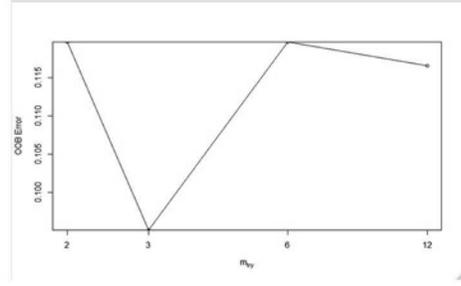


図表 12 ROC 曲線（ランダムフォレスト）

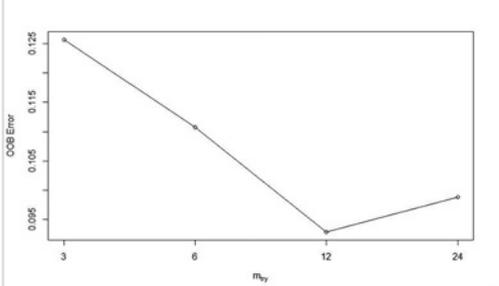
ホールドアウトテスト



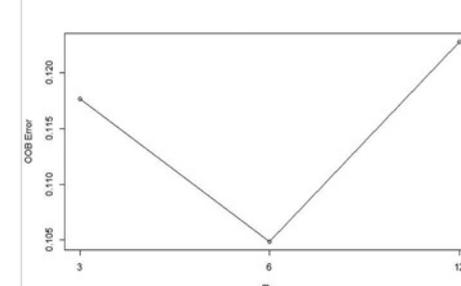
2012年3月まで



2012年11月まで

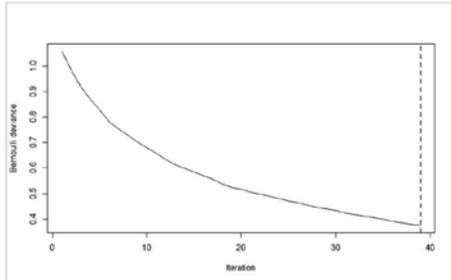


2017年8月まで

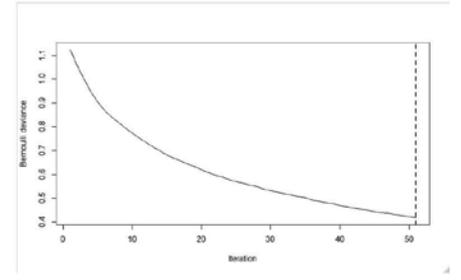


図表 13 説明変数の数

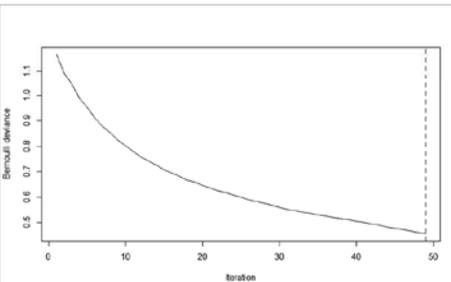
ホールドアウトテスト 39



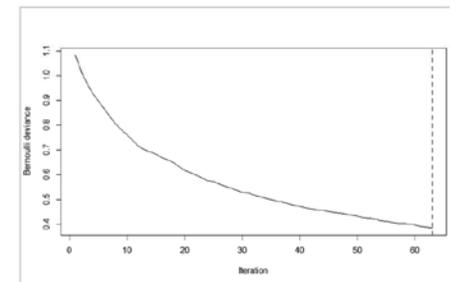
2012年3月まで 51



2012年11月まで 49

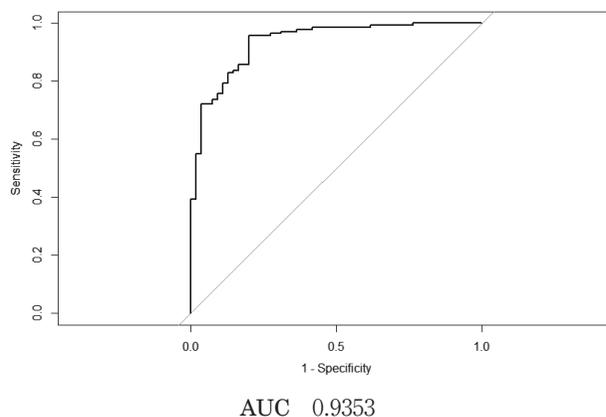


2017年8月まで 63



図表 14 木の数

機械学習による景気局面判断の分析



図表 15 ROC 曲線 ホールドアウトテスト

9.2 勾配ブースティングによるハイパーパラメーター

勾配ブースティングのハイパーパラメーターの決定には、`gbm` パッケージの `gbm.perf` 関数を使った。この関数を使うと縦軸が損失関数、横軸が木の個数が描かれ、最適な木の本数を決定できる。

