計量経済学における a priori information について

西　川　欽　也

目　　次

§ 1.　計量経済学における論理実証主義

いくつかの抽象的な仮定から出発して，ある結論を導びく数学的な理論が，経済理論の名に値するためには，そこで用いられる諸概念が経済学的な意味を附与された term で定義されているばかりでなく，その理論が現実の経済メカニズムの働きをなんらかの程度に表現しているという意味での実証性をもっていなければなるまい。

このような実証性を確保するためには，現実の経済メカニズムについての観察結果をもち，これとわれわれの理論とをつき合わせる手続が必要になる。Samuelson, Koopmans ならびに Stone [5] によって与えられた周知の定義にしたがえば，このような観察結果を提供するのが経済統計学であり，これを理論とつき合わせるのが計量経済学の役割であるということになろう。

しかし，われわれの観察結果が，自然科学の多くの領域に見られるよう

な，一定の管理された（既知の）条件の下に遂行された実験の結果でなく，仮りに調査対象や調査方法を明確に限定した統計調査の結果であるとしても，観察結果の実現過程になお多くの未知の，しばしば不確定の要素を含む統計データであるということは，計量経済学が提供する経済理論の実証性に大きな制約をもたらすことは否めない。そして，このことをさらにつきつめていくと，果して経済理論は実証性をもちうるかという根本的な設問にすら到達しかねないのである。

このような問題に対する計量経済学の立場からする解答は，基本的には，Haavelmo [3] によって提供された次のような考え方に代表される。すなわち，まず検証しようとする仮説に対して，その仮説を検証できるようなモデルをつくり，このモデルを仮説を検証すべく提供された観察結果＝統計データを発生する実験計画とみなすことである。いうまでもなく，われわれのデータ自体は passive に与えられたものであるから，この場合の実験はあくまでも ideal なものであり，研究者にとっての a priori な information の総括を示すものでしかない。しかし，このように考えることによって，ともかくわれわれが設定したモデルの下で，当の仮説に対する統計的推論を適用することが可能になり，また，その意味で仮説の検証が与えられることになる。ひとたびこのような手続きによって仮説が受容されたならば，この仮説を予測の基礎として採用することができ，さらにこの予測結果によって，われわれは必要ならばモデルを改善することができるであろう。そして，このようなアプローチが実際的にも fruitful であったというのが Haavelmo の主張であり，さらに Haavelmo 以後の膨大な経験の集積の上に立って，今日，多くの計量経済学者が explicit にもせよ implicit にもせよ依拠しているところの思想であるといってよかろう。

この思想が論理実証主義の計量経済学における counterpart であることは一目して明瞭である。しかし，実際的な計量経済学者たちは彼らの仕事のなかで論理実証主義を忠実に follow しているわけではない。計量経済学的

分析の多くが巨視的経済予測や政策効果の計測といったきわめて実用的な目的に利用されてきた事情とあいまって，既に Haavelmo の中にあった予測によるモデルの修正といった考え方は，よりプラグマティックな形で拡大されてきた。

すなわちモデルのなかに集約されている経済メカニズムに対する a priori な認識の絶対的な真否よりも，あるいはまた理論的な関心から出発して，ある仮説を検証し，そのことを通して理論の正否を判定するための効果的な実験計画としてモデルを組み立てるということよりも，当のモデル自体の予測性能の経験的な良さがモデル・ビルディングの基準とされ，もっぱら予測性能を高めるという目的から，必ずしも充分な理論的な吟味なしに変数が選択され，関数の型が決定されるという傾向が生まれている。

筆者はこのような傾向を非難するつもりはまったくない。経済メカニズムが現実に安定した法則性をもって働き，推移するという認識に立つならば，後述するように予測性能の良さはモデル選択の有力な基準となりうる。

Haavelmo 自身認めているように，モデルが表現している a priori information はすべてがすべて確実なわけではない。というよりも本来の理論検証という観点からは，ある検定の対象となる仮説の論理的な前提となるべき理論を，確実なものと仮定してモデルは組み立てられていると考えねばならない。だからこそ予測結果はモデル・ビルディングに feed back されねばならないのであり，予測はモデルの妥当性に対するテストなのである。しかも予測性能のより高い経験的なモデルに論理的な根拠を与える経済理論を，われわれが常にもっていたり，容易につくり出せるというわけではない以上，Haavelmo によって排撃された「データから理論を考える」という思考様式にも，正当な復権が認められなければなるまい。

モデルの妥当性に関するテストは，予測あるいは所与のデータそのものによって行なういくつかのものが知られている。しかし，すべての統計的テストがそうであるように，これらのテストもすべてそれぞれの前提条件をも

ち，この限られた条件の下でのテストである。これらの前提条件は当のテストにとっては確実な information となっていなければならない。もし，ここになにがしかの不確実性が混在しているならば，このテスト結果をモデル・ビルディングに feed back させる過程は，より慎重に吟味されなければなるまい。もちろん前提条件を異にするいくつかのテストをやって，その個々のテスト結果を個々ばらばらに考慮してモデルを修正することが，無意味であり，往々にしてモデル の 改悪 を 結果するであろうことはいうまでもない。(もっとも前提条件 の 異なるテストを行なうこと自体は，モデルのいわば頑健性 (robustness) を検知することになるから，無意味なわけではない。)

それゆえ，いくつかの alternative なモデルのなかから，より plausibility の高いモデルを選択するという問題，あるいはまた所与のデータなり，予測結果によって与えられるモデル自体に対する evidence をモデルの選択なり，モデルの修正なりにどのように利用するかについて統一的な手法を考えることは，実際的見地からも，きわめて重要なことと思われる。しかし，この問題は，ある意味では計量経済学の根抵にかかわる予想外 に 困難 な 問題であり，計量経済学の領域でもほとんど未開拓な分野である。したがって筆者もここで野心的な contribution を提出することは到底できない。本稿は，ただこの問題の性格について若干の吟味を行ない，未開の領域に踏み込むための手がかりを模索することを意図するに過ぎない。

