経済予測に使う統計ソフト

経済予測に使うソフトとして、Microsoft Excel(マイクロソフト・エクセル)、エクセル統計、EViews(イーヴュウズ)を紹介する。3つのソフトを使って回帰分析による予測法を説明する。経済予測の基本的な考え方をマスターした後は、計量経済学や統計学の知識を深めて、さまざまな予測に挑戦してほしい。

山澤 成康 Yamasawa Nariyasu

跡見学園女子大学マネジメント学部教授

1 ディズニーランド入場者を予測

本稿では、東京ディズニーランドの入場者数を予測することを試みる。オリエンタルランドは、毎年入場者数を発表しているが、2001年度以降は東京ディズニーランドと東京ディズニーシーの合計の入場者数しか発表していない。2001年度以降のディズニーランドのみの入場者数は現状ではわからないということだ。そこで、2001年度以降のディズニーランド入場者数の「予測」を統計ソフトを使って試みた。

統計解析ソフトとして紹介するのは、比較的安価で使いやすいものにした。マイクロソフト社のエクセル、エクセルにアドインとして組み込む形の社会情報サービス社の「エクセル統計」そして、米国IHS社のEViewsである。

2 回帰分析による予測

予測法としてよく使われるのが回帰分析である。経済変数を予測の前提となる変数である「説明変数」と説明変数に影響を受けて決まる「被説明変数」とに分ける。両者を最小二乗法などの統計的手法を使って関係づけ、説明変数を使って被説明変数を予測する。

ディズニーランドの入場者数の予測をする場合、ディズニーランドの入場者数が被説明変数で、それに影響を与える変数が説明変数だ。説明変数の候補としては、国内総生産(GDP)が考えられる。好景気で所得が多ければテーマパークに行く回数は増え、不景気で所得が少なくなればレジャー支出は抑制されるだろう。ディズニーランドの入場者数をY_t、実質GDPをX_tとして以下の式を想定する。e_tは誤差項である。

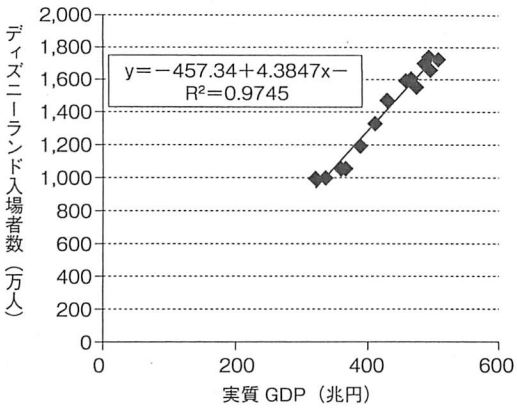
$$Y_t = a + bX_t + e_t$$

aやbは一定の数字となる係数で、最小二乗法で推計できる。最小二乗法は誤差の二乗の和が最も小さくなるように、aとbを選ぶ

著者紹介

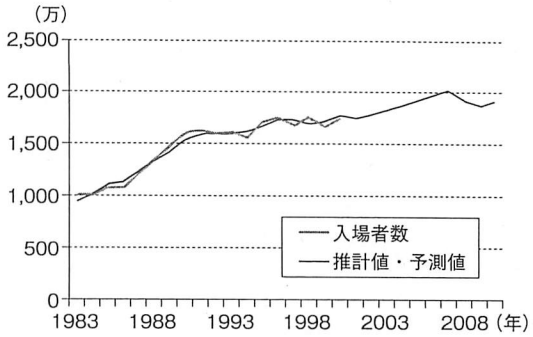
京都大学経済学部卒業。日本経済新聞社データバンク局経済情報部、スタンフォード大学客員研究員、日本経済研究センター短期・中期予測班総括などを経て現職。跡見学園女子大学大学院マネジメント研究科教授を兼務。主な著書・論文:『実戦計量経済学入門』日本評論社。「景気指標としての月例経済報告」『JCER Discussion Paper 124』日本経済研究センター。

図1　ディズニーランド入場者数の推計



出所）内閣府、オリエンタルランド

図2　ディズニーランド入場者数の予測



手法で、統計的に望ましい性質を持つことが知られている。推計期間を1983年度から2000年度として推計すると下記の式となる（係数は四捨五入）。

$$Y_t = -457 + 4.4X_t + e_t$$

マイナス457に、GDP（兆円単位）を4.4倍した値を加えた数がディズニーランドの入場者数（万人）となる。推計値を$\hat{Y}_t$とすると、以下の式が推計値を表す。上式から誤差項を除いたものである。

$$\hat{Y}_t = -457 + 4.4X_t$$

これを使うと将来の予測もできる。実質GDPの予測値があれば、ディズニーランドの入場者数の予測値が上式から計算できる。

説明変数の候補は、実質GDPの他にもある。実質民間最終消費や、実質雇用者報酬、実質賃金指数などである。これらの変数の中で、最も望ましいものを決めるためにさまざまな統計量を検討する。