ROC 曲線を描き、AOC をみると 0.9353 とサンプル内での精度は高い。

9.3 外挿と予測

外挿と予測のパフォーマンスをみると、決定木に比べて劣っていることがわかった。2012年3月までのデータで2012年4月以降を外挿すると、決定木ではほぼ景気後退期が予測できたが、ランダムフォレストでは、1ヵ月分誤りがふえている。勾配ブースティングでは、さらに誤りが増え、2012年4月からの景気後退期で正しく景気後退期と判定したのは2ヵ月分だけだった。

一般的にはアンサンブル予測の方が予測精度は上がるとされているが、このケースでは、決定木による予測の方が予測精度が高かった。

2017年9月以降の予測については、決定木予測では、景気後退を示すサインが多く出るが、ランダムフォレストでは全く出ず、勾配ブースティングでも少なかった。

図表 16 外挿と予測

| 日付 | データの種類 | 実績 | 決定木 | | | ランダムフォレスト | | | 勾配ブースティング | | |
|---------|-----------|----|-------------------------------|--------------------------------|-------------------------------|-------------------------------|--------------------------------|-------------------------------|-------------------------------|--------------------------------|-------------------------------|
| | | | 1: 2012 年4月 以降推 計 | 2: 2012 年12月 以降推 計 | 3: 2017 年9月 以降推 計 | 1: 2012 年4月 以降推 計 | 2: 2012 年12月 以降推 計 | 3: 2017 年9月 以降推 計 | 1: 2012 年4月 以降推 計 | 2: 2012 年12月 以降推 計 | 3: 2017 年9月 以降推 計 |
| 2012/04 | 学習 データ | 1 | 0 | | | 0 | | | 0 | | |
| 2012/05 | | 1 | 1 | | | 1 | | | 0 | | |
| 2012/06 | | 1 | 1 | | | 1 | | | 0 | | |
| 2012/07 | | 1 | 1 | | | 1 | | | 0 | | |
| 2012/08 | | 1 | 1 | | | 1 | | | 0 | | |
| 2012/09 | | 1 | 1 | | | 1 | | | 0 | | |
| 2012/10 | | 1 | 1 | | | 1 | | | 0 | | |
| 2012/11 | | 1 | 1 | | | 1 | | | 0 | | |
| 2012/12 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2013/01 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2013/02 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2013/03 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2013/04 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | | |
| 2013/05 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | | |
| 2013/06 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | | |
| 2013/07 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | | |
| 2013/08 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | | |
| 2013/09 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | | |
| 2013/10 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | | |
| 2013/11 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | | |
| 2013/12 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | | |
| 2014/01 | 外挿 期間 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2014/02 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2014/03 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2014/04 | | 0 | 0 | 0 | | 0 | 1 | | 1 | 1 | |
| 2014/05 | | 0 | 1 | 1 | | 0 | 0 | | 0 | 1 | |
| 2014/06 | | 0 | 1 | 1 | | 0 | 1 | | 0 | 1 | |
| 2014/07 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2014/08 | | 0 | 1 | 1 | | 0 | 0 | | 0 | 0 | |
| 2014/09 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2014/10 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2014/11 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2014/12 | | 0 | 1 | 1 | | 0 | 0 | | 0 | 0 | |
| 2015/01 | 外挿 期間 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/02 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/03 | | 0 | 1 | 1 | | 0 | 0 | | 0 | 0 | |
| 2015/04 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/05 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/06 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/07 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/08 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/09 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/10 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/11 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2015/12 | | 0 | 1 | 1 | | 0 | 0 | | 0 | 0 | |
| 2016/01 | 外挿 期間 | 0 | 0 | 0 | | 0 | 0 | | 0 | 1 | |
| 2016/02 | | 0 | 1 | 1 | | 0 | 0 | | 0 | 1 | |
| 2016/03 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2016/04 | | 0 | 1 | 1 | | 0 | 0 | | 0 | 0 | |
| 2016/05 | | 0 | 1 | 1 | | 0 | 0 | | 0 | 0 | |
| 2016/06 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2016/07 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2016/08 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2016/09 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2016/10 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2016/11 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2016/12 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/01 | 外挿 期間 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/02 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/03 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/04 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/05 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/06 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/07 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/08 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/09 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/10 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/11 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2017/12 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2018/01 | 予測 期間 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2018/02 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| 2018/03 | | 1 | 1 | 0 | | 0 | 0 | | 0 | 1 | 0 |
| 2018/04 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | 0 |
| 2018/05 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | 0 |
| 2018/06 | | 1 | 1 | 1 | | 0 | 0 | | 0 | 0 | 0 |
| 2018/07 | | 1 | 1 | 0 | | 0 | 0 | | 1 | 1 | 0 |
| 2018/08 | | 1 | 1 | 0 | | 0 | 0 | | 1 | 1 | 0 |
| 2018/09 | | 1 | 1 | 0 | | 0 | 0 | | 0 | 1 | 0 |
| 2018/10 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | 0 |
| 2018/11 | | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | 0 |
| 2018/12 | | 1 | 1 | 1 | | 0 | 0 | | 1 | 1 | 0 |
| 2019/01 | 1 | 1 | 0 | | 0 | 0 | | 0 | 1 | 0 | |
| 2019/02 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | 0 | |
| 2019/03 | 1 | 1 | 0 | | 0 | 0 | | 0 | 0 | 0 | |
| 2019/04 | 0 | 0 | 0 | | 0 | 0 | | 0 | 0 | 0 | |