§ 2.　モデル選択問題とモデルの識別可能性

予測結果にもとずいてモデルを修正する問題も，いったん修正を加えられたモデルはもとのモデルとは別のモデルになるのだから， alternative なモデルのなかから一つのモデルを，そのときわれわれがもっている情報にもとずいて選択する問題の一つと考えてよい。この種の問題をここではモデル選択問題 (model preference problem) と呼ぶことにする。

一般にモデルはいくつかの内生変数 (たとえばH個の) $y_1, y_2, \cdots, y_H$ に

関する連立方程式

(1)　$g_i(y_1, \cdots, y_H, x_1, \cdots, x_K, \boldsymbol{\theta}_{1i}, \cdots, \boldsymbol{\theta}_{Li}, u_i)=0,\ \ i=1, \cdots, H$

によって表わされる。ここに $x_1, \cdots, x_K$ は外生変数または先決変数であって，簡単化のため通常なされているように，これらはすべて確率誤差を伴わない確定変数であると仮定する。$\boldsymbol{\theta}_{1i}, \cdots, \boldsymbol{\theta}_{Li}$ は第 i 方程式に含まれている既知，または未知の構造パラメーターであり，u_i はこの方程式の random disturbance であり，ある確率分布に従う確率変数である。この確率分布は通常その分布パラメーターをもっているが，このパラメーターは先の構造パラメーターには含めないことにする。また，ここでは方程式の数は内生変数の総数に等しい H 個としてあるが，一般には必ずしも H 個である必要はない。ただ segmentable ないくつかの確定的な関係式（例えば定義式）を排除して考えると，モデルは多くの場合(1)を y_i を左辺に引き出すように書き改めて，explicit に y_i を決定する H 個の方程式の形で示されることが多いので，便宜上ここでは方程式の数をH個としたので，以下の議論はこれがH以外の任意の正の整数であっても影響されない。また，実際の問題では，符号条件等，構造方程式に表現されないいくつかのわれわれの a priori information を表わす確定，または不確定の制約条件が(1)に付加されることが多いが，ここではこのような制約条件はないものと仮定して議論を進める。

さて(1)は $y_1, \cdots, y_H$ を $x_1, \cdots, x_K, \boldsymbol{\theta}_{1_1}, \cdots, \boldsymbol{\theta}_{L_H}, u_1, \cdots, u_H$ が与えられたとき一義的に決定するに充分な consistency をもつと仮定しよう。（この仮定が充されているモデルを consistent なモデルと呼んで，以下の考察はこの consistent なモデルについて行なうと約束してもよい。）すると $y_1, \cdots, y_H$ は外生変数，構造パラメーター，ならびに random disturbance の確率分布によって定められたある分布法則をもつ，一般に確率的に相互に従属的な確率変数となる。したがって $y_1, \cdots, y_H$ をこのような確率変数としておけば，(1)の random disturbance はわれわれが random disturbance の分布に関心をもつ場合を除いては，省略してモデルを表現してもよい。実際，以下の

モデル選択問題の formulation では，random disturbance の分布が explicit にとり上げられることはあまりないから，これをモデルの表現から除くことにする。

表現をさらに簡潔にするため内生変数 $y_1, \cdots, y_H$, 外生変数 $x_1, \cdots, x_K$, 構造パラメーター $\boldsymbol{\theta}_{1_1}, \cdots, \boldsymbol{\theta}_{L_H}$ をそれぞれH次元，K次元，およびL次元（但し $L=\sum_{i=1}^{H} L_i$）のベクトル $\boldsymbol{Y}$, $\boldsymbol{X}$, および $\boldsymbol{\Theta}$ で表わすことにしよう。すると(1)はH＋K次元ベクトル $(\boldsymbol{Y}, \boldsymbol{X})$ をH次元ベクトルに $\mathfrak{G}$ を媒介にして変換するある変換 $\mathfrak{G}$ について，

(2)　$\mathfrak{G}(\boldsymbol{Y}, \boldsymbol{X} \mid \boldsymbol{\Theta})=\mathbf{0}$

が満足されることを示す。この等式は $\boldsymbol{Y}$ が確率変数であるから，$\boldsymbol{X}, \boldsymbol{\Theta}$ を所与とするならば，その条件の下での $\boldsymbol{Y}$ の確率法則に等しい確率，つまり $P=dP(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\Theta})$ なる確率で成立する。

それゆえ，われわれはモデルを次のように定義しよう。

定義 1.（モデル）H，K，L を任意の正の整数とするとき，それぞれH次元，K次元，L次元の実ベクトル $\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\Theta}$, ならびにH次元のゼロ・ベクトル $\mathbf{0}$ について，変換 $\mathfrak{G}$: $R^{H+K} \rightarrow R^H$ が確率 $dP(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\Theta})$ で(2)を満足し，かつ $\boldsymbol{Y}$ があるH次元の同時確率分布法則に従う確率変数，または確率過程となっているとき $\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\Theta}, \mathfrak{G}$ の組合せ $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\Theta}, \mathfrak{G})$ をモデルといい，あらゆる H，K，L にわたってこうして定義されたモデルを集めた集合をモデル空間 $\mathbf{M}$ で表わす。

ここで定義したモデル空間 $\mathbf{M}$ はきわめて一般的なものであり，それゆえ実りある implication もほとんどもつものではない。より進んだ考察のためには，以下の議論で実際やっているように，$\boldsymbol{Y}$ をすべてのモデルにわたって共通にしておくとか，変換 $\mathfrak{G}$ を $(\boldsymbol{Y}, \boldsymbol{X})$ の線形変換に限定するとか，H，K を一定にしておくといったような限定を加え，こうした限定によって与えら

れるモデル空間の部分空間に属するモデルについて議論することが必要であることは言うまでもない。

任意の二つのモデルにおいて，それぞれの内生変数 $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$ が共通成分をまったく含んでいないならば，これらのモデルはそれぞれまったく別の現象を説明するモデルであって，両者の間で，なんらかの共通のベースに立って優劣を争うということは考えられない。もっとも経済学的な観点からは，ある共通の定性的な問題を異種の変数を使って表現するという意味での比較が問題になりうるかも知れないが，この場合，抽象的には，他方の変数はいま一方の変数に，少なくとも確率的に変換可能でなければならず，この意味で $\boldsymbol{Y}_1$ と $\boldsymbol{Y}_2$ が共通成分をもつモデルに還元して比較することが可能なはずである。それゆえ，次に二つのモデルの比較可能性を次のように定義する。

