

テキスト分析を用いた米景気後退確率の検討

経済調査室 石川 亮

要旨

- オルタナティブデータとは、政府統計の補完が可能なデータであり、速報性や詳細性、異質性などの特徴を持つ。オルタナティブデータは、経済を多面的に分析できる可能性があり、注目されている。
- オルタナティブデータの一つであるテキストデータ(本稿ではページブック)に対して、機械学習の手法を用いて景気後退確率の算出を行うモデルを作成したところ、実際の景気を適切に捉えたモデルを得ることができた。また、このモデルによれば、足元の米国の景気後退確率は10%程度と考えられる。

近年、経済分析分野において、オルタナティブデータの活用が進んでいる。特にコロナ禍では、政府統計よりも公表が速いという「速報性」が注目を集め、活用が大きく進んだ。オルタナティブデータは、そのほかにも、「詳細性」や「異質性」といった特徴を有するため、利活用を進めることで経済を多面的に分析することが可能となる。本稿では、伝統的な数値データと異なるという意味での「異質性」の一つであるテキストデータを用いて、米国の景気分析を行った。

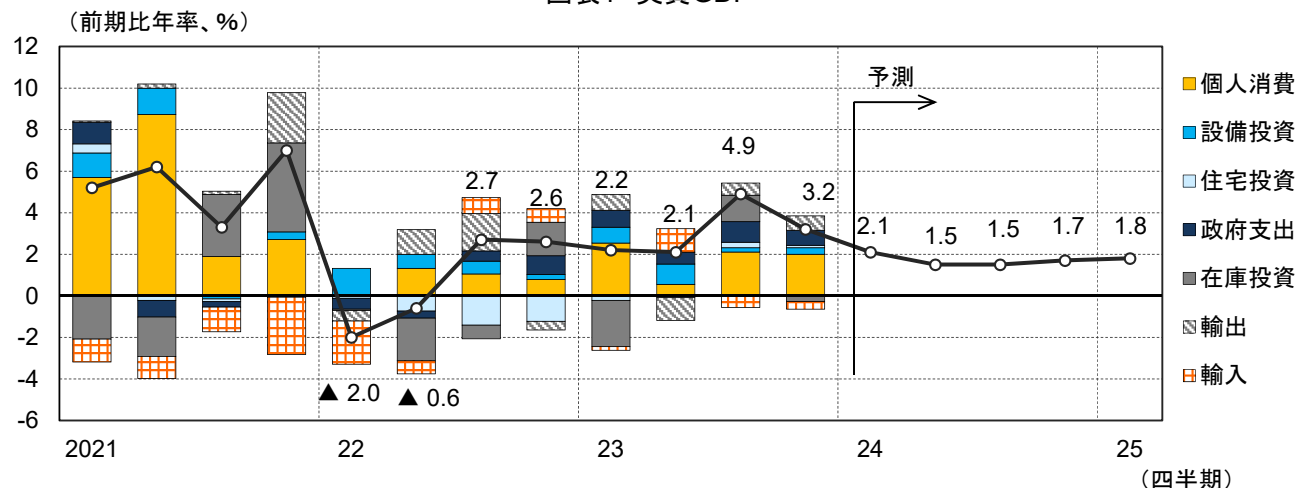
1. 堅調が続く米国経済

最初に、2020年以降の米国経済の状況を確認すると、コロナから景気が急回復する中でインフレが高進し、連邦準備制度理事会(FRB)は22年以

降、急速かつ大幅な利上げを行った。これによって景気が急減速するとの見方が広がり、22年夏にはエコノミストの半数が景気後退に陥ると予想した。ただし、22年後半以降の成長率は一貫して米国の潜在成長率とされる2%程度を上回り、予想に反して堅調を維持した。今後は、これまでの金利引き上げの影響などにより減速が見込まれているが、景気後退に陥るとの見方は少なく、軟着陸(ソフトランディング)可能であるとの予想が優勢となっている(図表1)。

