最小自乗法の基礎と計量経済学におけるその適用

宮川公男

同時的及び時系列データーからの経済的関係式の推定――それは計量経済学の主要な一部門であるが、そこでは先づ二つの問題を解決せねばならない。㈠関係式の函数型の決定。㈡そのパラメーターの推定法の決定、である。㈠の問題は、主として経済理論に依存し、それに数学的要素が交錯しているものといえようが、以下我々はこの問題を扱おうとはしないで、我々の議論を線型式を前提として進める。㈡の問題については、パラメーター決定の基準として何が最良であるか明白でない。最近エコノメトリシャンの間で勢を得ているのは最尤法の主張であるが（統計的決定論の研究もより盛んであるが、その実際的適用には尚大きな距離がある）、その他伝統的な最小自乗法、χ^2法、δ^2法等がある。しかし、これらの中で経済分析において伝統的に用いられており、我々に親しみの深いのは最小自乗法である。ガウスがその原理を確立して以来、あらゆる関係諸科学において、曲線のあてはめ、誤差の調整等に、揺ぐことのない王座を占めて来た最小自乗法が、経済データーの解析にも指導的な役割を演じて来たとしても何の不思議もないことに思われるであろう。しかし、最小自乗法の応用が、その基礎原理を省みることなく、無批判的に行われるとしたならば、そこには大きな落し穴が生ずるであろうということも、同様に不思議ではないといわなければなるまい。その最も顕著な理由として、最小自乗法を育てて来た実験可能な自然科学という土壌には、実験不可能な経済現象から生れたデーターという土壌には含まれていない「肥料」があるかも知れないからである。今日のエコノメトリシャンの最小自乗法批判の先駆となった一人ホーヴェルモーの論拠[1]も煎じ詰めればこの二つの土壌の異質性にあったということが出来る。しかし、同時方程式から最小自乗法を捨てて最尤法へ移ることだけが、二つの土壌の異質性の認識の唯一の方法ではない。そこにはまだ、間接最小自乗法と呼ばれる誘導型（reduced form）によるものあり、またウォルトの[2]如く逐次（recursive）模型に最小自乗法擁護の拠り所を求める方向もある。これらの諸方向が互いにどのような関係にあるのかを知ることは非常に重要なことであろう。もっとも、どれか一つの方向が排他的に正当であるという如き結論は、出すべきでもないし、又それは不可能でもあろうが。しかし、これらの諸論が最小自乗法の基礎原理といかなる関係をもつかについては、一つの明確なものがあるのであって、ここに考察しようとするのはそれである。ただここで取扱おうとすることは、根本的な問題の一部であって、他に経済データーの特徴たる時系列性や観察誤差の問題等をも論じ

なければ不十分なのであるが、紙数の都合で省略する。

(1) T. Haavelmo; The Probability Approach in Econometrics, *Econometrica*, Vol. 12, Supplement, July, 1944, p. 104. その他 Cowles Commission Monograph No. 10, No. 14 中の T. Koopmans, Haavelmo, 後掲の J. Bronfenbrenner 等の諸論文を見よ。

(2) H. Wold; *Demand Analysis*, 1953, pp. 49—53, 64—71, 202—204.

一　最小自乗法の数学的展開

先づ予備的な定義を述べよう。(1)

元 ξ、η、ζ……の集合は次の条件を満すとき線型空間Lをつくるという。即ち(1) 任意の二元 ξ、η の和 $\xi+\eta$ もまたLに属する。(2) Lの任意の元 ξ と任意のスカラー c とはLの中にスカラー倍 $c\cdot\xi$ を定める。(3) (1)(2)の演算は通常の代数の計算法則に従う。とりわけLは加法の恒等元たる零元Oを含む。次に、Lの任意の二元 ξ、η をとり、その内積（これを (ξ,η) と記す）を、次の法則を満すスカラーと定義する。即ち

$(c\xi,\eta)=c(\xi,\eta)$; $(\xi+\eta,\zeta)=(\xi,\zeta)+(\eta,\zeta)$ (双線型性); $(\xi,\eta)=(\eta,\xi)$ (対称性)

$\xi\neq 0$ ならば $(\xi,\xi)>0$ (正値性)

$\xi=0$ ならば $(\xi,\xi)=0$

内積の定義された線型空間Lにおいて、二元間の距離 $\|\xi-\eta\|$ を次のように定義する。(2)

$$\|\xi-\eta\|=+(\xi-\eta,\xi-\eta)^{\frac{1}{2}}$$

特に零元との距離 $\|\xi\|=(\xi,\xi)^{\frac{1}{2}}$ をノルムと呼ぶ。(2)

ここで次のことが成立つ。

$\|c\xi\|=|c|\,\|\xi\|$; $\xi\neq 0$ ならば $\|\xi\|>0$

$(\xi,\eta)\leqq\|\xi\|\,\|\eta\|$ (シュワルツの不等式)

$\|\xi+\eta\|\leqq\|\xi\|+\|\eta\|$ (三角形不等式)

以上で準備を終り、本題に移ろう。

ξ 及び ξ^1、$\xi^2\cdots\xi^p$ をLの一定の元とし、ξ^1、$\xi^2\cdots\xi^p$ の一つの線型式（勿論これもLの元である）