推計値がどれだけ実績値に近いかをチェックするには決定係数（R^2：アールスクエアーとも呼ぶ）を使う。決定係数はゼロから1までの値をとり、推計値が実績値をうまく追えていれば決定係数は1に近い。

3　エクセルによる予測

エクセルを使ってこれまでの手順を実行することができる。X軸を実質GDP、Y軸をディズニーランド入場者数とし、散布図を描いてみよう（図1）。2つの系列を選び、グラフの種類を散布図とすれば簡単に描けるはずだ。

グラフのオプションに、「近似曲線の追加」がある。実質GDPと入場者数の関係を、直線、2次曲線、指数関数などさまざまな曲線に当てはめてくれる。散布図の散らばり具合をみると、点の配列は直線に近そうだ。そこで、「線形近似」を選ぶ。グラフ内の系列を選んだ状態で右クリックして、「近似曲線の追加」→「近似曲線のオプション」とし、線形近似を選ぶ。

同じボックスの下の方に「グラフに数式を表示する」と「グラフにR－2乗値を表示する」のチェックボックスがあり、これらをチェックすると、2つの変数の関係式と決定係数が表示される。

グラフ内に表示された式は最小二乗法で計算されたものだ。この式からX軸に対応するY軸の値が計算できる。

推計値や予測値を計算するには、この計算式（$-457.34+4.3847X_t$）に従って計算すればよい。エクセル上で簡単に計算できる。この

結果を示したのが、図２である。

2000年までは、実質GDPを使った推計値と実績値が描かれ、2001年度以降は実質GDPを使った予測値が示されている。先に行くほど予測誤差は大きくなるので年度予測なら通常２、３年先までだが、実質GDPが入手可能な2010年度まで予測値を描いてみた。

エクセルのグラフ機能を使って予測できるのは、「説明変数が１つだけ」の場合に限られる。また、決定係数以外の詳しい統計量は出てこないので、本格的な分析をするにはほかのソフトを使うことになる。

4　エクセル統計による予測

エクセル統計は、エクセルに組み込むアドインソフトである。通常のエクセルのメニューに「エクセル統計」というメニューが加わる。エクセルの画面を使いながら、最小二乗法の推計結果に関するさらに細かい情報を得ることができる。

エクセル統計では、説明変数を複数に増やすことができる。Y_tを変数X_tのみで推計（回帰）する場合を単回帰分析と呼び、複数の変数で推計する場合を重回帰分析と呼ぶ。エクセル統計を使えば重回帰分析が簡単にできる。

被説明変数Y_tを実質GDP（X_{1t}）と、夏の気温のデータ（X_{2t}）で推計してみよう。夏の気温のデータは、７～９月の平均気温の平年値からの差とした。式は以下の形となる。定数項aのほかに、説明変数にかかる係数b_1、b_2を最小二乗法で計算する。

$$Y_t = a + b_1X_{1t} + b_2X_{2t} + e_t$$

最小二乗法を用いれば係数は計算できる。ただX_{1t}や、X_{2t}でうまくY_tを説明できる場合もあれば、そうでない場合もある。式全体の当てはまり具合を見るには決定係数を使うが、X_{1t}や、X_{2t}がそれぞれ予測に役立っているかどうかを調べるには、t値を見る。t値は、「係数がゼロ」という帰無仮説をもとに計算され、t値がおおむね絶対値で２より大きければ、仮説は棄却され、「係数がゼロでない」ことがわかり、絶対値で２より小さければ、係数がゼロの可能性を否定できないことを表す。

また、必要な説明変数が入っているかどうかを調べるには、ダービン・ワトソン比が２に近いかどうかをチェックする。

エクセル統計の結果は、表１である。「回帰式の精度」で、決定係数をみる。決定係数は0.9768と高い。ダービン・ワトソン比は、1.78で、比較的２に近い。

「回帰式に含まれる変数」で個別の説明変数の当てはまり度合いがわかる。「偏回帰係数」と書いてあるところが係数なので、係数を取り出して式の形に直すと次式となる（係数は四捨五入してある）。

$$Y_t = -484 + 4.4X_{1t} - 13.3X_{2t} + e_t$$

重要なのは、X_1（実質GDP）やX_2（夏の気温）のt値である。X_1（実質GDP）のt値は24.5と２よりかなり大きく、係数がゼロである可能性は少ない。一方、X_2（夏の気温）のt値は－1.2で絶対値で２より小さい。