定義 2.（比較可能なモデル）　二つのモデル（$\boldsymbol{Y}_1, \boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1$）と（$\boldsymbol{Y}_2, \boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2$）において恒常的に $\boldsymbol{Y}_1=\boldsymbol{Y}_2=\boldsymbol{Y}$ であるとき，これら二つのモデルはたがいに比較可能（comparable）であるという。また $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$ の成分中に，例えば $\boldsymbol{Y}_1$ の第 i 成分 y_{1i} と $\boldsymbol{Y}_2$ の第 j 成分 y_{2j} とが恒常的に $y_{1i}=y_{2j}$ であるという意味で，共通な成分が含まれているならば，これら二つのモデルはこの共通成分に関して部分的に比較可能（partially comparable）であるという。

なお，comparable なモデルの全体が作る集合を comparable モデル空間といい，partially comparable なモデルの全体がつくる集合を partially comparable モデル空間という。いうまでもなく comparable モデル空間は partially comparable モデル空間の部分空間である。

以下では上の定義の意味で comparable なモデルについてモデル選択問題をとりあつかう。（すなわち partially comparable なモデルのモデル選択問題はとりあげない。）

しかし，モデル選択問題に入るに先立って，モデルの識別可能性の問題

をとりあげておくのが便利である。いわゆる構造識別 (structure identification) の問題と区別されるモデル識別 (model identification) の問題を，はじめてとりあげたのは筆者の知る限りでは福地[2]である。しかし，彼の数学的 formulation は充分な厳密性を欠き，また，彼がモデルが「強く識別可能*」であるための十分条件として指摘した条件は誤っていると思われる。

彼の識別可能性をわれわれの formulation に従って言えば，二つのモデル $(\boldsymbol{Y}_1, \boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{S}_1)$ と $(\boldsymbol{Y}_2, \boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{S}_2)$ とにおいて，それぞれ $\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$ を動かしたときに得られる $\boldsymbol{Y}_1$ および $\boldsymbol{Y}_2$ の分布族が，互いに共通な分布を含まないとき二つのモデルは互いに他方に対して「強く識別可能」となり，それぞれの分布族が各々1個以上の共通でない分布をもっているならば，「弱く識別可能」であることになる。そして彼は，「強く識別可能」であるための一つの十分条件は，$\boldsymbol{Y}_1$ と $\boldsymbol{Y}_2$ または $\boldsymbol{X}_1$ と $\boldsymbol{X}_2$ が少なくとも1個以上の共通でない成分をもつことであるとしている。

しかし，$\boldsymbol{Y}_1$ と $\boldsymbol{Y}_2$ が少なくとも1個以上の共通でない成分をもつということは，筆者の見解では，二つのモデルは少なくとも部分的にはまったく別の現象を説明するモデルとなっているのだから，統計的な次元でこれを比較することは，$\boldsymbol{Y}_1$ と $\boldsymbol{Y}_2$ の共通成分に関する部分を除いては，およそ無意味である。彼の定義にしたがえば，もちろんこれは識別可能にはなるだろうが，モデル選択の問題に関連した意味での model identification を考える上では，われわれが関心を寄せる必要のない case に属する。また，筆者が理解するモデルの識別可能性について言えば，二つのモデルの comparable な部分をとり出して考えるとき，この部分モデルが識別可能であるかどうかの問題は，依然として未解決のままに残されるであろう。これに対して $\boldsymbol{X}_1$ と $\boldsymbol{X}_2$ が共通でない成分をもつということは，必ずしも二つのモデルを識別可能には

* 福地[2]では「識別可能」ということばを使わず「認定可能」ということばを用いているが，両者はともに identifiable の訳であって，まったく同じ意味で使われる。ここでは筆者の習慣にしたがい彼の「認定可能」をすべて「識別可能」という用語でおきかえる。

しない。なぜならば，いま内生変数 y_1, y_2 と外生変数 x_1, x_2, x_3, x_4 を考え

$$\left.\begin{aligned} y_1+\beta_{12}y_2+\gamma_{11}x_1+\gamma_{12}x_2+u_1=0 \\ \beta_{21}y_1+y_2+\gamma_{21}x_1+\gamma_{22}x_2+u_2=0 \end{aligned}\right\}$$

をモデルⅠ，

$$\left.\begin{aligned} y_1+\beta_{12}y_2+\gamma'_{11}x_3+\gamma'_{12}x_4+u_3=0 \\ \beta_{21}y_1+y_2+\gamma'_{21}x_3+\gamma'_{22}x_4+u_4=0 \end{aligned}\right\}$$

をモデルⅡとしよう。x_1, x_2, x_3, x_4 は確定変数であるから，もし，

$$\text{(3)}\quad \left.\begin{aligned} \gamma_{11}x_1+\gamma_{12}x_2=\gamma'_{11}x_3+\gamma'_{12}x_4 \\ \gamma_{21}x_1+\gamma_{22}x_2=\gamma'_{21}x_3+\gamma'_{22}x_4 \end{aligned}\right\}$$

が恒等的に成立し，かつ u_1 と u_3, u_4 と u_5 が同一の分布法則をもつならば（そして，このことは(3)が成立する限り真である。）モデルⅠとモデルⅡは $z_1=\gamma_{11}x_1+\gamma_{12}x_2$, $z_2=\gamma_{21}x_1+\gamma_{22}x_2$ なる二つの変数をあらためて外生変数としてとった

$$\left.\begin{aligned} y_1+\beta_{12}y_2+z_1+u'_1=0 \\ \beta_{21}y_1+y_2+z_2+u'_2=0 \end{aligned}\right\}$$