(1) $\xi^*=b_1\xi_1+b_2\xi_2+\cdots\cdots+b_p\xi_p$

で ξ を近似しようとするとき、ξ と ξ^* との距離を最小にするような ξ^* を求めようとするのである。

(2) $\xi=\xi^*+\varepsilon$　即ち　$\varepsilon=\xi-\xi^*$

と書けば、ε は残差と呼ばれるものである。従って、最小自乗法は残差のノルムを最小にするということと同じであるといえる。係数 b_1、$b_2\cdots\cdots b_p$ を決定するために $\|\varepsilon\|^2$ を計算すると、

(3) $\|\varepsilon\|^2=\|\xi-\xi^*\|^2=(\xi-\xi^*,\xi-\xi^*)$

$$=(\xi,\xi)+\sum_{i=1}^{p}\sum_{j=1}^{p}(\xi_i,\xi_j)\cdot b_ib_j-2\sum_{i=1}^{p}(\xi_1,\xi_i)b_i$$

となり、$b_1\cdots$、b_p に関して微分して o とおけば、

$$(4)\quad\begin{cases}(\xi_1,\xi_1)b_1+(\xi_1,\xi_2)b_2+\cdots\cdots+(\xi_1,\xi_p)b_p=(\xi,\xi_1)\\(\xi_2,\xi_1)b_1+(\xi_2,\xi_2)b_2+\cdots\cdots+(\xi_2,\xi_p)b_p=(\xi,\xi_2)\\\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\(\xi_p,\xi_1)b_1+(\xi_p,\xi_2)\cdot b_2+\cdots\cdots+(\xi_p,\xi_p)b_p=(\xi,\xi_p)\end{cases}$$

これが ξ^* の係数を決定する正規方程式といわれるものである。ここでこの正規方程式の係数の行列をAと書き、その行列式を A、A の余因子を A^{ij} と書こう。即ち

(5) $A=\begin{pmatrix}(\xi_1,\xi_1)\cdots\cdots(\xi_1,\xi_p)\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ (\xi_p,\xi_1)\cdots\cdots(\xi_p,\xi_p)\end{pmatrix}$ $A=\begin{vmatrix}(\xi_1,\xi_1)\cdots\cdots(\xi_1,\xi_p)\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ (\xi_p,\xi_1)\cdots\cdots(\xi_p,\xi_p)\end{vmatrix}$

である。前述の内積の性質から、Aは勿論対称行列である。又 $A\geqq 0$ なることも証明出来る。

さてここでAが non-singular 即ち $A\neq 0$ であれば、連立一次方程式の理論から(4)はξ^*の係数b_iを一意的に決定する。[3] このときb_i及びξ^*は次の如く与えられる。

(6) $b_i=\sum_{i=1}^{p}(\xi,\xi_j)\dfrac{A^{ij}}{A}$ $\xi^*=\begin{vmatrix}0\ (\xi,\xi_1)\cdots\cdots(\xi,\xi_p)\\ -\xi_1(\xi_1,\xi_1)\cdots\cdots(\xi_1,\xi_p)\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ -\xi p(\xi_p,\xi_1)\cdots(\xi_p,\xi_p)\end{vmatrix}\Big/A$

残差 ε は

(7) $\varepsilon=\xi-b_1\xi_1-\cdots\cdots b_p\xi_p$

これを行列式で書けば、

(8) $\varepsilon=\xi-\begin{vmatrix}0\ (\xi,\xi_1)\cdots\cdots(\xi,\xi_p)\\ -\xi_1(\xi_1,\xi_1)\cdots(\xi_1\xi_p)\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ -\xi_p(\xi_p,\xi_1)\cdots(\xi_p,\xi_p)\end{vmatrix}\Big/A$ $=\begin{vmatrix}\xi(\xi,\xi_1)\cdots\cdots(\xi,\xi_p)\\ \xi_1(\xi_1\xi_{,1})\cdots(\xi_1,\xi_p)\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ \xi_p(\xi_p,\xi_1)\cdots(\xi_p,\xi_p)\end{vmatrix}\Big/A$

ここで $-\infty<b_i<\infty$ とすればξ^*はLの中にp次元の部分空間を作り、ξ^*はξのこの空間上への射影である。さてもしもLの二元η及びξとの間に$(\eta,\xi)=0$なる関係が成立つとき、ηとξとは直交する (orthogonal) という。(8)からとεとξ_iの内積を作ると、分子の行列式において第一列と第i列が等しくなるから

(9) $(\varepsilon,\xi_i)=0$ $i=1,\ldots\ldots,p$

即ち、残差εはξ_1、$\xi_2\cdots\xi_p$の各々と直交する。従つてεはξ^*とも直交する。例えば、$\xi_1\xi^*\varepsilon$をベクターと考えれば、最小自乗法の原理は、ξ^*を直交する二つのベクターξ^*及びεとに分解することにある、といい得る。ここでε^2を計算しておこう。

(10) $\varepsilon\ \varepsilon_2(\varepsilon,\varepsilon)=(\varepsilon,\xi-b_1\xi_1-\cdots\cdots b_p\xi_p)$

