エコノメトリック・モデル作成の原理

Prof. H. Wold*

今日お話しようとするモデル作成の問題は，需要分析の分野ではじめ論ぜられたものでありますから，今日のセミナーの表題を「需要予測」とすることも出来るでしょうが，「エコノメトリック・モデル作成の原理」とでも呼んだ方が，内容が一般的な視野に立つものであるところから，より一層適切であろうかと思います．

エコノメトリック・モデルには，Tinbergen によって提唱された recursive system と，Haavelmo によって提唱された interdependent system との2つを考えることができますが，この二つの関連は非常に興味があり，また重要な問題であります．これについて新しい見地に立って，問題を解明しようとしたのが，ISI での私の報告でもあります．問題の難しさは，しかしながら，数学上の問題にあるのではなく，どういうモデルが意味があるかという点についてのいろいろな考え方をまとめることにあるといえましょう．

モデルを論ずる場合にはいろいろな観点が存在します．それをまとめますと，次のようになると思います．

1. Description versus Explanation
2. Experimental versus Non-experimental situations
3. Exact versus Distributed relations
4. Single versus Multiple relation models
5. Causal chains versus Interdependent systems

私がここで扱おうとするのはこの第5の点ですが，他の4つの見方も重要であります．これについては，J. R. S. S. vol. 119 (1956) pp. 28—61 で応用統計学についての広汎な展望を致しました．いまそれを簡単に説明しますと次の図のようになります．私は統計学の領域を四つのセクターに分けました．まず大きく description と explanation の2つに分け，それぞれがさらに experiments と non-experimental situations とに分けられます．

	Experiments	Non-experimental situations
Description	①	②
Explanation	③	④

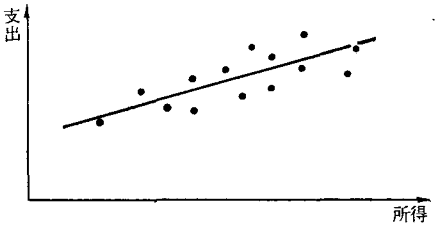
まず description と explanation との間の関係は，description は What is happening? という質問に対する答であるのに対して，explanation は Why is

* 昭和35年6月13日 ISI セミナーにおける要旨　昭和35年10月31日受理

it so？ という問に対する回答を与える分野であります．explanation は，あるいは，変数間の関係を論ずる領域であるともいえましょう．したがって下の欄では関係を扱い，上の欄では事実を扱うということができます．explanation のためのモデルを作る場合，まずそれに先立って事実を記述するという段階を通らなければならず，したがって上の欄は準備の段階であり，下の欄は分析の段階に相当するということになります．

第二の分類は，実験の可能な領域とそれが不可能な領域とに分けることです．薬学界で最も重要な実験といわれるオーストリアの Dietle の 1849 年の実験は，1つの典型的なものということができましょう．最近の統計学においては，実験計画や分析の手法が非常に発達してきました．R. A. Fisher により randomization の概念が 1920 年代に導入されましたが，これは2つの処理を比較するという典型的な実験の場合に用いられる重要な手法であります．randomized experiment を設計することによって，はじめて二つのグループが比較可能となります．そしてまた古典的な実験では比較的多数の対象についての実験によって結論を出していたのに対して，Fisher は少数の実験で結論を出しうるような理論を提供しました．生物学はこのような手法によって，大いに進歩しました．

一方右側の実験の出来ない分野には，社会科学や天文学などがあげられます．左下の分野には生物学があげられます．いうまでもなくこの四つの分野のそれぞれの中間の領域もあり，農業の分野には，実験ができないような面もあるわけで，中間的な分野も存在します．そしてこれら四つの領域は，内容の困難さにしたがい，統計学の歴史においても次々に段階的に研究がすすめられてきたといえましょう．左上の領域ではゲーム実験の記述，すなわちギャンブリングの研究として古くから開拓され，ついで人口などの社会的なデータの記述という右上の領域が 19 世紀に研究され，R. A. Fisher によって，左下の第三の領域が開拓されてきたのです．しかし第四の右下の領域は現在まだあまり研究されていない分野であります．今日お話しようとするのはこの領域に属する問題でして，ふつうの統計学の教科書で扱われているような実験は行うことができず，randomize されたデータをうることは困難です．私がここで説明しようとする causal chain や interdependent relation は以前の統計学でも，また自然科学のどんな領域においても用いられなかったエコノメトリックスにおける革新であるということができましょう．