(注) ゼロは景気拡大期、1は景気後退期。

10. おわりに

本稿では、決定木を使って景気局面判断を試みた。2012年3月までのデータを使って2012年4月以降の外挿予測をすると、良好な結果が得られた。

決定木の構造をみると、2012年3月まで、あるいは2012年11月までの木の構造は比較的単純だが、その後複雑になることがわかる。2014年4月に消費税率が引き上げられた、以前の判断基準では何度か景気後退のサインが出る。

次に決定木を発展させたランダムフォレストや勾配ブースティングを使って、外挿、予測を試みたが、決定木の方が予測パフォーマンスがよかった。ハイパーパラメーターの決め方によって予測パフォーマンスが変わることもあるため、条件を変えてチェックをすることが今後の課題である。

謝辞

本研究は科学研究費（基盤研究（C）『統計改革を反映したGDP予測—ビッグデータを利用したナウキャストリング』、研究課題／領域番号19K01680）の助成を受けたものである。

参考文献

- 有賀友紀・大橋俊介（2019）『RとPythonで学ぶ [実践的] データサイエンス&機械学習』技術評論社
- 浦沢聡士（2017）「構造変化の下での景気循環の動向：「定型化された事実（Stylized facts）」の再検証」
ESRI Discussion Paper No.341、内閣府経済社会総合研究所、2017年10月
- 大高一樹、菅和聖（2018）「機械学習による景気分析—「景気ウォッチャー調査」のテキストマイニング」
日本銀行ワーキングペーパーシリーズ No.18-J-8、2018年9月
- 金森敬文（2017）『Rによる機械学習入門』オーム社
- 金明哲（2017）『Rによるデータサイエンス（第2版）』森北出版
- 小寺信也・藤田隼平・井上祐介・新田克之（2018）「POS・テキストデータを用いた消費分析—機械学習を活用して—」経済財政分析ディスカッションペーパー DP/18-1、内閣府
- 小林雄一郎（2017）『Rによるやさしいテキストマイニング 機械学習編』オーム社
- 下川・杉本・後藤（2013）『樹木構造接近法』共立出版
- 内閣府経済社会総合研究所（2018）「第18回景気動向指数研究会議事概要」
- 長橋健吾（2017）『Rではじめる機械学習データサイズを抑えて軽量な環境で攻略法を探る』、インプレス
- 水門善之、坂地泰紀、和泉潔、島田尚、松島裕康（2019）『内閣府景気動向指数の先行系列に基づく機械学習を用いた短期経済予測』人工知能学会第34回社会におけるAI研究会、2019。

跡見学園女子大学マネジメント学部紀要 第31号 2021年2月25日

山本裕樹・松尾豊（2016）「景気ウォッチャー調査の深層学習を用いた金融レポートの指数化」第30回人工知能学会全国大会