$=(\xi,\varepsilon)=(\xi,\xi)-b_1(\xi,\xi_1)-\cdots\cdots-b_p(\xi,\xi_p)$

行列式で書けば

(11) $\varepsilon^2=(\xi,\varepsilon)=\begin{vmatrix}(\xi,\xi)\ (\xi,\xi_1)\cdots\cdots(\xi,\xi_p)\\ (\xi_1,\xi)\ (\xi_1,\xi_1)\cdots\cdots(\xi_1,\xi_p)\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ (\xi_p\xi_{1,})\ (\xi_p,\xi_1)\cdots\cdots(\xi_p,\xi_p)\end{vmatrix}\Big/A$

最後に、Aが singular な場合、即ち $A=0$ のときは、行列式の定理から$\xi_1\cdots\xi_p$が一次従属、即ち少くとも一つの0でないc_iに対し

(12) $c_1\xi_1+\cdots\cdots+c_p\xi_p=0$

なる関係が成立つ。この場合はいわゆる複共線性 (multicollinearity) といわれるもので、この時はξ_iの間に無駄があるわけで、$\xi_1\cdots\cdots\xi_p$の作る部分空間は $p-1$ 次元以下となる。しかし、この場合もξ_iを一次独立のものだけとすることによつて non-singular の場合に帰せしめることが出来る。

(1) 例えば Birkhoff & MacLane, *A Survey of Modern Algebra.* 邦訳一八二頁以下参照。

(2) 同右一九七頁。

(3) 例えば藤原松三郎『行列及び行列式』四五頁以下参照。

(4) 例えば T, Koopmans, ed, *Statistical Inference in Dynamic Economic Models* 中のホーヴェルモーの論文二六一頁以下参照。

二　最小自乗法の統計学的側面

前節においては、線型近似としての最小自乗法を展開したが、統計理論における最小自乗法はその一つの応用にすぎない。ここでは、前節における空間Lの元として次に定義する観察値のベクターを考え、その内積を同じく次に定義する積率でおき換えればよい。n回の観察で、各々 $p+1$ 個の測定値を含むと考えよう。即ち

$$x_{01}, x_{11}, \cdots, x_{p1}; x_{02}, x_{12} \cdots, x_{p2}; \cdots; x_{0n}, x_{1n}, \cdots x_{pn}$$

ここでLの $p+1$ 個の元を

(1) $x_i = [x_{i1}, x_{i2}, \cdots, x_{in}] \quad j=0, 1, \cdots, p$

なるベクター（n 項ベクターの集合は前節初めに述べた線型空間のための条件を満足する）とし、その内積を

(2) $(x_i, x_j) = \frac{1}{n}\sum_{l=1}^{n} x_{il}x_{jl} \qquad i, j=0, 1 \cdots, p$

なる積率と定義して

(3) $x^*_0 = b_1x_1 + \cdots b_px_p \qquad x_0 - x_0^* = e$

で x_0 を近似すると考えれば、前節と同様の結論が得られる。このとき、前節の諸公式はよく知られた回帰分析の公式となるのである。さてここで $x_{p+1} = [1, \cdots, 1]$ として常数項 b_{p+1} を (3) に附加して前節の結果に対応したものを導けば

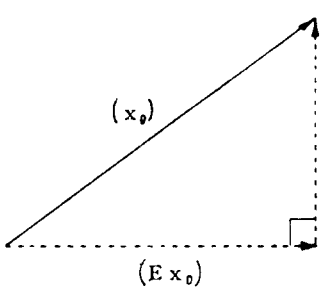
(4) $\frac{1}{n}\sum_{i=1}^{n} e_i = 0$

(5) $\frac{1}{n}\sum_{i=1}^{n} (x_{ii} - \bar{x}_i)\, e_i = \frac{1}{n}\sum_{i=1}^{n} x_{ii}e_i = 0$

$$(\bar{x}_i = \frac{1}{n}\sum_{i=1}^{n} x_{ii};\ i=1, \cdots, p)$$

即ち、最小自乗回帰の残差の平均は0であること及び残差は各々の回帰子（regressor）$x_1, \cdots x_p$ と無相関であることを示す。とりわけ、第二のことは前節での帰結たる直交性に対応するものであることは明らかであり、更に x_i と残差の同時分布に正規性の仮定があればこのことは残差と各回帰子との独立性を意味する。

以上のことを図示すれば第一図の如くなる。最小自乗法は観察値のベクター（x_0）（n 次元）を、最小自乗回帰のベクター（Ex_0）（p 次元）と残差ベクター（e）（$n-p$ 次元）とに分解することでありこのとき（Ex_0）と（e）とのなす角は直角である。n 次元空間のベクター（x_0）がその p 次元の部分空間に射影されたものが（Ex_0）なのである。



第一図

さてそれでは最小自乗法によって決定された係数 b_i は、統計学的にどのような好ましい性質をもつであろうか。この問題に入る前に、母数（ここで b_i のことを統計学でいうように母数といおう）を推定するための観察値の函数たる推定量 T の望ましい性質を若干述べる。[1]

㈠ 不偏性。観察値の大きさの任意の値に対し、Tの期待値$E(T)$が、母数の真の値θに等しいとき。即ち $E(T)=\theta$。このときTを不偏推定量という。