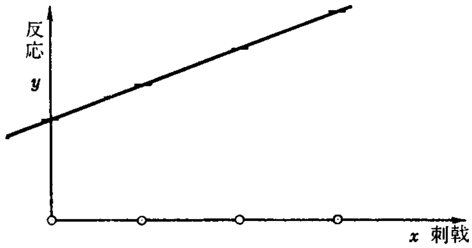
いま家計支出調査を例にとってみましょう．いろいろな所得をもつ家計の支出データが与えられれば，図のように回帰線を描くことができます．支出を食料品支出にとれば，Engel が 1895 年に発表した所得と食料品支出との間の関連を示すわけです．彼はベルギーやドイツにおける家庭のデータを用いて研究を行ったものですが，彼はこれを記述という観点から論じていたのであって，もしある家庭がこれだけの所得をもてばどうなるかということは考えていなかったのです．上のデータは実験的なデータではなく，原理的には

そのような実験は可能でしょうが，従来と違った所得水準で生活をしてもらうということは不可能ですから，ここで randomized experiment を行うということは，事実上できないといってよいでしょう．したがってこの第四の領域における問題は，第三の領域のそれとは，まったく異った状態にあります．データは randomize されていないのですから，所得が変ったときの家計支出に及ぼす影響を，この回帰線から論ずることは，家庭の人数とか構成など他の要因について考えることなしにはできないわけです．したがってわれわれはここで多重回帰分析を用いなければならないこととなります．しかしながらいかに多くの変数を導入したとしても，ある重要な要因が他になお存在するという可能性は常に残っており，他の要因の影響を取除くためには，家庭を random に割当てること以外にないといえます．第三の領域では randomize することが可能であり，もしある要因に興味があるならば，他の要因の影響を除くように対象を randomize して選ぶことによって実験することができるので，問題の解析が一層容易であります．この領域の問題の解析は routine として考えることができます．randomization という手法はどんな問題に対しても適用することができます．一方第四の領域では，それは routine とは考えられません．multiple regression はすべての状態について用いることができるというわけではないのです．たとえ残差項が小さくとも，新しい説明変数を導入することができますし，またそれによって他の回帰係数の傾斜がまったく違ったものになるということもありえます．この点については私の書物 “Demand Analysis” に詳述してありますのでそれを御参照下さい．

次に第三の観点に移りたいと思います．これも非常に単純な問題です．しかし一つ一つはさして面倒ではないのですが，これらの観点を常に同時に考慮しなければならないということから難しさが生じます．

いまある与えられた刺戟に対してある反応をしらべたとします．刺戟と反応との間に正確な関係があるならば，何回同じ実験をくりかえしても，いつも同じ反応が得られ，いろいろな処理の水準 x に対する反応 y は x の函数として正確に表現することができます．すなわち

$$y=f(x)$$

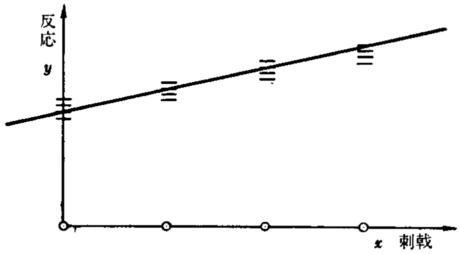


物理学の古典的な実験で見られるボイル・シャールの法則，圧力は絶対温度に比例し，体積に反比例するという法則は，実際上正確なものとして観察されます．すなわち

$$P=C\frac{T}{V}$$

なる関係が成立します．

これらの関係はほぼ正確ですが，このような実験において，誤差の要素を考慮に入れて分析するようになったのは比較的最近のことです．いくつかの刺戟の水準について，5，6 回のくりかえし実験を行いますと，いろいろちがった反応がえられます．この結果を表現する場合には，y の x に関する回帰線を作るのが通常です． 1850 年に行われた心理学の実験で，ある心理学的な



刺戟に対する反応は与えられた刺戟に比例することが認められました．この場合の関係は次の式を与えられます．

$$y=f(x)+\varepsilon$$

ここに $f(x)$ は与えられた x に対する y の期待値で，

$$f(x)=E(y|x)$$

と書くことができます．数学的には反応はこの平均値にあるちらばりを加えたものとして表現されます．通常の教科書にはこのようには規定されていないのですが，この点をはっきりすることが重要です．すなわち実験の結果からここに引かれた直線は，ある与えられた刺戟に対して期待される反応はどれ位かということを示します．これは一つの数学的なモデルであります．そしてこのモデルはそれを如何に利用しうるかということに対する基礎を与えるものであります．

