

計量テキスト分析による景気判断

—コーディングルールや主成分を使った時系列分析—

山澤成康¹

要旨

景気ウォッチャー調査の文章情報を使い、計量テキスト分析で景気動向の把握や予測法を検討した。「景気判断理由集（現状）」に使われる約 19 万件の文章から単語を抽出し、その単語が各月の総文章数に対してどれくらい出現するか（出現率）を集計して時系列データとして使用した。分析は、（1）コーディングルールを使った分析（2）相関分析（3）主成分分析（4）GDP 予測への応用——に分かれている。

コーディングルールを使った分析では、分析者が作成した単語の組み合わせ（コーディングルール）に従って出現率を計算して、グラフ化した。

相関分析では、景気ウォッチャー調査の現状判断 DI と各単語の出現率の相関係数をとり、どのような単語の相関係数が高いかを調べた。景気に順相関あるいは逆相関する単語を選び、景気指標を作成した。

主成分分析では、頻出 150 語の出現率を時系列データとみなし、主成分を抽出した。ウェートの高い語やその語と同時に使用される語などを検討して、各主成分がどのような性質を持っているのかを検討した。第 1 主成分が、景気ウォッチャー調査の現状判断 DI や景気動向指数・一致指数との相関が高いことがわかった。

GDP 予測への応用では、近似ダイナミックファクターモデルなどを使用して、実質 GDP 成長率が予測できるかどうかを検討した。まず、単語の出現率を説明変数として、どのような語が GDP の予測に役立つのかを検証した。次に主成分を使い、予測した。

J E L 分類番号：E 3 2

キーワード：景気循環、テキスト分析、景気ウォッチャー調査

¹ 本稿の執筆に際し、E S R I セミナーの討論者である松林洋一神戸大学大学院経済学研究科、また西崎文平所長をはじめ出席者の皆様より有益な指摘を頂いた。なお、本稿で示された内容や見解は筆者個人によるものであり、所属する機関のものではない。ありうべき誤りは筆者個人の責に帰するものである。