第3章

変数を関係づける

3.1 はじめに

x と y という 2 つのデータだけが手元にあって、この 2 つの関係を調べたいとする。グラフを描けば 2 つの変数には同じような傾向があるかどうかなど、大まかなことはわかる。相関係数を計算すれば、どの程度関係があるかという分析もできる。しかし、それ以上の分析はできない。

回帰分析は、2 つのデータだけから、$y = a + bx$ という式を作り出す。この式が推計できれば、x が変化した時の y の変化が定量的に分析できるし、さまざまなケースについて係数である b の大きさを比べることができ、分析の幅が飛躍的に広がる。

しかも、係数 a、b はほかの情報から作るのではなく、データの x と y だけから計算できるという、「魔法」のような推計法だ。こうした推計は最小二乗法によって可能になる。

3.2 最小二乗法

最小二乗法（OLS, Ordinary Least Squares）とは線形関係を仮定した 2 つの変数を結び付ける係数を推計するものだ。x が所得、y が消費を表すとき、x と y の間にどのような関係があるかを統計的に導き出す。たとえば、次のような式である。

$$y_t = a + bx_t + e_t$$

y は x によって説明されているので、被説明変数（または従属変数）と呼び、y を説明している x のことを説明変数（または独立変数）と呼ぶ。a や b のことを係数（またはパラメーター）と呼ぶ。特に、a のように定数だけの係数は定数項と呼ぶ。y と x が係数によって完全に一致するとは限らない。計算式によって計算される推計値（$\hat{y}_t = a + bx_t$）と実際の値との差を残差と呼ぶ。残差は確率変数である誤差項の実現値だと考える。

3.2.1 母集団と標本の違い

系列 x,y（時系列データでもクロスセクションデータでも良いが、ここでは時系列データとして考える）があるとき、両変数が次の式で表されるとする。

$$y_t = \alpha + \beta x_t + u_t$$

α と β は、すべてのデータを使って計算できる真のパラメータを表しているとする。しかし我々は、経済統計としてごく少ない y_t と x_t しか知らない。そこで、計算によって α と β の推計値を求めようとする。その推計値を a、b と表すと、同じ形だが次のようになる。

$$y_t = a + bx_t + e_t$$

この 2 つの式の違いは重要である。この 2 つの違いがわからないと係数の期待値、分散などを求めるときに混乱することになる。たとえば、1980 年から 2000 年までの消費と所得のデータを入手しており、その情報を使ってパラメーター a や b を推計しようとする。これは真のパラメータを表しているわけではない。70 年代のデータは省いているし、実際の消費や所得は毎日行われている。本当に知る必要があるのは、過去にも将来にもわたって変わらないパラメーターである。

この式の意味について詳しく考えてみよう。推計したパラメータ a が 100、b が 0.5 とすると、次のように書ける。

$$y_t = 100 + 0.5x_t + e_t$$

この 0.5 という値は、正確に 0.5 というわけではない。年によっては 0.4 になったり 0.6 になったりする。平均的に 0.5 になることを表している。平均して 0.5 といっても、その意味する可能性はさまざまだ。-1 と 1.5 などの値を平均した 0.5 と、0.49 や 0.51 などを平均した 0.5 とでは 0.5 の意味が違うだろう。係数の平均である 0.5 という値とともに、その分布が重要になることがわかる。

3.2.2 最小二乗法の係数の求め方

最小二乗法は、α と β の推計値のうち、残差の二乗 (e^2) を最小にするものを推計する。直感的には、散布図による説明がわかりやすい。図 3.1 は、Y の値を縦軸、X の値を横軸にしてデータを示した散布図である。この図にある直線（$Y = a + bX$）を引い

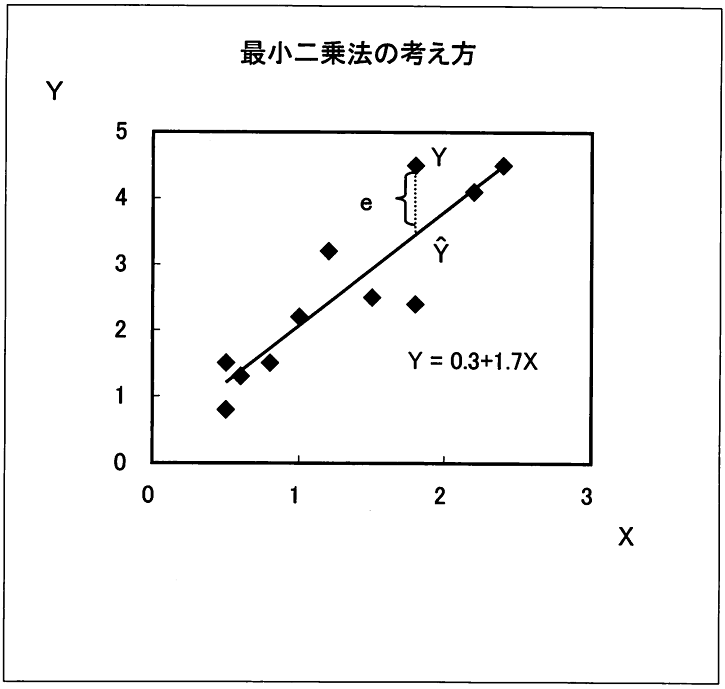

図 3.1: 最小二乗法の考え方

て、どの直線が最も両者の関係を表しているのかを考えることと同じである（この図では $Y = 0.3 + 1.7X$ になる)。推計値は直線上の点となり、$\hat{Y} = a + bX$ で計算できる。推計値と実際の Y の差が残差 (e) である。最小二乗法はすべてのサンプルの残差を二乗したものの和が最小になるように係数 a、b を決める方法である。