㈡ 一致性。$n\to\infty$ のときTがθに確率収斂するとき、即ち$Pr(|T_n-\theta|>\varepsilon)\to 0$ ($n\to\infty$ εは正任意)。いいかえれば $n\to\infty$ のとき $T_n\to\theta$ の絶対値がεより大きくなる確率が0となるとき、Tをθの一致推量という。

㈢ 有効性。次の二つの条件も満足するTをθの有効推定量という。(1) $\sqrt{n}\,(T-\theta)$ が漸近的に平均0、分散σ^2の規正分布$(N(0,\sigma^2)$と書く)に従って分布する。(2) Tを $\sqrt{n}\,(T-\theta)$ が漸近的に $N(0,\sigma'^2)$ 従って分布するような任意の推定量とするとき $\sigma^2\leqq\sigma'^2$。

尚、推定函数が線型で、不偏で、且つその分散が最小なるものを最良線型不偏推定量という。

ここで本題に戻り、回帰直線の係数の最小自乗推定値のもつ性質を調べよう。ここで視覚的なまぎらわしさを避けるため、前のx_0ベクターをyと、係数を慣例に従ってβと書き換え、観察値の作る線型回帰式を次の如く書く。

(6) $y=X\beta+e$

但し

$$y=\begin{pmatrix} y_1 \\ \vdots \\ y_\alpha \\ \vdots \\ y_n \end{pmatrix},\quad X=\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{p1} \\ \vdots & \vdots & & \vdots \\ x_{1\alpha} & x_{2\alpha} & & x_{p\alpha} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & & x_{pn} \end{pmatrix},\quad \beta=\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_\alpha \\ \vdots \\ \beta_p \end{pmatrix},\quad e=\begin{pmatrix} e_1 \\ \vdots \\ e_\alpha \\ \vdots \\ e_n \end{pmatrix}$$

ここで$e_1\cdots e_n$は互いに独立で平均0、有限な分散σ^2をもつ分布に従うものとする。尚、ベクターや行列の転置は′をつけて示す。[2]

さて最小自乗法によれば、最小にすべきものはeの二乗$(=e'e)$である。

(7) $e'e=(y-X\beta)'(y-X\beta)=y'y-\beta'X'y-y'X\beta+\beta'X'X\beta$

$=y'y-2y'X\beta+\beta'X'X\beta$

ここで $X'X=S$ と書き、β_iにつき微分して0とおけば、

(8) $S\beta=X'y$

(8)から求めたβが最小自乗推定量であるが、これを推定量を示す慣例に従って$\hat{\beta}$で表わそう。尚、二次形式$u'\hat{S}u=u'X'Xu=(Xu)'(Xu)$ は負であり得ないからSは非負である。

さてSが non-singular のときその逆行列 S^{-1} が存在して

(9) $\hat{\beta}=S^{-1}X'Y$

ここで期待値をとれば、

(10) $E(\hat{\beta})=E(S^{-1}X'Y)=E[S^{-1}X'(X\beta+e)]$

$=E(S^{-1}S\beta+S^{-1}X'e)=E(\beta)+S^{-1}X'E(e)=\beta$

故に$\hat{\beta}$は不偏推定量である。

不偏推定量の中で最小の分散をもつことを見よう。このために先づ$\hat{\beta}$の分散を計算する。[3]

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)']=E[(S^{-1}S\beta+S^{-1}X'e-\beta)(S^{-1}S\beta+S^{-1}X'e-\beta)']=E[(S^{-1}X'e)(S^{-1}X'e)']=E(S^{-1}X'ee'XS^{-1})$$

$$=S^{-1}X'[E(ee')]XS^{-1}=S^{-1}X'\sigma^2I_pXS^{-1}$$

$$=\sigma^2S^{-1}X'XS^{-1}=\sigma^2S^{-1}SS^{-1}$$

I_pは$p\times p$の単位行列

(e_iの独立性から)

(11) $=\sigma^2S^{-1}$

さてβの推定のために観察値 y の線型不偏推定量 Ay を作れば これは次の条件を満さねばならぬ。

(12) $E(Ay)=E(AX\beta+Ae)=\beta$

$E(e)=0$ なる故 A は次の条件を満たねばならぬはならぬ。

(13) $AX=I$ （I は単位行列）

ここで $(S^{-1}X'+B)$ なる行列を作り、$BX=O_p$ （O_p は $p\times p$ の零行列）を満足させるようにすれば、これは(13)を満足する。故に

(14) $(S^{-1}X'+B)y$ 但し $BX=y$

は β の任意の不偏推定量である。ここでこの推定量の分散共分散行列を計算すれば、

$$E[(S^{-1}X'+B)ee'(XS^{-1}+B')]=\sigma^2(S^{-1}X'+B)(XS^{-1}+B')$$

(15) $$=\sigma^2(S^{-1}+BB')$$

ここでこの行列の $i\cdot i$ 元が推定量の分散を表わすのであるが、BB' の $i\cdot i$ 元は B の第 i 行の元の二乗の和であるから、その行のすべての元が 0 でない限り正である。従って、最小自乗推定量(9)以外のどの線型不偏推定量もより大なる分散をもつ。即ち最小自乗推定量(9)は最良線型不偏推定量である。尚、このことには、残差の平均零その分散が有限一定、そして残差が互いに独立ということ以外の仮定はないということを注意しておかねばならない。