という形のまったく同じ形のモデルを表わしており，u'_1 は u_1, u_3 の分布法則，u'_2 は u_2, u_4 の分布法則に等しく，したがってこれから定まる y_1, y_2 の同時確率分布はモデルⅠ，モデルⅡが与える分布法則と等しく，したがって，モデルⅠが与える y_1, y_2 の同時確率分布とモデルⅡが与える同時確率分布とは同一になり，したがって，モデルⅠとモデルⅡは識別不能である。

一方(3)は x_1, x_2 がそれぞれ x_3, x_4 の一次結合によって与えられることを保証するにとどまり，x_3, x_4 が x_1, x_2 と異なる変数であること，つまり x_3 が x_1 または x_2 のいずれかと，x_4 が残りの一方の変数と恒常的に等しい変数であることを要求しない。彼がこのような case をどのように考えているのかは明らかでないが，少なくとも彼が［2］で示した上記の十分条件なるものは，彼の定義にしたがっても，モデル識別の十分条件でも，また必要条件でもない。（必要条件でないことは，たとえば内生変数・外生変数を共通としても，

その構造方程式が一つは線形式，いま一つは対数線形式というふうに変換 $\mathfrak{G}$ を変えると，一般に Y の分布法則が変って identifiable になることから明らかである。）

まず福地の定義を精密化して，われわれはモデルの識別不能性を次の定義によって与えよう。

定義 3．（モデルの識別不能性）比較可能な二つのモデル（$Y, X_1, \boldsymbol{\theta}_1, \mathfrak{G}_1$）と（$Y, X_2, \boldsymbol{\theta}_2, \mathfrak{G}_2$）において，任意の X_1, X_2 が与えられているとき，ある $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ で，Y の条件つき密度関数が Y の変域全体にわたって，

(4)　$f_1(Y \mid X_1, \boldsymbol{\theta}_1, \mathfrak{G}_1)=f_2(Y \mid X_2, \boldsymbol{\theta}_2, \mathfrak{G}_2)$

となるとき，二つのモデルは $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ で互いにモデル識別不能であるという。

ここで X_1 と X_2 は恒等的に等しくても，また，等しくなくとも差支えない。但し，$X_1=X_2$ で，かつ $\mathfrak{G}_1, \mathfrak{G}_2$ が変換として同一のものであり，さらに $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ のうちの既知のものが等しいならば，二つのモデルはモデルとして同一であり，構造識別だけが残されることになる。なぜならば $\mathfrak{G}_1, \mathfrak{G}_2$ の equivalence より，$\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ は同一の parameter space に属し，成分の値のみが異なるものとなり，このうち既知の成分が互いに等しいということになると，未知の parameter の差異によってのみ二つのモデルは区別されることになり，モデル識別は通常の構造識別問題に還元される。ここでは構造識別と区別される意味でのモデル識別問題をあつかうのが目的であるから，このような場合，二つのモデルはモデルとしては同一（equivalent）であるとしてよい。

定義 4．（モデルの識別可能性）比較可能な二つのモデルが与えられた任意の X_1, X_2 に対して，それぞれの parameter 空間において，識別不能になる点をもたないならば，二つのモデルは互いに強く識別可能（strongly identifiable）であるという。また，二つのモデルが与えられた X_1, X_2 に対して，それぞれの parameter 空間のすべての点で識別不能となることがないならば，二つのモデルは互いに弱く識別可能であるという。

モデルの選択が可能になるためには，モデルが比較可能なばかりでなく，上記定義のうちのいずれかの意味で識別可能なモデルとなっていなければならない。以下では，モデルが強く識別可能であることを前提して，モデル選択問題に入っていく。なお，パラメーターのうち既知のものはモデルの一部だと考えてよいから，以下の議論では，パラメーターはすべて未知であると仮定する。

§3. 予測を基準とするモデル選択の諸原則

モデル選択は一つの decision making であり，その意味で本来高度に behavioral な問題である。現実には簡単に数学的型式で表現することができない研究者の「かん」とか嗜好といった factor によってもそれは影響されるし，またこれを別としても，われわれのいわゆる a priori information は必ずしも数学的なモデルに表現できるとは限らず，この種の information が数学的に表現されたモデルの plausibility に関与することは，しばしば経験するところである。また，もしも同一の予測結果と予測精度をもたらすモデルならば，より簡単な，推定作業がより容易に行なえるモデルの方が，予測に関する限りよりよいモデルだと言ってよいだろうから，モデル選択問題には，実際上，われわれの計算能力（計算機の演算速度・記憶容量等）や計算費用も関与する。

したがってモデル選択問題では，その選択が行なわれるさいの与件が明確にされていなければ，選択の原則を定めることはできない。

まずわれわれは，計算能力や計算費用からする諸制約は，以下で選択の対象となるモデルについて，すべて同一の条件を与えると仮定しよう。また

a priori information としては，モデルの数学的型式の中に折り込まれた以上のものは利用できないと仮定する。

既に述べたように計量経済学的研究にあたって研究者は，多くの場合，予測性能のより優れたモデルを選択することに力を傾ける。そしてモデルの妥当性に対する統計的テストにおいても，予測結果にもとずくテストがもっとも重要な，あるいはモデルに対するもっとも厳しいテストと考えられている。このことが決して根拠のないものでない以上，われわれもモデルの予測性能を基準としたモデル選択の原則を追求することからはじめるのが適当であろう。

いま $\boldsymbol{X}$, $\boldsymbol{\Theta}$, $\mathfrak{G}$ を所与としたとき $\boldsymbol{Y}$ の prediction $\hat{\boldsymbol{Y}}$ は，predictor $\mathfrak{E}_{\boldsymbol{Y}}$ によって，

$$\hat{\boldsymbol{Y}}=\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}, \boldsymbol{\Theta}\,;\,\mathfrak{G})$$

と与えられるものとしょう。

さて，二つのモデル $(\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G})$ と $(\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$ において，それぞれ外生変数 $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ が定まり，また $\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$ も適当な方法で推定されて，ともかくすべての parameter value が固定されたとしたとき，こうした条件の下での $\boldsymbol{Y}$ の prediction が二つのモデルで共通となる場合，すなわち，

$$(5)\quad \mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_1, \boldsymbol{\Theta}_1\,;\,\mathfrak{G}_1)=\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_2, \boldsymbol{\Theta}_2\,;\,\mathfrak{G}_2)$$