Y と X との関係を表す係数はさまざまなものが考えられ、それによって直線の傾きや水準が変わるが、最小二乗法による係数を使った推定量が一定の仮定のもとでは、最も望ましい性質の推定量となる（第 5 章参照)。a と b を求めるには、誤差の二乗を表す式を作り、a と b で偏微分してゼロと置く。誤差の二乗和のことを SSR(sum of squared residual) または RSS(residual sum of squares) と呼ぶ。SSR と RSS は同じことを指す。方程式のパフォーマンスを調べる際に SSR は非常に重要で、今後も SSR を使ってさまざまな統計量を計算することになる。具体的には次のように計算する。

$$\text{残差二乗和 }(SSR) = \sum e_t^2 = \sum (y_t - a - bx_t)^2$$

これの最小値を求める。SSR は a と b の関数なので、これを展開した後、SSR を a と b で偏微分した結果をゼロと置くと、SSR が最小になる条件（1 階の条件）が導き出される。

$$\frac{\partial SSR}{\partial a} = -2\sum(y - a - bx) = 0$$

$$\frac{\partial SSR}{\partial b} = -2\sum(y - a - bx)x = 0$$

この2つの方程式のことを正規方程式と呼ぶ。x_t と y_t はすでに入手しているデータなので、未知数は a、b の2つだ。2つの方程式から a と b を求めることができる。

説明変数が増えても、「誤差の二乗を a と b など係数を使って作り、その最小値を各係数に関して偏微分して求める」という方法は変わらない。

この方程式を解くと、a と b が次のように求められる。$\bar{y}$ は y_t の平均を表しており、y_t のサンプル数を n とすると、$\bar{y} = \Sigma y_t / n$ で求められる。$\bar{x}$ も同様である。

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\Sigma(x_t - \bar{x})(y_t - \bar{y})}{\Sigma(x_t - \bar{x})^2}$$

3.2.3 説明変数が2つある場合

説明変数が1つの場合の推計を「単回帰」、2つ以上の説明変数がある場合の推計を「重回帰」と呼ぶ。説明変数が2つ以上ある場合も同様の計算をすれば、係数を求めることができる。説明変数が2つある次の式を想定する。

$$y_t = a + bx_t + cz_t + e \tag{3.1}$$

係数の求め方は説明変数が1つの場合と同じで、(3.1) 式を誤差の2乗和の式に変形する。展開して、a、b、c それぞれについて偏微分すれば正規方程式が3つ得られ、a、b、c を求めることができる。ただ、この方法はかなり煩雑な計算となる。以下のように計算すると、計算過程が簡単になる。まず、誤差の2乗和についての式を作る。これを a、b、c それぞれについて展開すると煩雑になるので、a について偏微分することを目指し、a についてまとめるように展開する。

$$\Sigma e_t^2 = \Sigma(y_t - a - bx_t - cz_t)(y_t - a - bx_t - cz_t)$$

$$= na^2 - 2a(\Sigma y_t - b\Sigma x_t - c\Sigma z_t) + \Sigma(y_t - bx_t - cz_t)^2$$

a について偏微分してゼロと置く。これが正規方程式の一つとなる。

$$\frac{\partial \Sigma e_t^2}{\partial a} = 2na - 2(\Sigma y_t - b\Sigma x_t - c\Sigma z_t) = 0 \tag{3.2}$$

(3.2) 式を Σy_t について変形すると次の式が得られる。

$$\Sigma y_t = na + b\Sigma x_t + c\Sigma z_t = 0$$

両辺を n で割ることで、各変数の平均の関係式が導かれる。

$$\frac{\Sigma y_t}{n} = \frac{na}{n} + b\frac{\Sigma x_t}{n} + c\frac{\Sigma z_t}{n}$$

x の平均を $\bar{x}$ とするなどすれば、次の式が成り立つ。

$$\bar{y} = a + b\bar{x} + c\bar{z} \tag{3.3}$$

次に、式を簡略するためにこれを各変数の偏差を用いて計算を進める。$\tilde{x}_t = x_t - \bar{x}$ である。もとの式 $y_t = a + bx_t + cz_t$ から (3.3) 式を差し引くと、定数項 a がない次の式ができる。

$$(y_t - \bar{y}) = (a - a) + b(x_t - \bar{x}) + c(z_t - \bar{z}) + (e_t - 0)$$

これを偏差の形で表すと次の式となる。

$$\tilde{y} = b\tilde{x} + c\tilde{z} + e_t$$

この式について誤差の二乗を最小にするように、b、c を求めればよい。

$$\Sigma e_t^2 = \Sigma(\tilde{y}_t - b\tilde{x}_t - c\tilde{z}_t)^2$$

この式を展開して、b、c について偏微分してゼロと置く。そこから b、c を求めると次のようになる。

$$b = \frac{\Sigma\tilde{x}_t\tilde{y}_t\Sigma\tilde{z}_t^2 - \Sigma\tilde{y}_t\tilde{z}_t\Sigma\tilde{x}_t\tilde{z}_t}{\Sigma\tilde{x}_t^2\Sigma\tilde{z}_t^2 - (\Sigma\tilde{x}_t\tilde{z}_t)^2} \tag{3.4}$$

$$c = \frac{\Sigma\tilde{z}_t\tilde{y}_t\Sigma\tilde{x}_t^2 - \Sigma\tilde{y}_t\tilde{x}_t\Sigma\tilde{x}_t\tilde{z}_t}{\Sigma\tilde{x}_t^2\Sigma\tilde{z}_t^2 - (\Sigma\tilde{x}_t\tilde{z}_t)^2} \tag{3.5}$$

a は (3.3) 式を変形すれば求められる。

$$a = \bar{y} - b\bar{x} - c\bar{z}$$

3.2.4 一般的な計算法

説明変数がさらに増えた場合など、より一般的な最小二乗法の計算は、行列計算を使う。具体的な計算法は、15 章参照。

	数式表示（定数項を含む説明変数が2つの場合）	行列表示（定数項を含む説明変数が k 個の場合）
真の推計式	$y_t = \alpha + \beta x_t + \varepsilon_t$	$\mathbf{Y} = \beta\mathbf{X} + \varepsilon$
最小二乗法の推計式	$y_t = a + bx_t + e_t$	$\mathbf{Y} = \mathbf{bX} + \mathbf{e}$
推計値	$\hat{y}_t = a + bx_t$	$\hat{\mathbf{Y}} = \mathbf{bX}$
係数 a	$\bar{y} - b\bar{x}$	$(\mathbf{X'X})^{-1}\mathbf{X'Y}$ の1行目
係数 b	$\frac{\sum(y_t - \bar{y})(x_t - \bar{x})}{\sum(x - \bar{x})^2}$	$(\mathbf{X'X})^{-1}\mathbf{X'Y}$ の2行目以下
残差の和はゼロ（仮定）	$\sum e_t = 0$	
誤差と x_t は無相関（仮定）	$\sum x_t e_t = 0$	$\mathbf{X'e} = \mathbf{0}$
誤差と推計値は無相関	$\sum \hat{y}_t e_t = 0$	$\hat{\mathbf{Y}}'\mathbf{e} = \mathbf{0}$
残差の分散	$\sigma^2 = \frac{\sum e_t^2}{n-2}$	$\boldsymbol{\sigma}^2 = \frac{\mathbf{e'e}}{\mathbf{n}-\mathbf{k}}$
係数の分散	$\frac{\sigma^2}{\sum(x_t - \bar{x})^2}$	$\boldsymbol{\sigma}^2(\mathbf{X'X})^{-1}$