この他、最小自乗推定量につき述べるべき性質もあるがここでは省略する。[4]

(1) 例えば S. S. Wilks, *Mathematical Statistics.* 邦訳二〇五頁以下参照。

(2) 行列の諸演算については例えば藤原前掲書参照。

(3) 正確には $\hat{\beta}$ の分散共分散行列である。

(4) 例えば T. Koopmans & W. C. Hood, "The Estimation of Simultaneous Linear Economic Relationships", in *Studies in Econometric Methods*, 1953 を見よ。

三　経済的関係式推定への適用における一つの問題

第二節で我々は最小自乗法の原理とその長所とを説明した。最小自乗法は元来実験的観察における誤差論から生れた。そしてそれが適用された多くの分野においては、後に説明する如く、その長所を生むべき原理的前提条件の満足が容易である。しからば経済的データーにおける誤差とストカスチックな経済モデルについてはどうであろうか。我々は先づこのことと、シカゴ学派の最小自乗法批判との関係を調べることにしよう。

前にも断った如く、我々は時系列や観察誤差の問題を他の機会に譲っているので、残る最大の問題はいわゆる同時的関係（連立方程式法）の問題である。経済データーは、一般に同時的諸関係の体系から生成されるものであって、これに対し最小自乗法による伝統的な接近法は、ホーヴェルモーの言葉によれば、単一方程式（single-equation）法であって、経済変数の相互依存関係を無視したものである、というのがシカゴのエコノメトリシャンの主張の基本である。我々はここで最小自乗法と同時方程式法との関係についてブロンフェンブレンナーの論文[1]を取上げてみよう。この問題については、彼女の論文は、具体的且つ明快であり、又二方程式モデルを取扱っていても一般的推論を可能ならしめるものであって、代表的に取上げることが出来るからである。

彼女の目的は、方程式体系内の一つの方程式に直接に最小自乗

を適用したときに生ずるバイアス（偏り）の源泉とその大きさをつきとめることにある。そしてそのために次の二つの方程式体系を考える。

(1) $y_1+\beta_{12}y_2+\gamma_1z_1+\varepsilon_1=u_1$

(2) $\beta_{21}y_1+y_2+\gamma_2z_2+\varepsilon_2=u_2$

ここで、y_1、y_2、z_1及びz_2は観察可能であり、観察値（y_{1i}, y_{2i}, z_{1i}, z_{2i}）（$i=1, 2, \cdots, n$）が得られ、u_1、u_2は観察出来ないランダムな攪乱である。又yは内生変数、zは外生変数とする。即ち、zは体系外で決定され、内生変数に影響を与えるが、それからは影響を受けない。このzについては次の三つのことを仮定しておく。㈠それは攪乱項uとは独立である。いいかえれば、uの標本変動はy_1及びyのみに基く。㈡標本の大きさが無限に大きくなるとき、その二次の積率は、有限な値をもつか、ないしは有限な値に確率収斂する。㈢これらの積率の行列は non-singular である。（これは複共線性の存在する一つの場合を除くことになる。）次に、uについては、それは観察誤差ではなく、各々小さい影響をもつ多の省略され変数の総効果であって、u_1とu_2は二変数正規分布をなし、平均に0、分散各々σ_1^2、σ_2^2、共分散σ_{12}である。ここで、(1)、(2)は共に ideutification の必要且つ十分条件を満している（$\gamma_1\neq 0, \gamma_2\neq 0$ ならば）。又、

(3) $y_1+\beta_{12}y_2+\gamma_1z_1+\varepsilon=0$

なる平面はy、y及びzの間の「真の」又は「攪乱されざる」関係を表わしていると考えて良い。(2)式についても同様である。

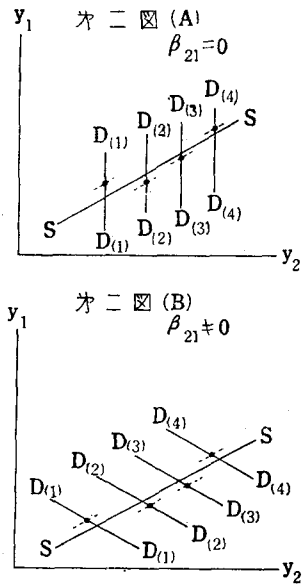
結論的に二つのバイアスの根源が証明されるのであるが、先づその一つ一つを考察するために、モデルに少しづつ制限条件を変えながら課していく。先づ(1)の方程式の推定のみを問題として取扱ってもよいから、$\gamma_1=0$ としても(1)の identifiability は変らぬ故 $\gamma_1=0$、$\gamma_2=\gamma$ とする。

第一のバイアス。（β_{21}から来るもの。即ちy_2とu_1との間の相関関係から来るもの。）これだけ切離して考えるため $\sigma_{12}=0$ とおく。