が所与の $\boldsymbol{X}_1$, $\boldsymbol{X}_2$. $\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$ について成立しているならば，二つのモデルは同一の prediction を与えるわけだから，予測精度という観点から，次のようなモデル選択の principle を考えることができるであろう。

定義 5. （モデル選択原則 1 ） (5)が成立しているとき，所与のいかなる $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ に対しても $\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_1, \boldsymbol{\Theta}_1\,;\,\mathfrak{G}_1)$のある近傍 S_r に含まれる任意の近傍Sで，

$$(6)\quad \int_{\mathrm{S}} d\mathrm{P}(\boldsymbol{Y}\,|\,\boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1) \geqq \int_{\mathrm{S}} d\mathrm{P}(\boldsymbol{Y}\,|\,\boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$$

が成立するならば，モデル $(\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1)$ はモデル $(\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$ に対して（予測精度に関して）preferable である。

しかし，異なる二つのモデルの下での $\boldsymbol{Y}$ のそれぞれの prediction が等しいという場合は，実際には稀であろう。予測モデル同志の間の distinctive な差異は，その予測精度以上に，多くの場合，prediction 自体に現われることが多いのである。しかもわれわれが真の $\boldsymbol{Y}$ の実現値を知らない以上，この場合どちらの prediction がより実現の可能性が高いかということは定められない。だが，$\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2$ は実際には過去のデータにもとずいて推定された値によって固定される。したがってこの場合，$\boldsymbol{Y}$ の確率分布を決定しているのは，真の random disturbance の分布ではなく，残差の分布である。

したがって，いま $\boldsymbol{Y}$ に関する prediction がこうした構造パラメーターの推定結果を条件とする conditional prediction の形で与えられるとすれば，その prediction 自体は二つのモデルの間で異なっていても，それぞれの prediction のまわりにより集中して $\boldsymbol{Y}$ が分布するモデルは，少なくとも構造推定の段階で用いられたデータに関しては，よりあてはまりの良いモデルになっているはずである。したがって構造パラメーターが安定しているとみなされ，かつ異なる $\boldsymbol{Y}$ の prediction に対して，われわれの態度がまったく無差別であるとすると，次のようなモデル選択の原則を立てることが許されるであろう。

定義 6.（モデル選択原則 2）

$$\boldsymbol{Z}_i=\boldsymbol{Y}-\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_i, \boldsymbol{\Theta}_i\,;\,\mathfrak{G}_i),\quad i=1,\ 2$$

とするとき，$\boldsymbol{0}\in \mathrm{R}^{\mathrm{H}}$ のある近傍 S_r に含まれる任意の近傍 S で，与えられたいかなる $\boldsymbol{X}_1, \boldsymbol{X}_2$ に対しても，

$$(7)\quad \int_{\mathrm{S}} d\mathrm{P}(\boldsymbol{Z}_1\,|\,\boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1) \geqq \int_{\mathrm{S}} d\mathrm{P}(\boldsymbol{Z}_2\,|\,\boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$$

が成立しているならば，モデル（$\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1$）はモデル（$\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2$）に対して（予測精度に関して）preferable である。

$\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_1, \boldsymbol{\Theta}_1\,;\,\mathfrak{G}_1)=\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_2, \boldsymbol{\Theta}_2\,;\,\mathfrak{G}_2)$ のとき(7)は(6)と同等であるから，定義 6.

は定義5. の拡張である。それゆえ，今後，「予測精度に関して preferable」というときは，一般に定義6. の選択原則の意味で言うことにしよう。

以上のモデル選択原則は構造推定に際して利用された情報以外の情報をもち合わせていない場合の選択原則である。したがって，あるモデルの下での prediction に対して，予測時点での実現値が知られた段階で，この情報と prediction とをつき合わせてモデルを修正し，あるいはモデルを選択しなおすという問題とは区別さるべきである。そこでわれわれは次に二つのモデルそれぞれの外生変数 $\boldsymbol{X}_1, \boldsymbol{X}_2$ の下で，$\boldsymbol{Y}$ に関するある observation $\boldsymbol{Y}_0$ が得られたときの二つのモデルの間の preference を定める原則を考えよう。(なお $\boldsymbol{Y}_0 \in \mathrm{R}^{\mathrm{H}}$ とする。) ここでそれぞれのモデルの構造パラメーター $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2$ は，あらかじめ適当な方法で推定され，モデル選択に対しては固定されたものと仮定してみよう。もし $\boldsymbol{Y}_0$ をなんらかの特殊な，偶発的な事情でもたらされた異常値と考えるべきなんらの理由もないならば，この observation に関する限り，prediction と observation とのひらきがより少なくなるようなモデルは，他方に対して preferable であると言えよう。

定義 7. (モデル選択原則 3) 二つのモデル $(\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1)$, $(\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$ において，所与の $\boldsymbol{X}_1, \boldsymbol{X}_2$ の下である observation $\boldsymbol{Y}_0$ が与えられたとき，もし，

$$(8) \quad |\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_1, \boldsymbol{\Theta}_1 ; \mathfrak{G}_1) - \boldsymbol{Y}_0| \leqq |\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_2, \boldsymbol{\Theta}_2 ; \mathfrak{G}_2) - \boldsymbol{Y}_0|$$

ならばモデル $(\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1)$ はモデル $(\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$ に対して，observation $\boldsymbol{Y}_0$ に関して preferable である。

しかし，この原則の致命的欠陥は，それが特定の一つの observation $\boldsymbol{Y}_0$ のみに関する preference を意味していることである。もっとも複数個の observation $\boldsymbol{Y}_{0i}$, $i=1, \dots, n$ がそれぞれの対応する外生変数 $\boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}$ が与えられている場合には，この原則は，

(9)　$\sum_{i=1}^{n}(Y_{0i}-\mathfrak{E}_Y(X_{1i},\boldsymbol{\Theta}_1;\mathfrak{G}_1))^2 \leqq \sum_{i=1}^{n}(Y_{0i}-\mathfrak{E}_Y(X_{2i},\boldsymbol{\Theta}_2;\mathfrak{G}_2))^{2^*}$