X_{2t}の符号はマイナスなので、「夏の気温が上がると入場者数が減る」という可能性を示している。しかし、それは統計的に有意ではなく、実際の予測には役立たないことを示している。

気象関係のデータは、夏の気温のほか春、秋、冬の気温のデータや、降水量を使ったデータがある。これらが予測に使えるかどうかは、重回帰分析をして、t値が有意になるかどうかをチェックすればよい。

表1　エクセル統計の推計値出力結果

重回帰分析							
結果							
回帰式の精度							
重相関係数		決定係数					
R	修正R	R2乗	修正R2乗	ﾀﾞｰﾋﾞﾝﾜﾄｿﾝ比	AIC		
0.9883	0.9868	0.9768	0.9737	1.7803	205.3097		
回帰式に含まれる変数（偏回帰係数・信頼区間等）							
				偏回帰係数の有意性の検定			
変　数	偏回帰係数	標準誤差	標準偏回帰係数	F　値	t　値	P　値	判　定
定数項	−483.9692	80.1205		36.4879	−6.0405	0.0000	**
x1	4.4442	0.1812	1.0006	601.4848	24.5252	0.0000	**
x2	−13.2503	10.8141	−0.0500	1.5013	−1.2253	0.2394	
回帰式の有意性（分散分析）							
要　因	平方和	自由度	平均平方	F　値	P　値		
回帰変動	1243060	2	621530	316.1023	0.0000		
誤差変動	29493	15	1966				
全体変動	1272554	17					

エクセル統計では、推計と同時に予測値を計算するオプションが付いているので、推計値や予測値を出力することが簡単にできる。

5　EViewsでの予測

EViewsは計量経済分析用に開発されたソフトで、さまざまな種類の推計を簡単に行うことができる。

データは、コマンドを使って読み込ませることもできるし、エクセルのデータをコピー・アンド・ペーストして持ってくることもできるので、便利である。

最小二乗法で推計する場合は、推計をする画面で図3のように、被説明変数、C（定数項）、説明変数の順に並べればよい。

y　c　$x1$　$x2$

また、log（y）とすれば対数で計算され、y（−1）とすれば、1期前のyの値が変数として使われる。こうした数式を使って、新たな変数を登録することもできる。「OK」ボタンを押すと、推計結果が出力される。

EViewsでは推計後、推計値と実績値のデータの出力はもちろん、グラフでの出力もできる。エクセルの項で紹介したグラフ（図2）を簡単な操作で作ることができる。

予測値は、「forecast」というボタンを押すことで作成される。実績値も予測値も、簡単にグラフで見ることができる。

6　ダミー変数

実証分析にはデータをよく見ることが重要だ。実績値と推計値グラフで比べると、気付くことがある。1988年12月に京葉線舞浜駅が開通したため、この頃からGDPの動きに比べて入場者が増えている。また、ディズニーランド開業5周年に当たる1988年や10周年に当たる1993年などで入園者が増えていることもわかる。

これらの要因も予測に取り込むことができる。基本的にはゼロが入り、対象となる事件

図3　EViewsの推計画面

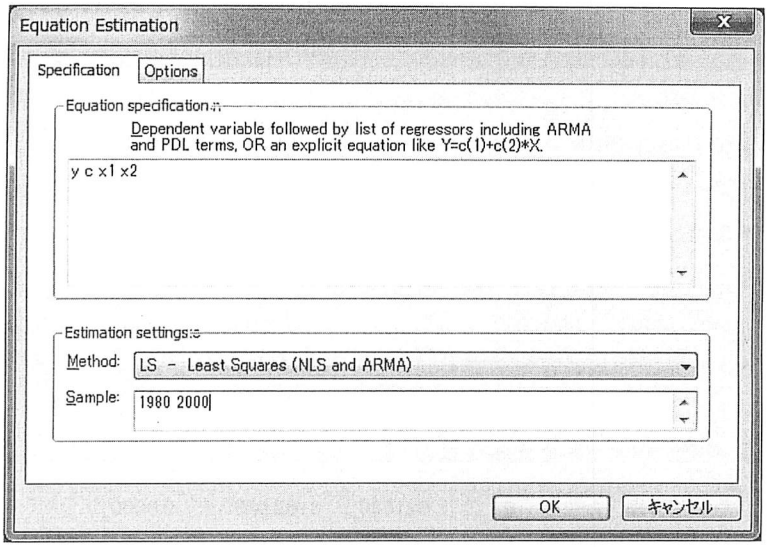
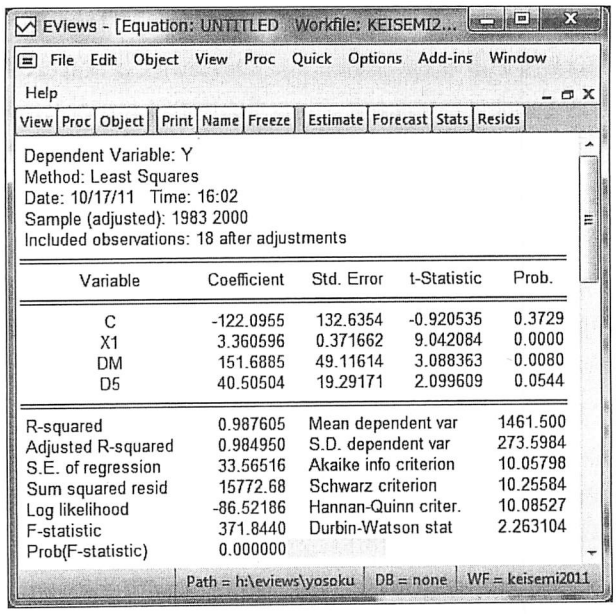