表 3.1: 最小二乗法の要点

3.2.5 最小二乗法の要点

最小二乗法の推計値がどのような特徴を持つかをまとめると表 3.1 となる。残差の和がゼロというのは、最小二乗法の仮定の一つである。定数項 a はこの仮定を満たすように決まる。誤差と x_t が無相関であることも重要な仮定で、これによって最小二乗法が統計として望ましい性質（不偏性 (第 5 章参照)）を持つことができる。

3.3 統計値の解説

この節では最小二乗法を統計ソフトウェアで推計した時に出力される統計値について解説する。統計値をまとめると表 3.2 となる。

3.3.1 回帰係数

回帰係数とは、推定された係数のことである。パラメーター、係数などとも呼ばれる。母集団からサンプルをとって推定したものなので、ある程度の広がりをもって分布していると考えられる。母集団のパラメーターの平均と差があるのかどうかを期待値をとって調べてみよう。x_t の平均を $\bar{x}$、平均からのかい離を $\tilde{x}_t(= x_t - \bar{x})$、$y_t$ の平均を $\bar{y}$、平均からのかい離を $\tilde{y}_t(= y_t - \bar{y})$ とすると、推定されたパラメーター b は次の式で表される。

用語	説明
回帰係数	係数。
回帰の標準誤差	回帰式の誤差の標準偏差。
係数の標準誤差	係数の標準偏差。
t 値	ある係数がゼロと違わないか、2 以上なら OK。
決定係数	当てはまりがどの程度良いか。1 が最も良い。
対数尤度	誤差が標準正規分布にどの程度近いか。大きいほど良い。
AIC	小さいほど良い。
SBIC	小さいほど良い。
ダービンワトソン比	誤差に系列相関があるかどうか。2 に近いほど良い。
F 値	全ての係数がゼロという仮説が棄却できるかどうか。大きいほど良い。

表 3.2: 最小二乗法の統計値のまとめ

$$b = \frac{\Sigma \tilde{x}_t \tilde{y}_t}{\Sigma \tilde{x}_t^2}$$

この式の $\tilde{y}_t$ に母集団の関係式を入れる。この操作は、回帰係数の期待値を調べる時に良く使う方法である。

$$b = \frac{\Sigma \tilde{x}_t (\beta \tilde{x}_t + u_t)}{\Sigma \tilde{x}_t^2}$$

式を変形していくと次の式となる。

$$b = \beta + \frac{\sum \tilde{x}_t u_t}{\sum \tilde{x}_t^2} \tag{3.6}$$

次に、b の期待値をとる。

$$E(b) = E(\beta) + E(\frac{\sum \tilde{x}_t u_t}{\sum \tilde{x}_t^2}) = \beta + \frac{\sum \tilde{x}_t E(u_t)}{\sum \tilde{x}_t^2} \tag{3.7}$$

最小二乗法の仮定により $E(u_t)$ がゼロなので $\tilde{x}_t$ が確率変数でなければ第 2 項はゼロである。つまり、b の期待値は β となる。

$$E(b) = \beta \tag{3.8}$$

最小二乗法の推計値の期待値は真のパラメーター β に等しいことを示している。

3.3.2 残差の分散

残差の分散は推計値がどの程度の幅で振れるのかを示す。残差の分散の平方根は、回帰の標準誤差と呼ばれる。標準誤差とは、「推計値の標準偏差」のことで、回帰の結果出てきた残差の標準偏差を調べている。

母集団の誤差の標準偏差は知ることができない場合がほとんどであり、残差の標準偏差を代用して使う場合が多い。

母集団に関する式が次のものであるとする。

$$y_t = \alpha + \beta x_t + u_t$$

母集団の誤差である u_t の標準偏差を σ とすると、誤差の分散は標準偏差の 2 乗だから次のように表される。

$$\sigma^2 = \frac{\Sigma u_t^2}{n}$$

しかし、母集団について α や β を知ることができないのと同様に、誤差 u_t がどのような動きをするのかを知ることはできない。誤差の平均がゼロであることは仮定できても、どのくらい散らばっているかは調べられない。

しかし、推計した方程式に関しては、推計残差が計算される。推計結果は次の式で表されるとする。

$$y_t = a + bx_t + e_t$$

残差の分散や標準偏差については e_t を使えば計算できる。残差の平均はゼロなので、偏差の二乗和は $\sum e_t^2$ である。母集団の場合はサンプル数 n で割れば分散が求まる。しかし係数 a と b を計算するのに自由度を 2 つ使っているので、分母はサンプル数から 2 を引いたものとなり、その分分散が大きくなる。一般的には定数項を含むパラメーターの数 k だけ自由度が減少する。残差の分散 s^2 は次の式で求めることができる。

$$s^2 = \frac{\Sigma e_t^2}{n-k} \quad (k \text{ は推計すべき定数項を含む係数の数})$$

3.3.3 係数の分散

推計値のばらつき（標準偏差）を表すものを標準誤差と呼ぶ。最小二乗法の計算で推計されるのは係数の平均値で、それはある程度ばらつきを持っている。推計式が

$y_t = a + bx_t + e_t$ で表される時、e_t が正規分布をすると仮定すると係数の分散はそれぞれ次のように計算できる（計算法は省略）。

$$var(a) = \sigma^2(\frac{1}{n} + \frac{\Sigma x_t^2}{\Sigma(x_t - \bar{x})^2})$$

$$var(b) = \frac{\sigma^2}{\Sigma(x_t - \bar{x}_t)^2}$$

標準誤差はそれぞれの平方根である。係数 a、b は正規分布をするため、係数を b、標準誤差を S とすると b/S は t 分布となり、サンプル数が 100 程度の場合は真の係数の値が $b \pm S$ に入る確率は 70 %、$b \pm 2S$ に入る確率が 95 %となる。たとえば係数の推計値が 1.32 で標準誤差が 0.068 なら、95 %の確率で b がとりうる範囲は、$1.32 \pm 2 \times 0.068$、つまり 1.18 から 1.50 までとなる。