この場合には，もしある指定された大きさの刺戟xを与えたとき，このモデルは，反応の期待値を与えます．これはモデルがそのように作られているからです．いまここで逆の推論をしようとすると，問題は困難になってきます．もし正確な関係が存在していれば，それは極めて容易です．ボイル・シャールの法則の場合なら，一方が判れば他はすぐ求められます．いまあるyの値 y_0 が与えられたとします．その時，どんな x_0 について，x_0 が与えられたときの y の期待値が y_0 になるか，というのがこの場合の逆の推論になります．y が変ったとき，x はどう変るかということについては何も示さないのです．この二通りの推論の差を区別しておくことが，これから私が論じようとすることを理解する上で，極めて重要と思います．

さらにこのようなモデルを考えたことから，その利用の方法の一つとして，この関係をどんな推定方法で推定したらよいか，ということも規定されます．モデルの両辺についてxが与えられたときの期待値をとれば $E(\varepsilon|X)=0$ となります．これより $E(\varepsilon)=0$ がえられ，xはコンタントですから，$\mathrm{Cov}(\varepsilon, X)=0$ となり，最小自乗推定値は不偏であることになります．モデルがきまれば，それによって推定方法も自からきまってくることになるわけです．

disturbed relation の operative use について今まで説明してきましたが，これはさらに次の single relation と multiple relation とに分けることができます．

single relition は次のようなものです．

$$y=f(x)+\varepsilon \qquad \text{その最も単純なものとして} \qquad y=\alpha+\beta x+\varepsilon$$

これは

$$E(y|x)=f(x) \qquad\qquad E(y|x)=\alpha+\beta x$$

なることを意味し，最小自乗法が不偏推定値を与えます．さてその operative use として，x から y への直接的な推論は，

1) x が与えられれば，y の期待値は $E(y|x)$ となる．また次に y から x への逆の推論は，

2) $y=y_0$ が与えられたとき，$E(y|x)=y_0$ となるような x の値はどんな値か

という形で示されます．これが single relation のモデルについて言えることのすべてであって，これ以上のことは何も言うことはできません．

次にいくつかの関係式で表わされるモデル，

$$\begin{cases} y_1=f_1(x_1,\ x_2,\ \cdots\cdots) \\ y_2=f_2(x_1,\ x_2,\ \cdots\cdots) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

すなわち multiple relation について考えましょう．ただここでは一般的な形で扱わず，二つの特別なタイプのモデル， causal chain と interdependent system について述べてみたいと思います．

まず causal chain ですが，例によって説明しましょう．この例ははじめ Tinbergen によって示され，私が 1947 年の論文で用いました．

$$\begin{cases} d_t=\alpha_1-\beta_1 p_t+u_t & \text{（需要函数）} \\ s_t=\alpha_2+\beta_2 p_{t-1}+u_t' & \text{（供給函数）} \\ p_t=p_{t-1}+\gamma(d_{t-1}-s_{t-1})+u_t'' & \text{（価格メカニズム）} \end{cases}$$

ここに　d_t：需要，s_t：供給，p_t：価格

三番目の式にある s_{t-1} の代りに s_t を用いたものも考えられます．内生変数はこの三つです．まず causal chain ということについて説明しましょう．もし $t-1$ 期の値が判っていれば，この式は t 期の値を１つ１つ与えることがわかります．第１に $t-1$ 期の価格によって，s_t が求められ，また第２の式から p_t が，そして最初の式から d_t がえられます．第３式の s_{t-1} 代りに s_t が用いられているとしても，同じです．これは Tinbergen の言葉を使えば， causation の流れ，ということになります．

さてこのようなモデルの operative use について，すなわちこの構造モデルを使って実際にどういうことがいえるか，という点を見てみましょう．

まず予測という点について，上で述べてきたように，x が与えられれば y の期待値が判るという条件付期待値の考え方が用いられることになります．次に代入という手続が用いられます．第３の式で p_t を求め，その値を第１の式に代入して d_t が求められます．ここで，

$$E(y|E(x|z))=E(y|z)$$

なる定理が成立する必要があります．そしてそのためには u_t'' が s_t，p_{t-1} と独立であると仮定する必要があります．そしてもし u_t'，u_t'' の間に相関がないならばこういうことがいえます．