(4) $y_1+\beta_{12}y_2+\varepsilon_1=u_1$

(5) $\beta_{21}+y_2+\gamma z+\varepsilon_2=u_2$　　$\sigma_{12}=0$

ここで一定のzに対しては、(4)と(5)はy_1とy_2の二次元的関係式に帰する。異なる値のzに関して、第二図の如くDD直線（二つの方程式(1)(2)を以後例としてそれぞれ供給函数及び需要函数と考えてこの記号を用いる）については、平行な直線群が出来る。観察点（y_{1i}, y_{2i}）に対応する攪乱のない真の点（$\bar{y}_{1i}$, $\bar{y}_{2i}$）はSS直線とDD直線群との交点として決定される。



ここで我々は(A)と(B)との間の重要な相違に気付くであろう。$u_1\neq 0$ なるときはSS直線が勾配β_{12}はそのままで平行移動する

だけであるから、$u_2=0$ と仮定すれば、観察される点は、平行移動したSS直線（図の点線）とDD直線との交点に定まる。(A)即ち $\beta_{21}=0$ の場合には、観察される点の y_2 軸座標はそのままで y_1 軸座標に u_{1i} が加えられるだけである。従って、このとき残差平方和を y_1 軸の方向に最小にするという伝統的方法は全く妥当であろう。しかし、(B)の場合には u_{1i} の標本変動の効果は y_1、y_2 の両方に存在する。従って、この時にはどちらか一方の方向にのみ残差平方和を最小にすべき理由はないし、どちらの方向に最小にしても β_{12} の不偏推定値が得られると考えるべき理由はない。y_1 の方向に最小にした時のバイアスは次の如くにして得られる。β_{12} の最小自乗推定値 $\beta^{(1)}_{12}$ は次の値に確率収斂する。その極限値を表わすのに *plim* と書くと、

(6) $$\beta_{12}^{(1)}=\underset{n\to\infty}{plim}\, b_{12}^{(1)}=\underset{n\to\infty}{plim}\frac{m^{o}_{y_1y_2}}{m^{o}_{y_2y_2}}$$

但し

(7) $$m_{y_jy_k}=\frac{1}{n}\sum_{i=1}^{n}(y_{ji}-\bar{y}_j)(y_{ki}-\bar{y}_k)=\frac{1}{n}\sum_{i=1}^{n}y_{ji}y_{ki}-\frac{1}{n^2}\sum_{i=1}^{n}y_{ji}\sum_{i=1}^{n}y_{ki}\left(\bar{y}_j=\frac{1}{n}\sum_{i=1}^{n}y_{ji},\ \bar{y}_k=\frac{1}{n}\sum_{i=1}^{n}y_{ki}\right)$$

他方(4)、(5)から誘導形を作ると、

(8) $$y^1=\frac{1}{1-\beta_{12}\beta_{21}}(\beta_{12}\gamma z+\beta_{12}\varepsilon_2-\varepsilon_1+u_1-\beta_{12}u_2)$$

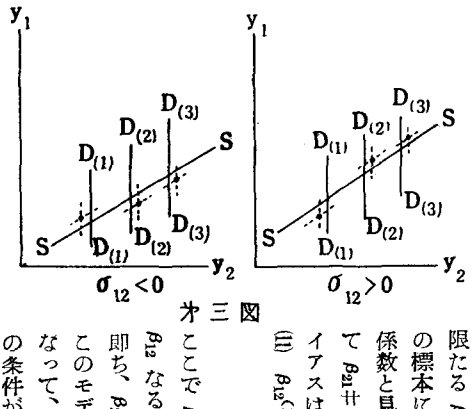
(9) $$y_2=\frac{1}{1-\beta_{12}\beta_{21}}(-\gamma z+\beta_{21}\varepsilon_1-\varepsilon_2-\beta_{21}u+u_2)$$

仮定から、u_1 と u_2、u_1 と z、u_2 と z の二次の積率の確率極限は0であり、u_1、u_2 及び z の分散の確率極限は σ_1^2、σ_2^2 及び μ_{zz} であるから、(8)、(9)を(6)に代入して、確率変数の函数の積率の計算法を適用して計算すれば、

(10) $$\beta_{12}^{(1)}=\frac{\beta_{21}\sigma_1^2+\beta_{12}\sigma_2^2+\beta_{12}\gamma^2\mu_{zz}}{\beta_{21}^2\sigma_1^2+\sigma_2^2+\gamma^2\mu_{zz}}$$

標本積率は母集団の積率の一致且つ漸近的に不偏な推定値であり、又積率の有理函数を取扱っているのであるから、$b_{12}^{(1)}$ の確率極限たる $\beta_{12}^{(1)}$ を大きさ無限大の標本における y_1 の y_2 への回帰係数と見ることが出来る。従って $\beta_{21}\neq 0$ の場合の漸近的なバイアスは

(11) $$\beta_{12}^{(1)}-\beta_{12}=\frac{(1-\beta_{12}\beta_{21})\beta_{21}\sigma_1^2}{\beta_{21}^2\sigma_1^2+\sigma_2^2+\gamma^2\mu_{zz}}$$



第三図

ここで $\beta_{21}=0$ ならば $\beta_{12}^{(1)}=\beta_{12}$ なることは明らかである。即ち、$\beta_{21}=0$, $\sigma_{12}=0$ のとき、このモデルでは y_2 が u_1 と独立となって、最小自乗法適用のための条件が満されるのである。

