



データを分析する際に重要なこと —カレーが売れると株価は上がるのか—^{*1}

神戸大学 経済経営研究所
准教授 柴本 昌彦

最近、妻があるテレビ番組を見ていた際に、耳を疑うような出来事があった。『カレーが売れると株価が上昇する』と専門家が紹介していたというのだ。「仕事が充実していて忙しい景気上昇時は、料理する暇がない為カレーを食べている」「カレーはポジティブな気分させるので、ポジティブな気分が株価を高める・景気を良くする」といったことをデータが支持している、というのである^{*2}。

カレーが大好物であり日本のマクロ経済に関する実証研究をしている私としては、大変興味深い仮説だ。早速、実際にデータを入手して仮説を検証することにした^{*3}。図1は1992年1月から2004年12月までの株価(東証株価指数、TOPIX)とカレールウ支出金額の時間推移を示したものである。データを見ると、1992年から2004年にかけてカレーの売れ行きは「持続的に」低下傾向にあり、株価も「持続的に」伸び悩んでいる様子が伺える。また相関係数という統計指標を計算すると0.7という高い数値を示しており、テレビ番組内で紹介されていた結果と整合的であった。

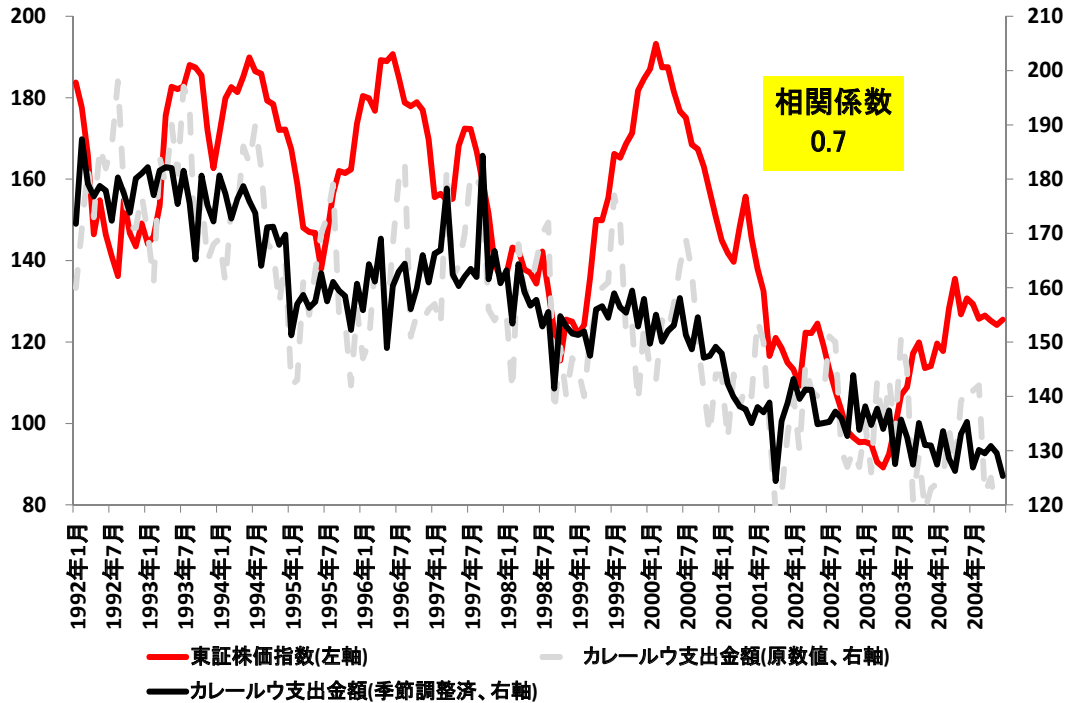
では、カレールウ支出金額と東証株価指数という2つの時系列データを使って『カレーが売れると株価が上昇する』と結論付けることができるのだろうか。もし分析の詳細にご関心があれば、計量経済学や時系列分析の教科書を読んでもらうなり学部・大学院の講義を受講してもらうなりで勉強してフォローして頂ければと思うが、本コラムでは今回の分析において特に重要な論点であると思われる『見せかけの回帰』について簡単にご紹介したい。