が成立するとき $(Y, X_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1)$ を $(Y, X_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$ に対して preferable とするものに拡張することができよう。しかし，これはあくまでも observation の組 $((Y_{01}, X_{1i}, X_{2i}),\ i=1, 2, \ldots, n)$ のみに限定した preference である。本質的な点は，この選択原則は構造推定において利用したデータをすべて無視した選択原則だということであって，この点はこのような拡張によってはいささかも影響されないのである。

Y のある prediction に対し，ある実現値 Y_0 がえられたとき，モデル選択テストに対して本質的な問題は，Christ [1] が指摘するように，Y_0 が構造推定の段階で利用したデータと本質的に同一の条件の下に実現したかどうか。つまりわれわれにとっては，未知であるところのある真の構造方程式とその構造パラメーターが，構造推定を行なって $\boldsymbol{\Theta}_i$ を固定させた段階と Y_0 が得られた段階とで，同一であるかどうかということである。

もし同一であるとするならば，Y_0 自身は既になにがしかの random disturbance を含むものであるから，これだけに着目して prediction が Y_0 に，より近いモデルを prefer するというのは，モデル選択の原則として充分とは言えない。なぜならば，この場合，構造推定で用いられたデータの構造方程式に対する適合度を Y_0 の適合度に対して軽視すべき理由はないからである。したがって，このような条件の下では，より妥当な選択原則は次のように述べられねばなるまい。

定義 8．（モデル選択原則 4）　二つのモデル $(Y, X_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1)$ と $(Y, X_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$ とにおいて，$\boldsymbol{\Theta}_1, \mathfrak{G}_1$ および $\boldsymbol{\Theta}_2, \mathfrak{G}_2$ が時間の経過に対して不変であるならば，構造推定の段階で外生変数 X_1, X_2 がそれぞれ n 個の $X_{1i}, X_{2i},\ i=1, \ldots,$

---

* 2乗の演算はここではベクトルの成分ごとに行なうものと約束する。すなわち X^2 は X の各成分を平方したベクトルを表わすものとする。

n と決定され，これに対応して $\boldsymbol{Y}$ の observation Y_{0i}, $i=1, \ldots, n$ が与えられ，これに対して prediction の時点について外生変数 $\boldsymbol{X}_1, \boldsymbol{X}_2$ が X_{1i}, X_{2i}, $i=n+1, \ldots, n+\mathrm{T}$ と与えられ，対応する $\boldsymbol{Y}$ が Y_{0i}, $i=n+1, \ldots, n+\mathrm{T}$ と実現したとき，

$$(10)\quad \sum_{i=1}^{n+\mathrm{T}}(Y_{0i}-\mathfrak{E}_{\boldsymbol{Y}}(X_{1i}, \boldsymbol{\theta}_1\,;\,\mathfrak{G}_1))^2 \leqq \sum_{i=1}^{n+\mathrm{T}}(Y_{0i}-\mathfrak{E}_{\boldsymbol{Y}}(X_{2i}, \boldsymbol{\theta}_2\,;\,\mathfrak{G}_2))^2$$

が成立しているならば，モデル（$\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\theta}_1, \mathfrak{G}_1$）はモデル（$\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\theta}_2, \mathfrak{G}_2$）に対して，observation（$(Y_{0i}, X_{1i}, X_{2i})$, $i=1, \ldots, n, n+1, \ldots, n+\mathrm{T}$）に関して preferable である。

これは構造推定時の observation と予測時点での observation をプールして，preditition に対する observation の適合度を比較することを意味している。しかし，ここで一定とした $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ は構造推定時の observation にもとずいて推定されたものであって，仮りにモデルのいずれかが真であるとしても，ここで固定された $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ までが真の parameter values を与えているという保証はないのが通常である。したがって上の条件の下でもより正しくは，構造推定時と予測時点での observation を全部使って，それぞれのモデルの下での構造推定をやり直し，このようにして新たに定められた構造パラメーターの下で，モデル選択原則 2 を適用するのが，計算の繁雑さを別とすれば，より正当なやり方と言えよう。

もちろんこれらの原則は，構造推定時点と予測時点との間に構造変動が生じている場合には適用できない。いまそれぞれのモデルにおいて，構造推定時点のパラメーター $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ は，予測時点でそれぞれ $\boldsymbol{\theta}_1+\varDelta\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2+\varDelta\boldsymbol{\theta}_2$ と変化すると考えられるとすると，これをあらためて $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ として，予測時点のみのデータにもとずいてこれを推定し，(8)ないし(9)を適用してモデル選択を行なうことが正当化されよう。しかしこの場合，もし構造推定時点ではモデル（$\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\theta}_1, \mathfrak{G}_1$）が（$\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\theta}_2, \mathfrak{G}_2$）に対して prefer され，予測時点では逆に（$\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\theta}_2, \mathfrak{G}_2$）が（$\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\theta}_1, \mathfrak{G}_1$）に対して prefer されるというこ

とが起りうる。こうした preference の変化は情報の蓄積を利用してもたらされる限りでは差支えないが，この場合は予測時点のみの observation にもとずいて preference を変更するわけで，この際，構造推定時点での情報がまったく無視されることになるから，適当とは言えまい。すなわち，先にモデル選択原則 3 の致命的欠陥として指摘した問題が，ここでも当然問題になるわけである。

だがこの場合，構造推定時点での情報と予測時点での情報を結合して行なわれるモデル選択の原則は，どのように考えるべきであろうか。これは本質的には，構造推定時点での observation. と予測時点での observation が，モデルの justification に対してもつ evidence としての相対的な重みに依存する問題である。いま適当な重み w, w' $(0\leqq w,\ w'\leqq 1,\ w+w'=1)$ がそれぞれ構造推定時点の observation と予測時点の observation につけることができるとすれば，モデル選択原則 3 は，構造変化がある場合にも適用できるよう，次のように拡張できるだろう。