第二のバイアス。（σ_{12}即ちu_1とu_2との間の相関々係から来るもの。）この場合もβ_{21}の効果と分離して考察するために$\beta_{21}=0$とする。（第三図参照）。ここで前と同様に考えれば、

(12) $$\beta_{12}(1)=\frac{\beta_{12}\sigma_2^2+\beta_{12}\gamma^2\mu_{zz}-\sigma_{12}}{\sigma_2^2+\gamma^2\mu_{zz}}$$

バイアスは

(13) $$\beta_{12}(1)-\beta_{12}=\frac{-\sigma_{12}}{\sigma_2^2+\gamma^2\mu_{zz}}$$

$\sigma_{12}=0$ならば勿論$\beta_{12}(1)-\beta_{12}=0$となり、前と同じ結論に達する。

さて以上と同様な考察を一般的な二方程式モデル(1)(2)について行ってみよう。前と同じように計算すれば、

(14) $$\beta_{12}(1)=\frac{\beta_{21}\sigma_1^2-(1+\beta_{12}\beta_{21})\sigma_{12}+\beta_{12}\sigma_2^2+\beta_{12}\sigma_2^2\mu_{z2}{}^{\circ}z_2{}^{\circ}}{\beta_{21}{}^2\sigma_1^2-2\beta_{21}\sigma_{12}+\sigma_2^2+\gamma_2{}^2\mu_{z2}{}^{\circ}z_2{}^{\circ}}$$

$$\mu_{z2}{}^{\circ}z_2{}^{\circ}=\mu_{z2}z_2-\frac{\mu^2_{z1z2}}{\mu_{z1}z_1}$$

ここでは。z_2のz_2に対する最小自乗回帰の残差である。この場合の漸近的なバイアスは、

(15) $$\beta_{12}(1)-\beta_{12}=\frac{(1-\beta_{12}\beta_{21})(\beta_{21}\sigma_1^2-\sigma_{12})}{\beta_{21}{}^2\sigma_2-2\beta_{21}\sigma_{12}+\sigma_2^2+\gamma_2{}^2\mu_{z2}{}^{\circ}z_2{}^{\circ}}$$

である。さてここで(15)を分析してみるとy_2の変動が根本的に重要な役割を果していることが次の如くにして分る。

$$z_2=\frac{\mu_{z1z2}}{\mu_{z1}z_1}z_1+z_2{}^{\circ}$$

であるから、これをy_2の誘導形

(16) $$y_2=\frac{1}{1-\beta_{12}\beta_{21}}(\beta_{21}\gamma_1z_1-\gamma_2z_2+\beta_{21}\varepsilon_1-\varepsilon_2-\beta_{21}u_1+u_2)$$

に代入すれば

(17) $$y_2=\frac{1}{1-\beta_{12}\beta_{21}}\left[\left(\beta_{21}\gamma_1-\gamma_2\frac{\mu_{z1z2}}{\mu_{z1}z_2}\right)z_1-\mu_2z_2{}^{\circ}-(\beta_{21}u_1-u_2)\right]$$
$$+\eta=y_2{}^{(z)}+y_2{}^{(z^{\circ})}+y_2{}^{(u)}+\eta\quad \eta=\text{常数}$$

これは変動の原因に対応するようにy_2を分解したものと解し得る。ここで第二項で第二項と第三項

(18) $$-\frac{\gamma_2}{1-\beta_{12}\beta_{21}}z^{\circ}-\frac{\beta_{21}u_1-u_2}{1-\beta_{12}\beta_{21}}=y_2{}^{(z_2{}^{\circ})}+y_2{}^{(u)}$$

はy_2のz_1と無相関の成分と考え得る。又第三項$y_2{}^{(u)}$はz_1、zの両方と独立なy_2の成分で、回帰式によっては「説明されず」に残されたyの成分である。(15)の分母子を$(1-\beta_{12}\beta_{21})^2$で割れば、分子は$u_1$と$y_2{}^{(u)}$との共分散の符号を変えたものとなる。従って、バイアスは$\mathrm{cov}(u_1, y_2{}^{(u)})$に比例する。又$y_2$の三つの成分は互いに独立であるから、

(19) $$\mathrm{var}(y_2)=\mathrm{var}(y_2{}^{(z)})+\mathrm{var}(y_2{}^{(z^{\circ})})+\mathrm{var}(y_2{}^{(u)})$$

(15)の分母は$(1-\beta_{12}\beta_{21})^2$で割れば(18)の分散となるから、バイアスは、$\mathrm{var}(y_2{}^{(z^{\circ})}+y_2{}^{(u)})$と反比例することが分る。以上を要約すれば、次の公式で表わされる。

(20) $$\beta_{12}(1)-\beta_{12}=\frac{-\mathrm{cov}(u_1, y_2{}^{(u)})}{\mathrm{var}(y_2{}^{(z_2{}^{\circ})}+y_2{}^{(u)})}$$
$$=\frac{-\mathrm{cov}(u_1, y_2{}^{(u)})}{\mathrm{var}(y_2{}^{(z_2{}^{\circ})})+\mathrm{var}(y_2{}^{(u)})}$$