図4　EViewsによる推計結果



やイベントが起こった時だけ1となるような変数（ダミー変数）を説明変数として使う。

舞浜駅開業ダミー（DM）は、1987年度まではゼロ、1988年度は0.33（＝4カ月／12カ月）、1989年度以降は1という変数である。この変数は、舞浜駅が開業したことによる入場者数増を反映するはずだ。

また、5周年イベントダミー（D5）は、開園時の1983年を含めて、5年ごとに1、そのほかの年はゼロという変数である。

式の形は以下の通りである。

$$Y_t = a + b_1 X_{1t} + b_2\,\mathrm{DM}_t + b_3\,\mathrm{D5}_t + e_t$$

EViewsで推計する時は、以下のように入力する。

y　c　$x1$　DM　D5

その結果が図4で、下記の式を表している（係数は四捨五入）。

$$Y_t = -122+3.4X_{1t}+151.7DM_t+40.5D5_t+e_t$$

被説明変数（＝従属変数）は、Dpendent Variableに示されている。説明変数は、Variableに列記される。Cは定数項を表し、その下に説明変数が続く。

推計結果を見ると舞浜ダミーのt値は3.1、5周年イベントダミーのt値は2.1となり、両方とも有意である。

この推計式によれば、舞浜駅開業で、毎年152万人入場者が増え、5周年ごとのイベントで41万人入場者が増えることがわかる。

7　さらに進んで

経済予測の世界にはほかにもさまざまな推計法がある。EViewsで推計することができるものについて、紹介しておこう。

実際には何の関係もない変数どうしの相関が高そうに見える「見せかけの相関」の問題を回避するために、階差（＝前期差のこと）をとって推計することが多い。階差は、Δ（デルタ）と表す場合が多い。多くの経済系列は、階差をとると定常になると考えられているためである。

$$\Delta Y_t = a+b\Delta X_t+e_t$$

また、過去の誤差を修正するようにモデルが組まれるという、エラーコレクションモデルもよく使われる。

$$\Delta Y_t = a\Delta X_t-b\varepsilon_{t-1}+u_t \text{（短期的関係）}$$

$$\varepsilon_t = Y_t-c-dX_t \text{（長期的関係）}$$

長期的関係の式は、X_tを現実の値、$c+dX_t$を長期的な均衡状態の値と考えると、ε_tだけかい離があることを表している。短期的関係の式は、長期的な均衡値からのかい離ε_tを小さくするようΔY_tが修正されていくことを表している。このメカニズムが成り立つためには係数bは正（$-b$は負）でなければならない。

そのほか、複数の方程式を用いたマクロ計量経済モデルがある。

また、経済理論にとらわれずに、系列自身の性質を使って予測する手法に時系列モデルがある。単変数で代表的なのは自己回帰・移動平均（ARMA）モデルだ。それを複数の変数に応用した多変量自己回帰（VAR）モデルなどもある。

EViewsを使った予測については、飯塚・加藤（2006）が詳しく、分析については北岡・矢野・高橋（2008）が詳しい。山澤（2004）はさまざまな手法についてわかりやすく説明したものだ。経済予測一般の解説は山澤（2011）を参考にしてほしい。

参考文献

飯塚信夫・加藤久和（2006）『EViewsによる経済予測とシミュレーション入門』日本評論社

北岡孝義・矢野順治・高橋青天（2008）『EViewsで学ぶ実証分析入門　基礎編・応用編』日本評論社

山澤成康（2004）『実戦計量経済学入門』日本評論社

山澤成康（2011）『新しい経済予測論』日本評論社