3.3.4 t 値

t 値は、その変数が推計式に必要かどうかを判断するために使う。係数÷標準誤差で計算できる。2 以上なら係数が有意だと判断する。ただ、サンプル数が 10 や 20 と少ない場合は、有意水準が変わるので t 分布表で判断する。t 値が大きければその係数がゼロでない可能性が高い。しかし、その係数が特定の推計された係数の値（たとえば 1.32）である可能性が高いことを示しているわけではない。

図 3.2 は推計した係数が同じ値（この場合は 1）でも、係数の標準誤差の違いによって係数の分布が変わることを示している。標準誤差が大きければ、1 を中心として裾野が広く、係数がゼロである可能性を否定できない。一方標準誤差が小さければある程度の分布を勘案しても係数がゼロである確率は小さい。推計した係数が 1 を中心としたゼロではない何らかの値であることを示している。t 値が大きいということは標準誤差が小さく、推計されたパラメーターがゼロから離れていることを示し、t 値が小さいということは標準誤差がゼロである可能性が否定できないことを表わしている。

標準的に出力される t 値は係数がゼロという検定をするものだが、ある特定のパラメーターについても検定できる。ある係数がある定数（k）であるという検定は、次の検定統計量で求めることができ、t 検定を行う。t 検定の詳しいプロセスは 15 章検定参照。

$$t = \frac{\text{係数の推計値} - k}{\text{標準誤差}}$$

統計ソフトウェアで標準的に出力される t 値は $k = 0$ という特殊な場合の検定である（表 3.3）。

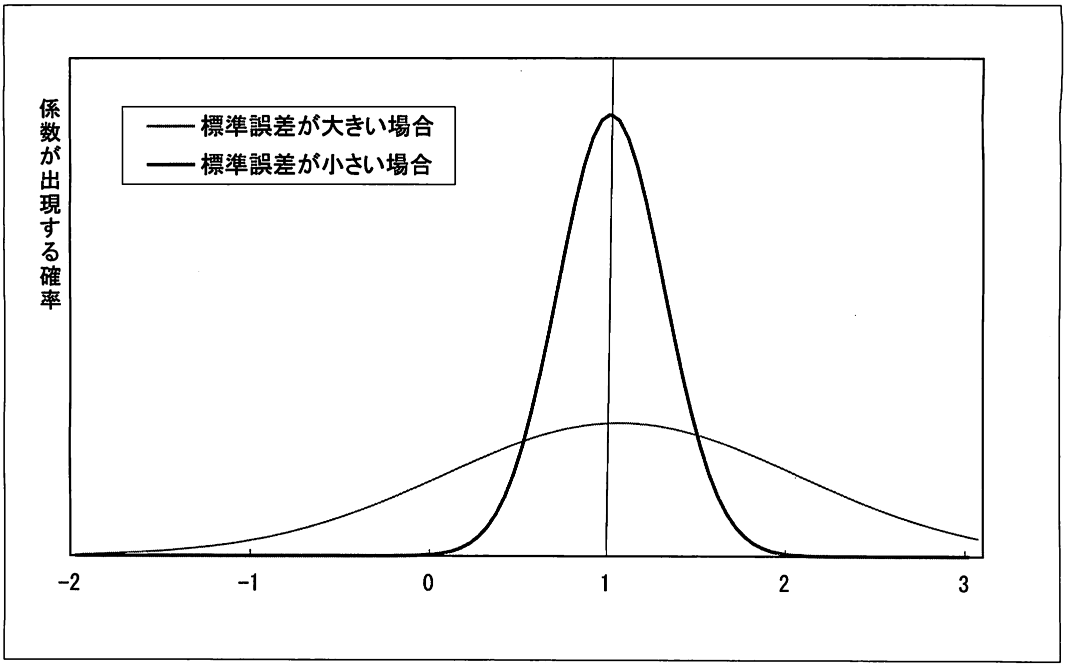

図 3.2: 係数分布の概念図

帰無仮説		係数がゼロである
対立仮説		係数がゼロでない
検定統計量		t 値
結論	t 値の絶対値が大きい (p 値が小さい)	係数がゼロでない
	t 値の絶対値が小さい (p 値が大きい)	係数ゼロの可能性がある

表 3.3: t 検定

3.3.5　決定係数

決定係数は推計値がどの程度実績値を再現できるかの尺度である。1 に近いほどあてはまりがよい。予測では決定係数が高いことが非常に重要だが、経済分析でパラメーターの大きさや有意かどうかに興味の対象がある場合では決定係数はそれほど重視されず、それよりも係数の t 値に注目する。偏差の二乗和である全変動を推定値の変動と残差の二乗和に分解し、全変動に対する推定値の変動の割合を決定係数と呼ぶ。決定係数はゼロ以上 1 以下の値をとる。

$\sum(y_t - \bar{y})^2$ 全変動 (SST)

$\sum(\hat{y}_t - \bar{y})^2$ 推定値の変動 (SSE)

$\sum(y_t - \hat{y})^2$ 誤差の二乗和 (SSR)

全変動 (SST) = 推定値の変動 (SSE) + 誤差の二乗和 (SSR)

決定係数＝推定値の変動／全変動 $= 1 - \dfrac{\text{誤差の二乗和}}{\text{全変動}}$

3.3.6 決定係数と相関係数の関係

決定係数は、被説明変数と推計値との相関係数の二乗となる。

(被説明変数と推計値の相関係数)^2＝決定係数

相関係数を R で表した時、決定係数は R^2 となるため、決定係数の略号として R^2 (アールスクエア) が使われる。また決定係数のことを重相関係数と呼ぶこともある。相関係数が 0.8 の 2 変数を回帰させると、決定係数は 0.64 となる。