ブロンフェンブレンナーはこのように最小自乗推定を直接的に方程式体系内の一方程式に適用した場合のバイアスを計算したのであるが、これと同様な計算が、より簡単な場合についてではあるがウォルトによって行われている。[3] ただ彼の場合との計算が、逆に最小自乗法擁護の重要な一点となる定理の説明例として行わ

れていることは興味深いことである。以下それをここに紹介するのは、最小自乗法をめぐる諸論者の立場がそれぞれどのような考え方を代表しているかを知るのに大いに役立つと考えられるからである。

もし攪乱項が無視し得る程小なるものであるか、又は回帰子（前述の式では y と z）が攪乱項と無相関であれば、最小自乗回帰係数が不偏であることは明らかである。ウォルトはこの二つの条件が相互補強的に働いて最小自乗回帰数のバイアスを減少せしめるということを云わんとするのである。

母集団回帰関係、最小自乗回帰関係を

(21) $y=\beta x+z^*$　$y=bx+z$

とする。z^*、z は各々残差である。又簡単のために変数は皆平均からの偏差をとってあるから $E(x)=E(y)=0$ である。すると

(22) $$E(b)=\frac{E(xy)}{E(x^2)}=\frac{E(x(\beta x+z^*))}{E(x^2)}=\frac{\beta E(x^2)+E(xz^*)}{E(x^2)}$$

$$=\beta+\frac{E(xy^*)}{E(x^2)}=\beta+r_{xz}^*\frac{\sigma_{z^*}}{\sigma_x}$$

$$\left(r_{xz}^*=\frac{E(xz^*)}{\sqrt{E(x^2)E(z^{*2})}}\right)$$

これで証明されたことは、$\sigma(z^*)$ と r_{xz}^* とが一次のオーダーの小さであれば、バイアスは二次のオーダーの小ささとなるということである。これをウォルトは Proximity Theorem と呼んでいるが、この定理に基づき、実際的に残差の分散は回帰子の分散に比べて大体非常に小さいこと、及び回帰子の取捨選択によってその残差との相関を小ならしめ得ることとを考慮して最小自乗法の経済分析における妥当性健全性を主張するのである。(Proximity Theorem を第一節での数学的形式で証明することは容易であるが省略する。)

大方紙数も尽きてしまったので、終りの部分は前の部分と比較すれば尚論じ足りない恨みがあるが、この辺で以上の問題を整理してみれば、結局、問題は、最小自乗法の原理から来る残差と回帰子との直交性、無相関性に帰することが明瞭である。とすれば、真の問題は最小自乗法それだけを切離して考えるべきものではなく、その原理的前提の実現が、経済的関係の分析において困難であるという点にあるというよう。物理学、医学、農学等の分野においては、一般に、回帰関係の統計的導出に当って実験的な観察が可能であり、さし当って問題とならぬ要因の固定ないしは中立化が可能であるから、そのような実験からは最小自乗法の前提を十分満足するデーターが得られる。しかるに、経済現象について得られるデーターはかかる性格のものではなく、同時的諸関係の切離しが不可能であって、色々な力が綜合的に働いて出来たものと考えらるべきである。即ちそのデーターは非実験的データーである。しかも、経済分析に統計を用いる場合、二次的統計（他の目的のために得られたもの）を用いる場合が圧倒的に多いのであるから、理論モデルの構成の仕方とデーター（例えそれが正確なものであっても）の選び方によっては、最小自乗法の原理的前提が満されるのに程遠い場合が多いのである。

この事態に対処するのに二つの方向がある。一つは統計的推定方法からの適応、他は経済理論モデルからの適応である。我々は、共に最小自乗法の原理を認識しながらも、同時的経済関係の分析

における最小自乗法への評価が異なった二つの立場を対比してみたのであるが、両者の分岐点はまさにこの適応の仕方にあるといってもよいであろう。シカゴ学派は可能な限り統計的推定方法からの適応を考えて、同時関係式の最尤推定法へと進んだ。これに対してウォルトは経済モデルからの適応を主張して、逐次モデルに拠り、推定方法としては仮定の最も少ない最小自乗法を健全なものとした。この二つのどちらかにここで軍配をあげようとするには余りにも多くを論じ残している。最小自乗法を全く又はほぼバイアスなしに適用出来る誘導形や逐次モデルについて、更に最尤法の性格について等である。しかし、いづれにせよ次のようなウォルトの言葉は非常に示唆に富んだものといえよう。(4)『回帰法に対する最近の批判の多くは、疑もなく、近代統計学の諸方法の勝利を経済統計学社会統計学にまで持ち込もうとするものである。しかし、このような計画は大きな困難にぶつかる。非実験的データへの適用においては、我々は攪乱や残差の適当なスペシフィケーションのために必要な経験と知識をもっていないからである。……よりよい方法が実際に適用される時には、その利点は常に、最小自乗法の、根抵にある仮定が非常に弾力的であること及び数値計算がごく簡単であるという実質的利益と、秤にのせて比べられなければならない。』

(1) J. Bronfenbrenner; "Sources and Size of Least-Squares Bias in a Two-Equation Model," in *Studies in Econometric Methods.*

(2) identification の分り易い解説としては *Studies* 中の第二章を見よ。

(3) H. Wold, *Demand Analysis*, p. 37.

(4) ibid., p. 59.