こうしたシナリオには、どの程度の蓋然性があるだろうか。本稿では、通常の経済分析とは異なる「オルタナティブデータ」を用いた定量分析を複数行い、米国経済の景気後退確率の把握を試みる。

図表1 実質GDP



2.オルタナティブデータとは政府統計の補完が可能なデータ

分析に入る前に、オルタナティブデータについて簡単にまとめておく。オルタナティブデータとは、政府統計の補完が可能なデータとされる。最近では、日本経済学会や、内閣府の論文集である「経済分析」において、オルタナティブデータの特集が組まれるなど、注目が高まっている。

オルタナティブデータへの関心が高まっている理由は、オルタナティブデータが、①詳細性(細かい)、②速報性(公表が速い)、③異質性(新しいデータを扱う)などの特徴を持つためである。このうち、分析事例が多いのは、エコノミストや市場参加者などの関心に沿った②の速報性を生かした分析となっている。(図表2-1)。

オルタナティブデータ利活用の代表例に、人流を用いたサービス消費の分析がある。コロナ禍では、活動制限による飲食や宿泊などのサービス消費への影響が関心を集めたが、統計局のサービス産業売上高は公表が翌々月下旬と遅い。サービス消費と人流の動きはかなり似通っていることが確認されたことにより、携帯端末などから得られた人流データを用いて飲食、宿泊などのサービス消費を把握する試みが政府を含めて広くみられた(図表2-2)。

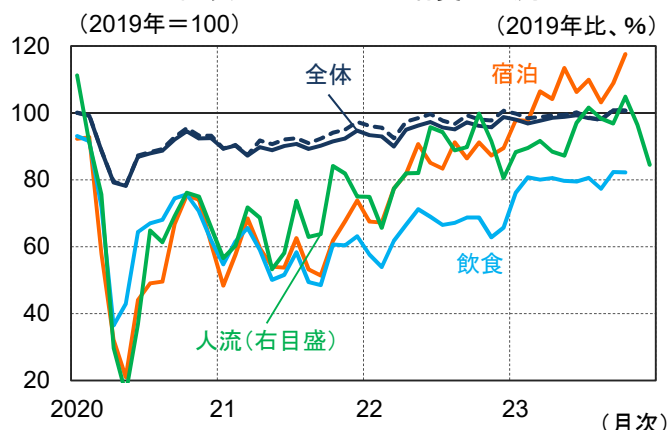
また、政府統計は基本的に数値データであるものの、テキストや画像など、数値以外の異質なデータ(③)を活用する試みも進んでいる。例えば、サンフランシスコ連銀は、新聞記事のテキスト分析を行い、日次の景気のセンチメント(景況感)指数を作成・公表している(図表2-3)。

図表2-1 オルタナティブデータの特徴と利用目的、分析例

データの特徴		分析に利用する目的	具体的な分析例
量が多い	①細かい(頻度/濃度)	政府統計よりも高頻度なデータや粒度の細かいデータを用いて、経済の状況をより詳細に把握	- 家計簿データを用いた定額給付金の影響分析(内閣府、2023) - クレジットカードデータを用いた小売売上高の予測(Dunn et al., 2020)
	②公表が速い	政府統計の公表前に経済の状況を把握	- 人流データを用いた景気分析(日本銀行、2020年4月「経済・物価情勢の展望」) - 新聞記事データを用いた景況感分析(サンフランシスコ連銀、2020)
質が異なる	③新しいデータを扱う	分析可能なデータの幅を広げ、多面的に経済の状況を把握	- 決算説明会の音声データを用いた景況感分析(橋本、2021) - 衛星画像を用いた新興国のGDP予測(Martinez, 2022)