この関係は以下の式の展開によって説明できる。y と推計値 $\hat{y}_t$ の相関係数は、次の式で表される。

$$y_t と \hat{y}_t の相関係数 = \frac{\sum(\hat{y}_t - \bar{\hat{y}})\sum(y - \bar{y})}{\sqrt{\sum(\hat{y}_t - \bar{\hat{y}})^2\sum(y_t - \bar{y})^2}}$$

一方決定係数は、次の式である。

$$R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2} = \frac{(\sum(\hat{y}_t - \bar{y})^2)^2}{\sum(y_t - \bar{y})^2\sum(\hat{y}_t - \bar{y})^2}$$

$\hat{y}_t = y_t - e_t$ なので、分子のカッコの中は次のように変形できる。

$$\sum(\hat{y}_t - \bar{y})^2 = \sum(\hat{y}_t - \bar{y})(y_t - e_t - \bar{y}) = \sum(\hat{y}_t - \bar{y})(y_t - \bar{y} - e_t)$$

$$= \sum(\hat{y}_t - \bar{y})(y_t - \bar{y}) - \sum(\hat{y}_t - \bar{y})e_t = \sum(\hat{y}_t - \bar{y})(y_t - \bar{y})$$

最後の等号の変形には、$\sum(\hat{y}_t)e_t = 0, \sum(\bar{y})e_t = 0$ を用いる。この式を決定係数の式に代入すれば、相関係数の二乗になっていることがわかる。

$$R^2 = \frac{(\sum(\hat{y}_t - \bar{y})(y_t - \bar{y}))^2}{\sum(y_t - \bar{y})^2\sum(\hat{y}_t - \bar{y})^2} = \left(\frac{\sum(\hat{y}_t - \bar{y})(y_t - \bar{y})}{\sqrt{\sum(y_t - \bar{y})^2\sum(\hat{y}_t - \bar{y})^2}}\right)^2$$

3.3.7 自由度修正済み決定係数

サンプル数が少ないと決定係数は高くなる。極端な場合では、サンプル数が2個であれば回帰直線は必ず2点を通るので決定係数は1となる。そこで、自由度修正済みの決定係数（*Adjusted* $R^2, \bar{R}^2$ などと略す）を次のように定義する。推計結果を見るときは、自由度修正済み決定係数を見る方がよい。

$$\text{自由度修正済み決定係数} = 1 - \frac{\text{誤差の二乗和}/(\text{サンプル数}-\text{定数項を含む説明変数の数})}{\text{全変動}/(\text{サンプル数}-1)}$$

3.3.8 残差の二乗和 (SSR)

残差の二乗和は各期の残差を二乗して足したものである。推計誤差はなるべく小さい方がよいので残差の二乗和は当てはまりの尺度を示す基本データである。決定係数では、全変動に対する残差の二乗和の比を出して1から引く。そのほか、次に挙げる方程式の当てはまりを表す指標では残差の二乗和が重要な役割を果たす。また、方程式の残差の二乗和の大きさを比べることで検定する方法（F 検定）もよく用いられる。

3.3.9 その他式の当てはまりを表わす指標

決定係数以外にも推計式の当てはまりを表わす指標がある。残差の二乗和を SSR、n をサンプル数、K を定数項を含む説明変数の数とすると、次の式で表される。対数尤度の計算式については6章参照。

対数尤度 (log likelihood)

$$lnL = -\frac{n}{2}(1 + \log 2\pi + \log \frac{SSR}{n})$$

赤池情報規準（AIC, Akaike Information Criterion ）

$$AIC = \log(SSR/n) + 2K/n$$

シュワルツベイズ情報規準 (SBIC, Schwartz Bayesian Imformation Criterion)

$$SBIC = \log(SSR/n) + (\log n)K/n$$

いずれも SSR の関数である。誤差が大きくなれば SSR も大きくなることから、対数尤度は大きければ大きいほど当てはまりがよいことを示す。残差の二乗和は説明変数を増やせば必ず減るので、説明変数を増やせば対数尤度は必ず大きくなる。その意味では当てはまりを表す指標としては適当ではない。AIC は、SSR が増えたり、説明変数の数が増えたりすると大きくなる。AIC で最も当てはまり度合いのよい方程式を選ぶには、AIC が最も小さい方程式を探せばよい。シュワルツベイズ情報基準は、$SIC, BIC, SBIC$ などと呼ばれる。サンプル数が 100 の時、$\log 100 = 4.6$ なので、AIC は K（定数項を含む説明変数の数）を 2 倍しているのに対し、SBIC では 4.6 倍していることになる。係数の追加に対して SBIC の方が大きくなりやすく、説明変数を追加することに関してより厳しい基準だといえる。

3.3.10 ダービン・ワトソン比

ダービン・ワトソン比（Durbin-Watson statistic, DW などと略す）は、誤差が 1 期前の誤差とどの程度相関しているのかを示す。過去の残差との相関（系列相関）が高い場合、最小二乗法の推計値は望ましい性質を示さないため何らかの工夫が必要となる（5 章参照）。ダービン・ワトソン比はゼロから 4 までの値をとり、2 になる場合が系列相関がまったくないことを示す。これは、誤差系列とその 1 期ラグをとった系列との相関係数を ρ とすると、次の式が近似的に成り立つためだ。

$$\text{ダービン・ワトソン比} \approx 2(1-\rho) \tag{3.9}$$

相関係数はマイナス 1 から 1 までの値をとり、完全に逆相関している時（ρ がマイナス 1 の時）、ダービンワトソン比は 4 になり、完全に正の相関をしている時（ρ が 1 の時）、ゼロになる。誤差が 1 期前の誤差と無相関の時は 2 となる。

残差を e_t とすると、1 期前の系列（e_{t-1}）との相関係数は、次の式で表される。

$$\rho = \frac{\sum_2^n e_t e_{t-1}}{\sqrt{\sum_2^n e_t^2}\sqrt{\sum_2^n e_{t-1}^2}}$$