定義 9.（モデル選択原則 5 ） 構造推定時点の observation $((\boldsymbol{Y}_{0i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}),\ i=1, \ldots, n)$ と予測時点での observation $((\boldsymbol{Y}_{0i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}),\ i=n+1, \ldots, n+\mathrm{T})$ が二つのモデル $(\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1)$，$(\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$ に関して与えられ，これらの observation のモデルの justification に対する evidence としての重み w, w' $(0\leqq w,\ w\leqq 1,\ w+w'=1)$ が与えられているならば，構造推定時点での構造パラメーターが，その時点での observation のみによって，それぞれのモデルで $\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$ と推定され，また予測時点での構造パラメーターが，その時点での observation のみによって，それぞれ $\boldsymbol{\Theta}_1'$, $\boldsymbol{\Theta}_2'$ と推定されているとき，もし，

$$(11)\quad w\sum_{i=1}^{n}(\boldsymbol{Y}_{0i}-\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_{1i}, \boldsymbol{\Theta}_1\,;\,\mathfrak{G}_1))^2+w'\sum_{i=n+1}^{n+\mathrm{T}}(\boldsymbol{Y}_{0i}-\mathfrak{G}_{\boldsymbol{Y}}(\boldsymbol{X}_{1i}, \boldsymbol{\Theta}_1'\,;\,\mathfrak{G}_1))^2$$
$$\leqq w\sum_{i=1}^{n}(\boldsymbol{Y}_{0i}-\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_{2i}, \boldsymbol{\Theta}_2\,;\,\mathfrak{G}_2))^2+w'\sum_{i=n+1}^{n+\mathrm{T}}(\boldsymbol{Y}_{0i}-\mathfrak{E}_{\boldsymbol{Y}}(\boldsymbol{X}_{2i}, \boldsymbol{\Theta}_2'\,;\,\mathfrak{G}_2))^2$$

ならば，モデル $(\boldsymbol{Y}, \boldsymbol{X}_1, \boldsymbol{\Theta}_1, \mathfrak{G}_1)$ はモデル $(\boldsymbol{Y}, \boldsymbol{X}_2, \boldsymbol{\Theta}_2, \mathfrak{G}_2)$ に対して所与の observation に関して重み w, w' の下で preferable である。

だが，この重み w, w' を定める充分に説得的な根拠は与えにくいように思われる。また，構造変化がどの時点で起ったかということは，実際には不確実である。特に observation が時系列データの形で与えられているときには，構造変化はもっと連続的に生じているのではないかと考えることもできるが，この場合には実際問題としてそれぞれの時点での構造パラメーターすら決定することができなくなるであろう。これらはいずれもモデル選択問題を困難な問題にする要因と考えられる。

§4. モデル選択問題におけるベイジアン・アプローチ

前節にかかげたモデル選択の principle はなんら新奇なものではなく，多くの econometric researcher たちが暗黙のうちに採用しているところのものである。だが現実の evidence をもとにして，どの principle を実際に採用するかは，そのとき確実とみなされる情報の種類，evidence の性格に依存する。そしてモデル選択問題は，まさにモデルを構成している a priori information の不確実さのゆえに発生することを思えば，この問題の本質は，こうした不確実な土台の上に示される個々のモデルの予測性能の比較といったプラグマティックな点にあるのではなく，それぞれのモデルの正しさに対するわれわれの a priori な確信が，時間の経過ないしは research の前進にともなってもたらされたもろもろの evidence によって，いかに修正され，また修正されねばならぬかという点に求められねばなるまい。モデル選択の基準として，予測性能をとるということは，あるいは予測値と実現値を比較するということは，これがわれわれのモデルの正しさに対する確信に影響する evidence とみなしうるという前提の下に許されるのである。

たとえば言い古された例ではあるが，景気変動を説明するモデルにおい

て，ある時期については所与のデータの下では，投資や消質，あるいは人口等のいわゆる economic term によるモデルよりも，太陽黒点の数のような non-economic term によるモデルの方が fit がよい，つまり経験的な予測性能が少なくともこの時期に限ってはよいという結果が生じたとき，研究者はこの evidence にもとずいて後者を prefer するであろうか。本来，経済現象は多くの random とみなされうるかなり大幅な disturbance をともなうということを知っている経済学者は，恐らくこうした evidence にもかかわらず，景気変動を前者のモデルにおける economic term が random disturbance を伴ないながらも規定していく mechanism の存在を否定する気にはならないであろう。

モデル選択問題のこうした本質に即して，考察を進めてこそ，モデル選択の principle を正しく把握し，これを適切に適用することができるであろう。そして，このことはモデル選択問題においては，ベイジアン・アプローチが理論の基本的な骨組とならねばならないということを示唆するのである。以下ではこうしたベイジアンの立場に立ったモデル選択問題の formulation を試みてみよう。

いまモデル空間 M は discrete であるとしょう。このときわれわれは適当な方法で M の各要素 $\omega \in \mathsf{M}$, $\omega=(Y, X, \Theta, \mathfrak{S})$ に対して確率 $\pi(\omega)$ を定めることができる。この $\pi(\omega)$ をわれわれは，モデル ω の正しさに対する研究者の a priori な degree of belief を表わす主観確率と考えよう。ここで注意すべきことは，M は $Y, X, \Theta, \mathfrak{S}$ がそれぞれ属する空間の直積空間というよりは（そう考えることもできないわけではないが），研究者によって行なわれる内生変数，外生変数およびそれらの間の関数関係の選択の仕方の違いによって，その要素が区別されるところの空間であって，つまり choice of model という研究者の activity の集合を表わしていることである。

モデル選択問題に対するベイジアン・アプローチは，ある任意の observation Z によって，a priori な ω の degree of belief $\pi(\omega)$ がどのような手

続きで，また ω に関するどのような a posteriori probability に変換されるかを示すことであり，このプロセスでベイスの定理を活用することにある。