^{*1} 本コラムは、テレビ番組内で紹介されていた研究成果を否定することが目的ではなく、あくまで応用計量経済学者が日々行っていることや応用計量経済分析において心掛けるべき点を一般読者の方々へ紹介することを意図したものである。また、本コラムの分析は応用計量経済学・時系列分析の視点のみに基づいた私一人の見解によるものであり、市場エコノミストの方々や行動経済学をご専門にされている方々のこれまでの知見を考慮できていない。本コラムは、当該分野の更なる発展を行う上で、計量経済学・時系列分析が有用な分析ツールである可能性を指摘することを意図したものとしてお汲み取り頂きたい。

^{*2} テレビに出ていた専門家の解説を聞く限り、異なる2つの因果関係が混在していると考えられる。しかしながら、本コラムでは後者の因果関係に焦点を当てて分析を行う。

^{*3} 『家計調査』(二人以上の世帯、全国)に収録されているカレールウの支出金額(円)。東証株価指数(TOPIX)はNikkei Needs Financial Questより。カレールウ支出金額はX-12ARIMAを用いて季節調整を行った。

図1 株価とカレーウ支出金額の推移



そもそも回帰とは何か。(経済学に基づく)因果関係を統計的に検証する場合、我々応用計量経済学者は仮説に基づいて実証モデルを構築する。具体的には、『カレーが売れると株価が上昇する』という因果関係を統計的に調べるために、以下のような回帰モデルを考える。

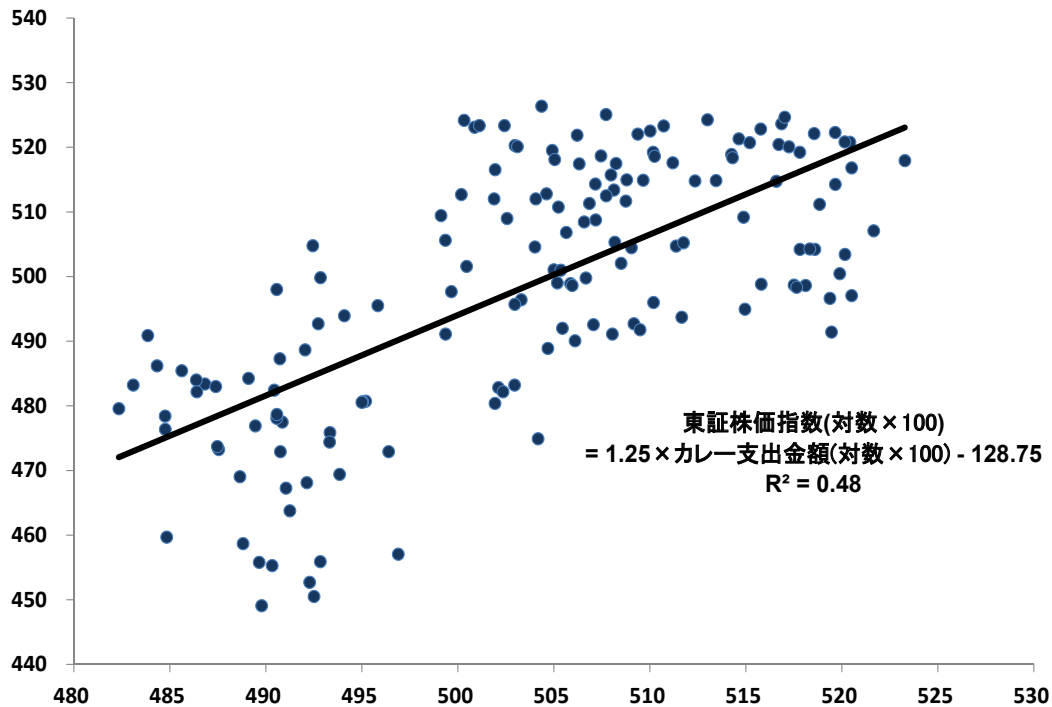
$$\text{東証株価指数 (対数)} = \alpha + \beta \text{カレーウ支出金額 (対数)} + \text{誤差} \quad (1)$$