次にこのモデルの誘導形について考えてみましょう．この誘導形は代入を行なってゆけばよく，

$$\begin{cases} s_t=\alpha_2+\beta_2 p_{t-1}+u_t^{*} \\ p_t=t-1\text{ 期の項}+u_t^{**} \\ d_t=t-1\text{ 期の項}+u_t^{***} \end{cases}$$

といった形になります．ここに＊をつけた項は n_t, u_t', u_t'' の線型結合です．これらの誘導形は線型式で，重要なことは，誘導形が条件付期待値で表わされるという点です．例えば，

$$p_t=E(p_t|p_{t-1},\ d_{t-1})+u_t^{**}$$

となります．したがってわれわれは構造モデル（最初のモデル）についても，誘導形についても同じように用いることができます．これが causal chain と interdependent system との本質的な相違点であります．

次に interdependent system を例によって説明しましょう．

$$\begin{cases} q_t=\dfrac{3}{5\rho}p_t+v_t & \text{（需要函数）} \\ q_t=\dfrac{3}{5}p_{t-1}+v_t' & \text{（供給函数）} \end{cases} \qquad \text{ただし}\quad -0.96\leq\rho\leq 0$$

ここで内生変数は q_t と p_t です．interdependent system は現在時点の内生変数の間の implicit な関係とみることができます．その explicit な関係は次の誘導形で示されます．

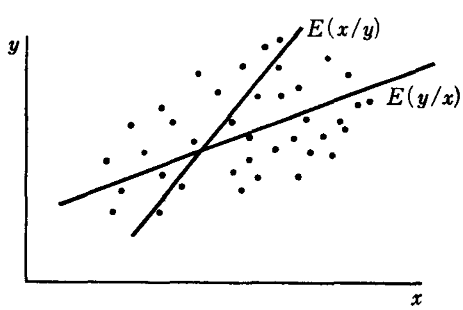
$$\begin{cases} q_t=\dfrac{3}{5}p_{t-1}+v_t'=E(q_t|p_{t-1})+v_t' \\ p_t=\rho p_{t-1}+v_t^*=E(p_t|p_{t-1})+v_t^* \end{cases}$$

ここでの仮定は，v_t, v_t' が先決変数と相関がないということです．そうすると右側のような条件付期待値で表わすことができます．この誘導形は，もし先決変数が決れば現在の変数の値を推定するために用いることができます．しかしながらここで第1の式から予想されるように，

$$E(q_t|p_t)\neq\frac{3}{5\rho}p_t$$

上の関係の等号は決して成立しません．それは誘導形を求めるに際して，単なる代入の手続きのみならず，移項をしたり割算を行っているからである，ということができます．代入だけの操作では誘導形は求められません．

回帰の関係も，条件付確率という観点からすれば，逆転可能とはいえません．y の x に関する回帰線と，x の y に関する回帰線とは違った線になることは衆知の通りです．これはこの関係が逆転できないものであることを示しています．したがって interdependent system においては，誘導形においてのみ実際への適用が可能であるということになります．はじめの式において $\dfrac{3}{5\rho}$ はよく需要弾力性というように言われていますが，そういう解釈はできないわけです．一方 causal chain の場合には，はじめの体系も，その誘導形も，どちらも用いることもできます．



パラメターの推定もこの考え方に立って行われなければなりません．causal chain system の

場合には，はじめの体系について推定を行なっても，誘導形について推定を行ってもよく，最小自乗法を用いることができ，大標本の場合には同じ結果をうることができます．interdependent system については limited information method などいろいろなテクニックが用いられますが，その基礎はやはりこの条件付確率という観点と関連して論ぜられなければなりません．

今日のセミナーで問題になっている需要予測という点について考えてみますと，予測もやはりモデルの operative use ということです．causal chain のモデルの場合にははじめの体系でも，誘導形でも，どちらでも予測に使えますが， interdependent system のモデルの場合には，誘導形がその役に立ちます．私がここで強調したい点は極めて単純なことで，あるモデルが構成されれば，それに伴って，理論的な応用であれ実際問題への適用であれ，その用い方を規定する基盤が与えられ，そしてまた同時に古典的な最小自乗法によってそのパラメターを推定することの根拠が与えられる，という考え方です．そしてその場合，条件付確率という概念が中心的な役割を果しているということです．

大(沢　豊)