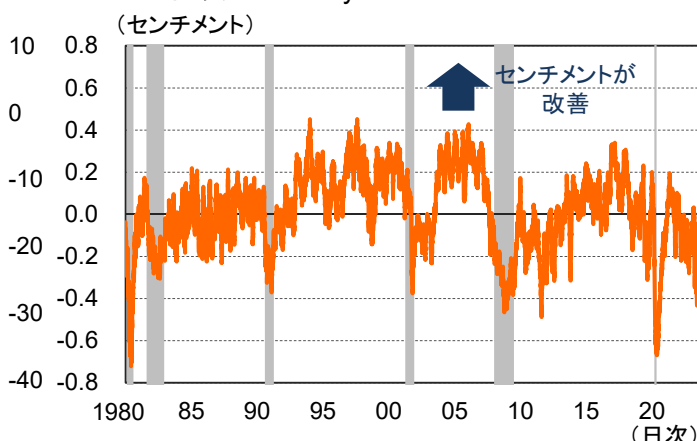
(備考) 日本政策投資銀行作成

図表2-2 サービス消費と人流



(備考) 1. 総務省、V-RESAS、株式会社Agoop「流動人口データ」により日本政策投資銀行作成 2. 破線は実質

図表2-3 Daily News Sentiment Index



(備考) 1. サンフランシスコ連銀、NBER(全米経済研究所) 2. 網掛けは景気後退期

3.テキストデータであるページブックを用いる

本稿では、オルタナティブデータの一つであるテキストデータを用いて、米国の景気分析を行った。分析データには、米FRBがFOMC(連邦公開市場委員会)の基礎資料として1ヵ月半ごとに公表するページブック(地区連銀景況報告)を用いた。これは12の地区連銀がそれぞれの地域の景況を文書形式でまとめたものであり(図表3-1)、その名前は表紙のページ色に由来する(図表3-2)。

4.単語の出現頻度から文書の特徴を確認可能

ページブックのテキストを単語レベルに分解し、その出現頻度を確認すると、例えば、トランプ政権

になって輸入関税が大きな問題となった2018年～19年は、「tariff(関税)」という単語の出現頻度が急増した。同様に、コロナ禍では、「covid」などのコロナ関連の単語の出現数が急増した。このように、単語の出現頻度に注目することで、その時々の特徴を掴むことができる(図表4-1)。

今回の分析の流れを整理すると、まずデータの収集・加工を行い、サンフランシスコ連銀の指数や先行研究と同様に、各回のページブックについてセンチメントを判定する。その後、文書間の内容の類似度を定量的に測定し、最後に、文書のテキストデータを用いて景気後退確率の試算を行う(図表4-2)。

図表3-1 ページブック(2023/11/29公表)の主な内容

全体として経済活動は減速した<総括> (前回報告(10/18公表)では横ばい)
今後6-12カ月の経済の見通しは低下した<総括> (前回報告では横ばい、もしくはわずかに軟化)
地域別にみると、経済は、4地域が緩やかに成長し、2地域がほぼ横ばい、6地域がわずかに縮小した<総括>
節約のため、高価な商品避けたり、ルームシェアを行う動きがみられた<カンザスシティ>
消費者は一層価格に敏感になっている<フィラデルフィア>
建設業では、高金利のために遅延、キャンセルとなった案件がある<セントルイス>
需要の減少などにより価格支配力が弱まった<クリーブランド>