これは「1% カレーウの支出金額が増えると株価が $\beta\%$ 上昇する」ということを意味する式である。 β の推定値がゼロとは異なり正の値が得られるかどうかを統計的に分析することで、『カレーが売れると株価が上昇する』という因果関係を統計的に示唆しているかどうか検証することができる。

データを一見すると、『カレーが売れると株価が上昇する』という因果関係を統計的に支持しているようにも見える。最小 2 乗推定法 (OLS) という手法を用いて (1) 式の回帰を行うと、 β の推定値は約 1.25 という結果が得られる。このことは、1% カレーウの支出金額が増えると株価が 1.25% 上昇するということの意味する。図 2 は統計ソフト Excel を用いて横軸をカレーウ支出金額、縦軸を株価にした際の散布図を作成したもので、黒線は散布図の近似線というものを挿入したものである。散布図から何らかの正の関係があるように見るとともに、近似線の数値を出力すると 1.25 という推定値が出てくる。こ

これは、その近似線の傾きであり、OLS という手法を用いた時の β の推定値を表している。

図 2 株価とカレー支出金額の散布図



ただし、私の分析によると、カレーの支出金額と東証株価指数という 2 つの時系列データから『カレーが売れると株価が上昇する』という因果関係を統計的に示唆していると結論付けることはできない。というのも、「全く背後にそんな因果関係が存在しない」にも関わらず、「単なる偶然の産物として」カレー支出金額と株価との間に 0.7 という高い相関係数なり 1.25 という β の推定値が得られた、と分析結果から示唆されるからである。

ここまでの相関係数や散布図を利用した分析には非常に重要なステップが抜け落ちており、これらの統計手法の欠点を見逃して因果関係を評価してしまっている。我々が現実の時系列データの分析を行う場合、まずは「使用するデータの特性」を確認することから始める。分析結果は表 1 をご参照頂きたいが、カレー支出金額と株価というのは単位根過程という時系列データ特性を持っている、ということを経験的に支持している。単位根過程に従っている 2 つの系列に関して (1) 式のような線形回帰モデルを OLS を用いて推定すると、現実にはそんな因果関係は全く存在しないにもかかわらず、あたかも統計学的に意味がある推定値であるかのような回帰結果が『見せかけ』として得られる『見せか

けの回帰』という問題が生じる可能性があることが広く知られている*4。表2の上部の結果は、『見せかけの回帰』が生じる可能性がある、ということを経済的に示唆している。更に、『見せかけの回帰』の可能性を考慮に入れて β の推定法を変えた分析結果(表2のVARとDGLS)によると、統計的には『カレーが売れると株価が上昇する』という因果関係を全く支持しない。まとめると、カレールウ支出金額と株価がそれぞれ単位根過程というデータの特性をもっているせいで、『カレーが売れると株価が上昇する』という因果関係が現実には存在するとは統計的には全く言えないにもかかわらず、『見せかけ』として $\hat{\beta} = 1.25$ という推定値が得られてしまった、というわけである。

表1 単位根検定

	ADF
カレールウ支出金額(季調済、対数)	-0.56
東証株価指数(対数)	-2.05

注: ADF は定数項を含んだ Augmented Dickey-Fuller (1979) 検定統計量を表す。ラグ数は4期を想定。サンプル期間は1992年1月から2004年12月まで。

表2 共和分検定と β の推定値

	EG	JOH		
検定統計量	-2.84	11.52		
	OLS	VAR	DGLS	
$\hat{\beta}$	1.25	-0.06	0.44	
	(0.19)	(0.10)	(0.41)	