この相関係数をみれば、基本的には 1 期前の誤差との相関の度合いがわかるが、サンプルが少ない場合に統計的な問題があるためダービン・ワトソン比が使われる。ダービン・ワトソン比は次のように計算される。

$$\text{ダービン・ワトソン比} = \frac{\sum_2^n (e_t - e_{t-1})^2}{\sum_1^n e_t^2}$$

ダービン・ワトソン比は次のように展開できる。

$$\frac{\sum_1^{n-1} e_t^2 + \sum_2^n e_t^2 - 2\sum_2^n e_t e_{t-1}}{\sum_1^n e_t^2} = \frac{\sum_1^{n-1} e_t^2}{\sum_1^n e_t^2} + \frac{\sum_2^n e_t^2}{\sum_1^n e_t^2} - 2\frac{\sum_2^n e_t e_{t-1}}{\sum_1^n e_t^2}$$

$$\approx 1 + 1 - 2\rho = 2(1-\rho)$$

これは、近似的に (3.9) 式が成り立つことを示している。

3.3.11　F 値

最小二乗法の推計で推計値などと同時に出力される F 値は、「式のパラメーター β_1 や β_2 のいずれかに意味があるかどうか」という検定である。t 値が個別の係数の意味があるかどうかの検定であるのに対し、F 検定は式全体に意味があるかどうかの検定をする。一つでも有意な係数があれば F 値は有意になり、すべての係数がゼロの可能性がある場合は有意にならない。自由度によって判断する数字が変わるので、確率（p 値）を見て判断するのが便利だ。

p 値は「係数がすべてゼロである」という仮定のもとで、統計量（F 値）がどのくらいの確率で起るかを示す。確率が低ければ、「係数がすべてゼロ」でないことを示す。

具体的には、「すべての係数がゼロ」という制約のある推計と制約のない推計のそれぞれの残差二乗和（SSR）を比較することで検定する（表 3.4）。次の式のテストをするとする。まず通常の式を推計する。

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + e_t \tag{3.10}$$

次に制約のついた式を推計する。制約条件は次の二つである。

$$\beta_1 = 0$$

$$\beta_2 = 0$$

つまり、y_t を定数項だけに回帰させる。

$$y_t = \tilde{\beta}_0(\text{定数項}) + \tilde{e}_t \tag{3.11}$$

無制約の推定（(3.10) 式）による残差二乗和 SSR と制約付きの推定（(3.11)）による残差二乗和 SSR_r を計算し、次の F 分布にしたがう検定統計量を計算する。

$$F = \frac{(SSR_r - SSR)/q}{(SSR)/(n-p)}$$

q は制約の数、n はサンプル数、p は無制約モデルの定数項を含めた説明変数の数である。

<table>
<tr><td colspan="2">帰無仮説</td><td>回帰係数がすべてゼロ</td></tr>
<tr><td colspan="2">対立仮説</td><td>回帰係数がすべてゼロではない</td></tr>
<tr><td colspan="2">検定統計量</td><td>F 値</td></tr>
<tr><td rowspan="2">結論</td><td>F 値が大きい
(p 値が小さい)</td><td>少なくとも一つの係数は有意である</td></tr>
<tr><td>F 値が小さい
(p 値が大きい)</td><td>すべての係数が有意でない</td></tr>
</table>

表 3.4: F 検定

3.3.12 最小二乗法の計算例

これまでの説明を具体的な例で解説しよう。まず、説明変数は実質 GDP、被説明変数は実質民間消費支出とする (図 3.3)。推計期間は 1980 年度から 2002 年度までの 23 サンプルである。統計のサンプル数としては少ないが、一つの例として説明する。

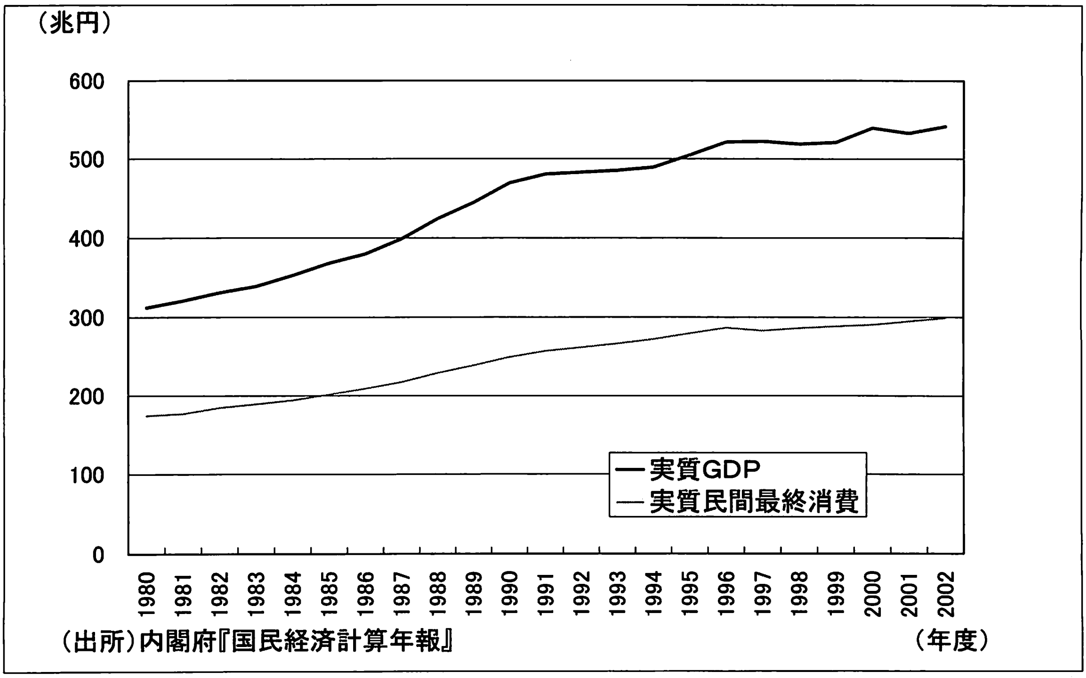

図 3.3: 所得と消費の関係

実質 GDP の一定割合が消費に回るとして、次のモデルを想定する。y_t が実質民間最終消費、x_t が実質 GDP である。

$$y_t = a + bx_t \quad (t = 1, 2, \ldots, T)$$

この式に最小二乗法を適用して計算した結果が次の表 3.5 である。出力には EViews という計量経済分析ソフトを用いた。

Dependent Variable: CP95
Method: Least Squares

Sample(adjusted): 1980 2002
Included observations: 23 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3845.475	4316.08	0.890965	0.383
GDP95	0.538952	0.009514	56.646	0