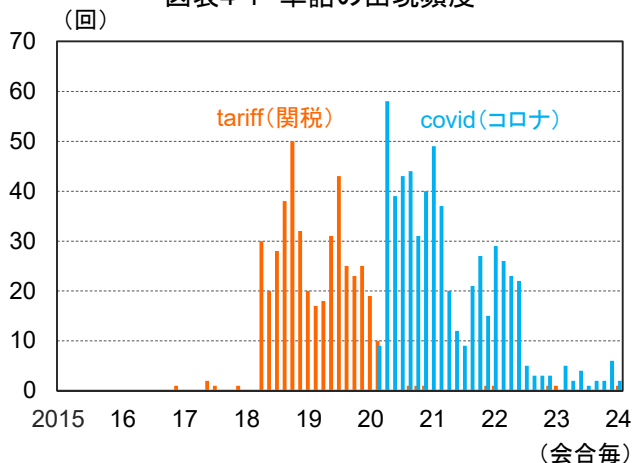
(備考) FRB資料により日本政策投資銀行作成

図表3-2 ページブック



(備考) ニューヨーク連銀

図表4-1 単語の出現頻度



(備考) 1. FRB資料により日本政策投資銀行作成
2. 「tariff」は少数ながら2015年以前もみられる

図表4-2 分析内容

データの収集・加工	- ページブックをPDF形式で入手 - 加工(文章を単語に分解するなど)
① センチメント分析	文書ごとに、センチメント(景況感)を計算
② 類似度分析	文書に出現する語句をもとに、文書と文書の類似度を計算
③ 景気後退確率の試算	機械学習の手法で、テキストデータから景気後退確率を試算

(備考) 日本政策投資銀行作成

5. センチメント分析：ページブックから景気変動を把握

ページブックに対してテキスト分析を活用した先行研究としては、Armesto et al. (2009)やSadique et al. (2013)が挙げられ、いずれも、ページブックの出現単語を、ポジティブかネガティブか、意味によって分類し、出現頻度の差をとることで、文書のセンチメント(景況感)を確認するセンチメント分析を行っている。センチメント分析は計算が比較的容易なため、サンフランシスコ連銀のセンチメント指数を含め、経済分析に用いられる例は比較的多い(図表5-1)。

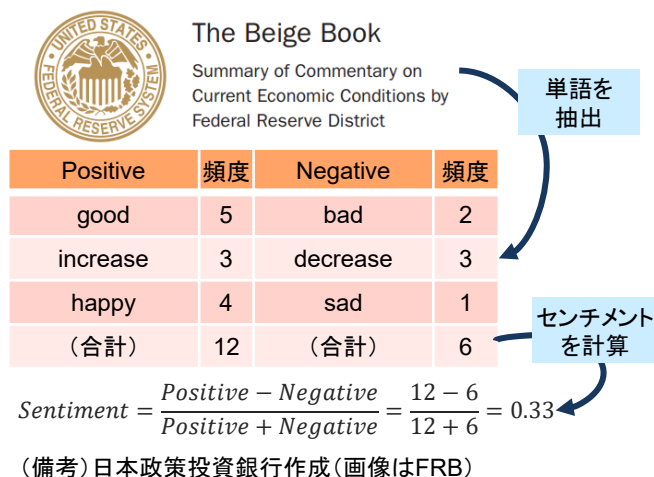
先行研究に倣って、ページブックのセンチメントを計算すると、代表的な景況指数であるISM製造

業景況指数にかなり近い動きを示し、景気後退局面においてセンチメントは悪化している。このことから、ページブックは、景気変動の把握に有用な情報を内包していることがわかる(図表5-2)。

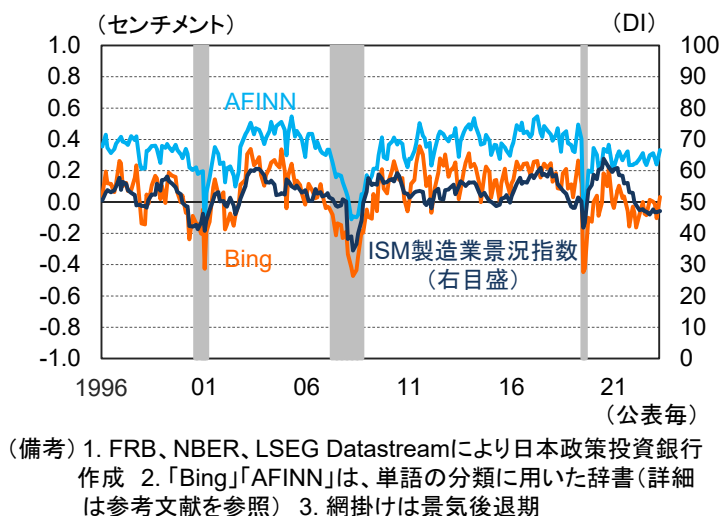
6. 類似度分析：景気後退時には文書の類似度が高い

センチメント分析は計算が容易な一方、分析対象が分類可能な単語に限られるといった欠点もあるため、文書全体のデータを用いて、文書同士の類似度を計算する。最も広く用いられるコサイン類似度は、文書ごとに単語の出現頻度を成分とするベクトルを作成し、ベクトル同士の近さ(コサイン値)を測定する(図表6-1)。