注: EG は(1)式を最小2乗推定した際の残差に Augmented Dickey-Fuller テストを応用した際の検定統計量を表す。JOH は Johansen の共和分検定(maximal eigenvalue テスト) 統計量を表す。 $\hat{\beta}$ は(1)式を推定した際の β の推定値を掲載している。OLS は最小2乗推定値を表す。VAR はコントロール変数としてカレールウ支出金額(対数)及び東証株価指数(対数)の1期から4期までのラグ値を加えた際の最小2乗推定値を表す。DGLS は Choi, Hu, and Ogaki (2008) によるダイナミック GLS 法(リードとラグ次数は4期を想定)を用いた推定値を表す。Newey-West の標準誤差(ラグ数は4期に設定)が計算され、括弧内に示されている。サンプル期間は1992年1月から2004年12月まで。

*4 単位根過程に従う時系列データには確率的に変化するトレンドが含まれている。そのために、全く因果関係がないにも拘らず「たまたま」相関関係が生じてしまう、ということがある。

統計手法にはそれぞれ利点や欠点があり、それを正しく理解しないと誤った結論を導いてしまう。カレールウ支出金額と株価の時系列データを用いた分析例による私からの教訓とは、「データを用いて因果関係の検証を行う際、統計手法には利点と欠点・限界があり、それを正しく理解した上で使用・分析しなければならない」ということである。私の知る限り、どんな因果関係でもどんなデータでも‘ある’統計手法さえ使えば分析可能だ、といった万能の手法は存在していない。「経済学理論や自分自身の関心に基づき仮説を考える」「時系列データの統計的な特性を理解する」「仮説や時系列データ特性と照らし合わせて回帰モデルの構築や推定法の選択を行う」「分析結果を吟味する」というプロセスが、実証分析を用いて経済現象を明らかにする上で極めて重要なのである。

近年、現実のデータを利用したエビデンスに基づく評価の重要性が強調されるとともに、データ分析する環境も飛躍的に向上している。データをインターネット等で入手しやすくなったことに加え、Excelをはじめとする統計ソフトも次々と開発され、一般の人や学部生・大学院生も応用しやすい形でサービスが提供されている。データ分析の重要性が認識され始めたことに加え、データ分析する環境が飛躍的に向上したことは、我々研究者にとっても大いに歓迎すべきことである。

ただし、統計手法の正しい理解をせずに使用したせいで、「真実を明らかにする」という本来の目的を達成できていない、というケースも増えてきているのではないだろうか。「データさえ手に入れば、統計手法や理論はよく分からないけれども統計ソフトを使って分析結果は出てくるので、その結果に対して都合の良いように何らかの解釈を加えた」といった印象の研究も散見される。高度かつ複雑な統計手法を駆使することで何らかの結果が得られたけれども、手法を理解していないせいで分析者自身が分析結果を咀嚼できていない、分析者の関心とは全く別のものを分析している、といった研究に遭遇する場合もある。「データは入手できた」「統計ソフトも使える」けれども肝心の「現実の因果関係の検証」にはなっていない、というような研究や分析結果に基づく主張が近年増えている、という印象をもっているのはおそらく私だけではないだろう。

「データを使ったエビデンスに基づく評価の重要性が高まっている」「データ分析を比較的容易に行える環境になった」今だからこそ、現実データを使って誤った結論や都合の良い考えに誘導されてしまう、といったことにもなりかねない。新聞、テレビ、インターネット等で報道されているニュースを見ると、専門家と呼ばれている人ですら正しい統計処理や分析を行わずに（もしくは、知らずに）主張を展開し、それを大々的に報道しているのではないかと感じることもある。本や教科書なり学部・大学院で正しいデータ分析の作法を学ぶことで、専門家の意見を鵜呑みにすることなく自分自身の力で社会現象を理解する、といったことが、今後益々大事になってくるのではないだろうか。