R-squared	0.993498	Mean dependent var	244750.6
Adjusted R-squared	0.993188	S.D. dependent var	42783.02
S.E. of regression	3530.994	Akaike info criterion	19.25949
Sum squared resid	2.62E+08	Schwarz criterion	19.35823
Log likelihood	-219.484	F-statistic	3208.77
Durbin-Watson stat	0.919831	Prob(F-statistic)	0

表 3.5: 最小二乗法の出力例

この表の意味しているのは、次の式である。

$$CP95 = 3845.475 + 0.538952 \times GDP95 + e$$

被説明変数 (Dependent Variable) は CP95 であり、最小二乗法（Least Squares）で推計していることが示されている。サンプルは 1980 年度から 2002 年度で、サンプル数は 23 である。GDP95 にかかる係数（Coefficient）は 0.538952 で標準誤差（Std. Error）は 0.009514 である。t 分布表によれば自由度 21（サンプル数 23 －係数の数 2）の時、t 値が 2 なら約 95 %をカバーする。つまり、$0.538952 \quad \pm 2 \times 0.009514$ の範囲で 95 %をカバーすることになる。具体的に計算すると、係数を β として、$0.519924 < \beta < 0.55798$ の範囲である。この範囲はゼロからは程遠いので、GDP95 にかかる係数がゼロではないといえる。t 値（t-Statistic）が 56.646 と高いのはゼロから遠く離れていることを示している。対応する P 値（Prob.）もかなりゼロに近く、ここではゼロとなっている（完全にゼロというわけではない)。一方、定数項で同じ計算をすると、係数を α として、$-4786.69 < \alpha < 12477.64$ となり、95 %で取りうる範囲にゼロが入る。したがって、定数項がゼロであるという仮説は棄却できない。t 値は 0.890965 と、棄却できるおおよその水準 2 よりも小さくなっている。p 値も 0.3830 となっており、「40 %有意水準」という通常は想定しない高い有意水準でやっと棄却できるレベルであることがわかる。

決定係数（R-squared）を見ると、0.993498、自由度修正済み決定係数 (Adjusted R-squared) も 0.993188 と高い。サンプルが少なく両変数とも上方トレンドを持っているのでかなり高い数値となっている。回帰の標準誤差 (S.E. of regression) は 3530.994、残差の二乗和（Sum squared resid）は 2.62E+08（E+08 は 10 の 8 乗の意味）だ。対数尤度 (Log likelihood) は-219.4841 となっている。ダービンワトソン比 (Durbin-Watson stat) は 0.919831 と正の相関があり、2 から離れている。誤差の系列相関が疑われる。被説明変数の平均 (Mean dependent var)244750.6、標準偏差 (S.D.dependent var)42783.02 に続いて、当てはまりを示す指標として、AIC(Akaike info criterion)19.25949、SIC(Schwarz criterion)19.35823 が出力されている。これらの統計値はこの結果だけでは特に何かを判断することはできず、ほかの推計結果と優劣をつける場合に使用する。F 値 (F-statistic) は 3208.770 と大きくそれに対応する p 値 (Prob) もゼロに近い（表示上はすべてゼロとなっている)。

3.4 予測

ある経済変数 x_t と y_t に関して次のようにモデルが推計できたとする。

$$y_t = a + bx_t + e_t \quad (t = 1, 2, \ldots, T) \tag{3.12}$$

推計期間は 1980 年から 2002 年までとする。2003 年の y の予測値は、2003 年の x_t がわかれば上の式に代入することで求めることができる。誤差項の期待値はゼロなので、予測値の計算は誤差項のない次の式を計算することである。x_t の予測値を x_{ft}、y_t の推計された予測値を $\hat{y}_{ft}$ とする。

$$\hat{y}_{ft} = a + bx_{ft}$$

真の値は、次のように書ける。ε_t は標準的な誤差の仮定を満たすものとする。

$$y_{ft} = \alpha + \beta x_{ft} + \varepsilon_t$$

真のパラメーターによる値と推計したパラメータによる予測値の予測誤差は次の式で表される。

$$\hat{y}_{ft} - y_{ft} = (a - \alpha) + (b - \beta)x_{ft} - \varepsilon_t$$

a の期待値は α、b の期待値は β、ε_t の期待値はゼロなので、予測誤差の期待値はゼロとなる。

$$E(\hat{y}_{ft} - y_{ft}) = 0$$

予測誤差の分散は次の式で表される。

$$V(\hat{y}_{ft} - y_{ft}) = \sigma(1 + \frac{1}{n} + \frac{(x_{ft} - \bar{x}_t)^2}{\sum(x_t - \bar{x}_t)^2})$$

これは、推計式の誤差の分散が小さければ誤差の分散も小さく、大きければ誤差が大きくなることを示している。また、x_t の予測値（x_{ft}）が x_t の平均から離れたものであればあるほど、予測誤差の分散は大きくなる。

3.4.1 説明変数の予測

最小二乗法による予測の大きな問題は、(3.12) 式を使って予測するとき、x_t の予測値がわかっていなければならないことである。説明変数の予測法は、式の性格や予測する対象によってさまざまなものが考えられる。