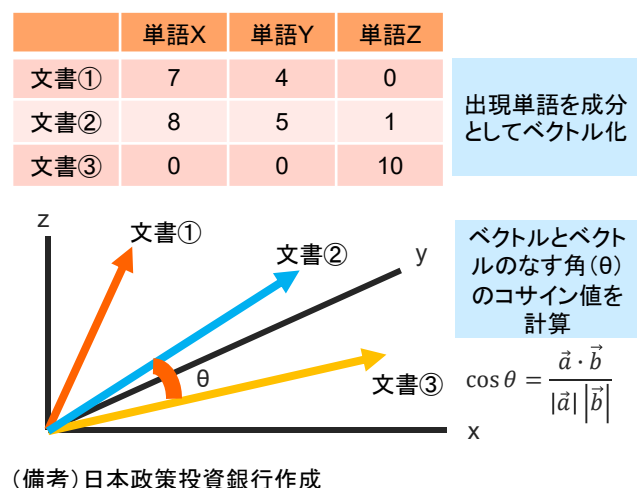
図表5-1 センチメント分析の進め方(例)



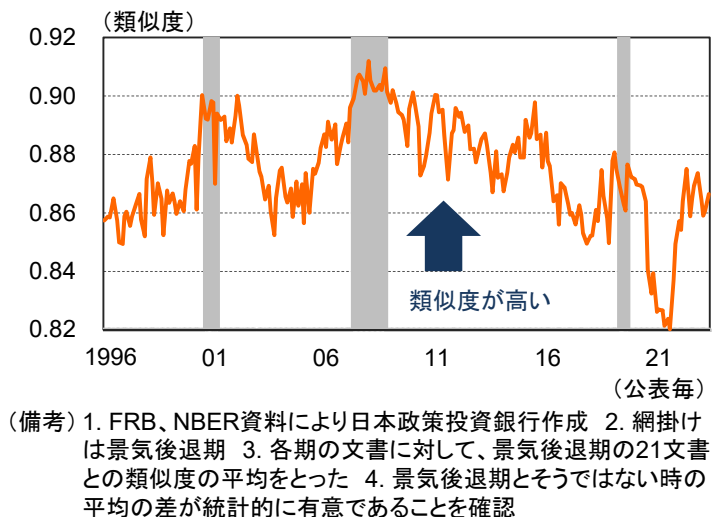
図表5-2 ページブックのセンチメント



図表6-1 類似度の計算手法(例)



図表6-2 ページブックの文書の類似度(景気後退期)



文章のベクトル化により意味を把握する手法は機械学習分野で一般的であり、検索エンジンやレコメンド機能、生成AI(ChatGPT)などに応用されているが、経済分析分野での適用例は少ない。

ここでの分析では、景気後退時の文書同士の類似度が高いことが確認でき、ページブックは、景気後退判定に利用できる可能性が示唆された(図表6-2)。

7.景気後退確率分析:足元の米国では10%程度

ページブックの文書同士の類似度をみることで、景気後退判定を実施できる可能性が示唆されたが、最後に、景気後退判定をより正確に行うため、ページブックにおける単語の出現頻度を説明変数(特微量)、景気後退かどうかを被説明変数として、ランダムフォレストの手法を用いて景気後退確率の推計を行った。ランダムフォレストは、Breiman(2001)によって提示された機械学習の手法であり、論文の引用数は12万回(24年3月時点)を超えており、内閣府(2018)の分析でも用いられるなど、幅広く利用される手法となっている。

結果をみると、過去の景気後退期に確率が上昇するモデルを得ることができた(図表7-1)。今回作成したモデルでは、足元の景気後退確率は10%程度にとどまると試算され、想定外に堅調とされる現在の米景気の認識を裏付けるものとなった。

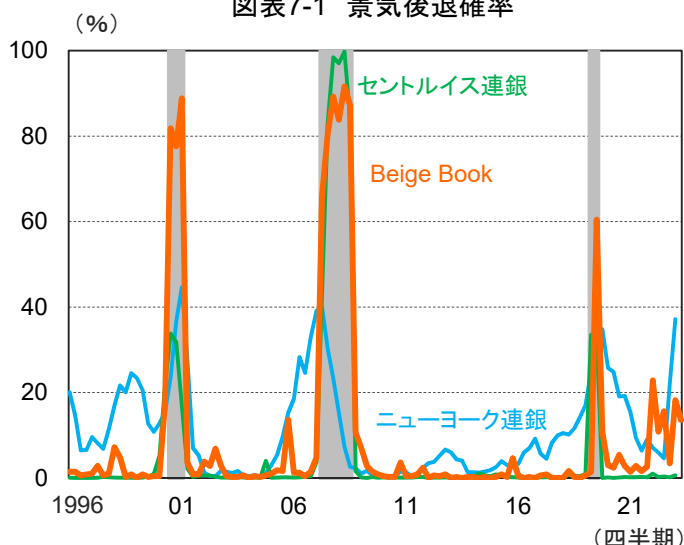
他のモデルと比較すると、ニューヨーク連銀のモデルは先行性に利点があるものの、精度の面で、①そもそも景気後退期の確率が低い、②足元でリーマンショック並みの景気後退確率にまで上昇してしまっている、などの課題がある。また、セントルイス連銀のモデルは、精度の面では信頼できるものの、公表がやや遅れるといった課題があり、本稿での分析は、先行性と精度の面でバランスの良いモデルといえる(図表7-2)。

8.重要単語を確認すれば景気後退の簡易的な予測が可能

テキスト分析では、どの単語が景気後退判定に重要視されたかを確認することができる。それを見ると、「weaken(弱まる)」や「reduce(減らす)」 「close(縮小する)」などの単語が最上位に来る。なお、単語の重要度は、景気後退ではない時に特徴的に出現頻度が増加する単語(「optimistic(楽観的な)」や「standard(普通の)」など)も上位に来ることには注意が必要である(図表8-1)。

上位三つの単語について出現頻度を確認すると、景気後退期に急増する傾向が確認できる。ページブックをみるにあたっては、本稿で用いた分析より簡便な方法として、これらの単語の出現頻度に注目すると、簡易的に景気判断を行うことも可能であると考えられる(図表8-2)。

図表7-1 景気後退確率



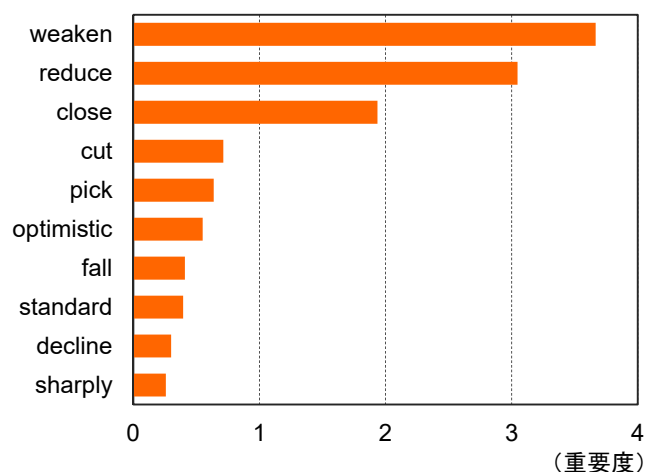
(備考) 1. FRB、ニューヨーク連銀、セントルイス連銀、NBER資料により日本政策投資銀行作成 2. 網掛けは景気後退期

図表7-2 景気後退確率モデルの比較

	ニューヨーク連銀のモデル	セントルイス連銀のモデル	今回のモデル
データ頻度	月次	月次	1ヵ月半
時期	月央に12ヵ月先が更新	月初に2ヵ月前のものが更新	FOMCの2週間前
使用データ	長短金利差	实体经济のデータ	Beige Bookのテキストデータ
先行性	◎	△	○
精度	△	○	○