- サーベイデータの利用
- 直近の値で横伸ばし
- 時系列モデルを利用
- トレンドで伸ばす

サーベイデータとはアンケート調査によるデータのことである。たとえば、米国の実質 GDP の予測値が必要な場合は、エコノミストによる予測値の平均を使うことが考えられる。また、直近の値で横伸ばしをすることも考えられる。変動することはわかっていても予測が非常に困難な場合や、特にこれまでと環境が変わらない状態で予測したい場合である。たとえば直近の原油価格が 1 バレル＝ 20 ドルであったとすると、その値が将来まで続くとして予測する。理論的な根拠が乏しい場合は、過去の系列の動きから機械的に先行きを予測することが考えられる。ARMA モデル（9 章参照）による予測が代表的なものである。また、トレンド傾向が明確な場合は、トレンド変数（4 章参照）を使った予測も考えられる。

x_t がほかの方程式によって決まる場合には、1 本だけの方程式での予測ではなくなり、同時方程式モデル（7 章参照）となる。この場合は、そもそも単一方程式による推計では推計した係数にバイアスがでることが知られており、新たな工夫が必要となる（5 章参照）。

3.4.2 定数項修正

方程式を使って予測する場合、最近時点での誤差をそのままにして 1 期先を予測すると段差が生じてしまう場合がある。その場合は誤差の分だけ定数項を修正して予測値とする。仮想的な値で考えてみよう（図 3.4）。第 5 期までは実績値がわかっているとする。第 6 期の説明変数を予測すれば、推計式に当てはめて、第 6 期が予測できる。その値が B 点である。

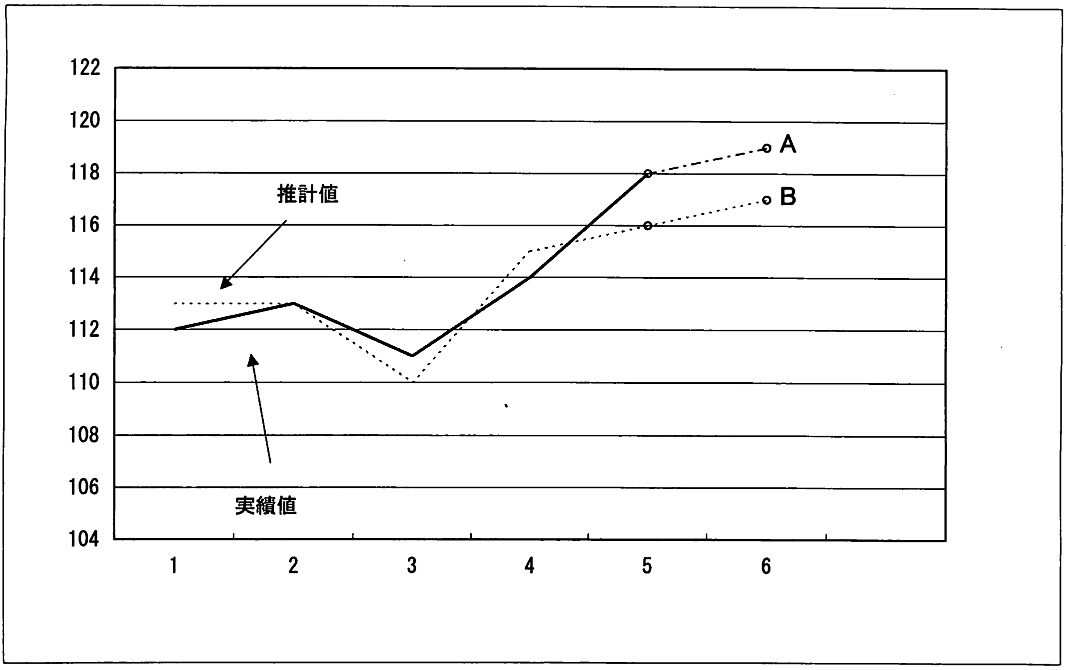

図 3.4: 定数項修正による予測値の修正

これをそのまま予測値として使っていいだろうか。実績値の最新期（第 5 期）では、推計値よりも実績値が 2 だけ上にはずれている。予測値 B は最新の実績値より低い。推計値の数字は拡大傾向が続いているのに実績値と予測値を接続すると予測は拡大傾向が止まって低下に転じるという予測になる。そこで、最新の実績値に推計値の増加を反映させた A 点も予測値の候補として考えられる。最新の実績値と予測値のかい離の分だけ推計式の定数項を上乗せすることを意味しており、こうした修正を定数項修正と呼ぶ。

予測値 A と予測値 B のどちらが良いかは、経済変数の性質や、現状判断（異常な事態が起きて誤差が増えたのか、たんなる偶然なのかなど）によって総合的に決める必要がある。最新期の誤差が、第 5 期だけの特別な理由で生じたもので、第 7 期にはその影響はなくなると考えるなら B 点を予測値として採用する。

一方、第 6 期の誤差がすぐに消えるとは考えられない場合は、その誤差分だけ修正した予測値 A を採用する。B 点の予測は第 6 期よりも低下するという予測だが、さまざまな経済情勢による判断から実績値よりも予測値の低下が考えられない場合である。また、誤差に系列相関がある場合は、急激に誤差が動くことは考えにくいので、予測値 A を採用するという考え方もある。

3.5　まとめ

推計の基本は最小二乗法である。最小二乗法で推計したら、まず実績値と推計値のグラフ、誤差のグラフを見る。ここでもグラフによる確認は大事である。あまりにもおかしい結果の場合はデータに戻って検討し直す必要がある。統計値としては、まず決定係数を見る。予測に使う場合は決定係数は高いほどよい。筆者の経験からはかなり精密な予測をする場合は決定係数が 0.99 以上は必要だと思う。ただ、係数が有意かどうかの検定などの場合は決定係数がかならずしも高くなくてもかまわない。クロスセクションデータの場合なら、決定係数が 0.5 あれば十分だろう。次に、ダービンワトソン比を見る。2 に近いほど残差に自己相関がないことを示す。2 から大きくはずれて自己相関がある場合は何らかの対処が必要だ (5 章参照)。次に、個々の説明変数を検討する。まず、t 値をみる。t 値が 2 より小さい場合はその変数は有意でない。たとえば、30 という数値が推計されても推計誤差が大きく統計的にはゼロとみなせると考え得る。逆に係数が 0.01 であっても t 値が 2 以上あるということはその変数はゼロと明確に異なり 0.01 である、ということを示す。

説明変数が有意であれば、説明変数のパラメーターの符号が経済理論の想定通りか、パラメーターの大きさは適当かどうかなどを検討する。対数線形で推計していれば、係数は弾性値を表し、チェックがしやすい。