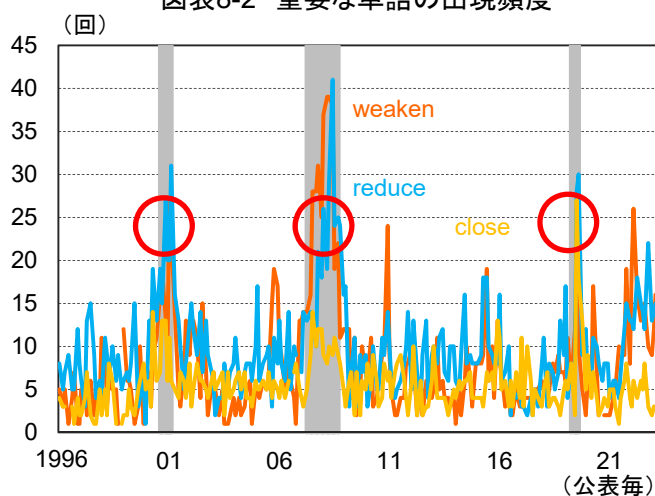
(備考) ニューヨーク連銀、セントルイス連銀資料により日本政策投資銀行作成

図表8-1 単語の重要度



(備考) 1. FRB、NBER資料により日本政策投資銀行作成
2. 上位10単語

図表8-2 重要な単語の出現頻度



(備考) 1. FRB、NBERにより日本政策投資銀行作成
2. 網掛けは景気後退期

参考文献

- Armesto, M. T., Hernández-Murillo, R., Owyang, M. T., & Piger, J. (2009). Measuring the information content of the beige book: A mixed data sampling approach. *Journal of Money, Credit and Banking*, 41(1), 35-55.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Buckman, S. R., Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). News sentiment in the time of COVID-19. *FRBSF Economic Letter*, 8(1), 5-10.
- Dunn, A., Hood, K., & Driessen, A. (2020). Measuring the effects of the COVID-19 pandemic on consumer spending using card transaction data. US Department of Commerce, Bureau of Economic Analysis.
- Liu, B. (2022). Sentiment analysis and opinion mining. Springer Nature.

- Martinez, L. R. (2022). How much should we trust the dictator's GDP growth estimates?. *Journal of Political Economy*, 130(10), 2731-2769.
- Nielsen, F. Å. (2011). Afinn. Richard Petersens Plads, Building, 321.
- Sadique, S., In, F., Veeraraghavan, M., & Wachtel, P. (2013). Soft information and economic activity: Evidence from the Beige Book. *Journal of Macroeconomics*, 37, 81-92.
- 内閣府 (2023) 特別定額給付金が家計消費に与えた影響—リアルタイムに記録される家計簿アプリデータを活用した分析—(政策課題分析シリーズ 22)
- 内閣府 (2018) 日本経済2017—2018 —成長力強化に向けた課題と展望—
- 橋本英樹 (2021) “国内企業の決算説明会の感情分析.” 第55 回 2021 年度夏季JAFEE大会発表論文

©Development Bank of Japan Inc.2024

本資料は情報提供のみを目的として作成されたものであり、取引などを勧誘するものではありません。
本資料は当行が信頼に足ると判断した情報に基づいて作成されていますが、当行はその正確性・確実性を保証するものではありません。本資料のご利用に際しましては、ご自身のご判断でなされますようお願い致します。本資料は著作物であり、著作権法に基づき保護されています。本資料の全文または一部を転載・複製する際は、著作権者の許諾が必要です。当行までご連絡下さい。著作権法の定めに従い引用・転載・複製する際には、必ず、『出所：日本政策投資銀行』と明記して下さい。

お問い合わせ先 株式会社日本政策投資銀行 産業調査部

Tel: 03-3244-1840

e-mail(産業調査部): report@dbj.jp