もっとも直接的なアプローチは，パラメーターの推論の場合と平行的に，$\pi(\omega|\boldsymbol{Z})$ を考えることである。

いま $p(\boldsymbol{\theta}|\omega)$ をモデル ω を選択したときの，そのモデルのパラメーター $\boldsymbol{\theta}$ の条件つき確率密度としょう。これはパラメーターの推論に関してベイジアンが用いるところのパラメーター $\boldsymbol{\theta}$ の a priori probability density に他ならず，モデル ω の下で，パラメーター $\boldsymbol{\theta}$ を与えたときの observation $\boldsymbol{Z}$ の確率密度を $f(\boldsymbol{Z}|\boldsymbol{\theta}, \omega)$ で表わせば，ベイスの定理により $\boldsymbol{\theta}$ の a posteriori probability density $p(\boldsymbol{\theta}|\boldsymbol{Z}, \omega)$ は，

(12)　$p(\boldsymbol{\theta}|\boldsymbol{Z}, \omega) \propto p(\boldsymbol{\theta}|\omega)f(\boldsymbol{Z}|\boldsymbol{\theta}, \omega)$

によって定められる。他方，ω と $\boldsymbol{\theta}$ の同時確率密度は，

(13)　$\pi(\omega, \boldsymbol{\theta}) = \pi(\omega)p(\boldsymbol{\theta}|\omega)$

であり，これを a priori probability density として，ベイスの定理を再び適用すれば，$\boldsymbol{Z}$ が与えられたとき，

(14)　$\pi(\omega, \boldsymbol{\theta}|\boldsymbol{Z}) \propto \pi(\omega, \boldsymbol{\theta}) \cdot f(\boldsymbol{Z}|\boldsymbol{\theta}, \omega)$

によって $\pi(\omega, \boldsymbol{\theta}|\boldsymbol{Z})$ を定めることができ，

$$\pi(\omega, \boldsymbol{\theta}|\boldsymbol{Z}) = \pi(\omega|\boldsymbol{Z})p(\boldsymbol{\theta}|\boldsymbol{Z}, \omega)$$

より，

(15)　$\pi(\omega|\boldsymbol{Z}) = \pi(\omega, \dot{\boldsymbol{\theta}}|\boldsymbol{Z})/p(\boldsymbol{\theta}|\boldsymbol{Z}, \omega)$

を得る。ここで(12)，(14)の比例定数はそれぞれ，パラメーター空間を Θ で表わせば，$(\int_\Theta p(\boldsymbol{\theta}|\omega)f(\boldsymbol{Z}|\boldsymbol{\theta}, \omega)d\boldsymbol{\theta})^{-1}$，$(\sum_\omega\int_\Theta \pi(\omega, \boldsymbol{\theta}) \cdot f(\boldsymbol{Z}|\boldsymbol{\theta}, \omega)d\boldsymbol{\theta})^{-1}$ であり，また(13)より(14)の比例定数がさらに $(\sum_\omega \pi(\omega) \cdot \int_\Theta p(\boldsymbol{\theta}|\omega)f(\boldsymbol{Z}|\boldsymbol{\theta}, \omega)d\boldsymbol{\theta})^{-1}$ となることを考慮して，(15)を書き改めると，(12)，(13)，(14)より，

(16)　$$\pi(\omega|\boldsymbol{Z}) = \frac{\pi(\omega)\int_\Theta p(\boldsymbol{\theta}|\omega)f(\boldsymbol{Z}|\boldsymbol{\theta}, \omega)d\boldsymbol{\theta}}{\sum_\omega \pi(\omega)\int_\Theta p(\boldsymbol{\theta}|\omega)f(\boldsymbol{Z}|\boldsymbol{\theta}, \omega)d\boldsymbol{\theta}}$$

$$\propto \pi(\omega)\int_{\Theta} p(\boldsymbol{\theta}\,|\,\omega)f(\boldsymbol{Z}\,|\,\boldsymbol{\theta},\ \omega)d\boldsymbol{\theta}$$

を得る。

すなわちモデル ω の下でのパラメーター $\boldsymbol{\theta}$ に対するベイジアンの統計的推論⑿における比例定数の逆数を observation $\boldsymbol{Z}$ のモデル ω に対する weight of evidence として $\pi(\omega)$ に乗じたものに比例するように，ω の事後確率 $\pi(\omega\,|\,\boldsymbol{Z})$ は定められることになるのである。そこで，

定義 10.（モデル選択のベイジアン・ルール1）

二つのモデル ω_1 と ω_2 とにおいて，それぞれのモデルのパラメーターを $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, それらが属するそれぞれのパラメーター空間を Θ_1, Θ_2 とするとき，いま ω_1 に対する事前確率 $\pi(\omega_1)$, ω_2 に対する事前確率 $\pi(\omega_2)$ が与えられ，また，それぞれのモデルのパラメーターに対する推論が，observation $\boldsymbol{Z}$ の下で，

$$p(\boldsymbol{\theta}_i\,|\,\boldsymbol{Z},\ \omega_i)\propto p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}\,|\,\boldsymbol{\theta}_i,\ \omega_i),\ i=1,\ 2$$

によって与えられているとき，

$$\pi(\omega_i\,|\,\boldsymbol{Z})\propto\pi(\omega_i)\int_{\Theta_i} p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}\,|\,\boldsymbol{\theta}_i,\ \omega_i)d\boldsymbol{\theta}_i,\ i=1,\ 2$$

によって定められる事後確率 $\pi(\omega_i\,|\,\boldsymbol{Z})$, $i=1,\ 2$ が

$$\pi(\omega_1\,|\,\boldsymbol{Z})\geqq\pi(\omega_2\,|\,\boldsymbol{Z})$$

となっているならば，モデル ω_1 はモデル ω_2 に対して preferable である。

というモデル選択のベイジアン・ルールを考えることができる。

この選択ルールは，それぞれのモデルのパラメーターが，そのパラメーター空間上のどの位置にあるかということとは indifferent に与えられる。つまり $\int_{\Theta} p(\boldsymbol{\theta}\,|\,\omega)f(\boldsymbol{Z}\,|\,\boldsymbol{\theta},\ \omega)d\boldsymbol{\theta}$ は，observation $\boldsymbol{Z}$ が $\boldsymbol{\theta}$ とは indifferent に ω に対して与える weight of evidence となっており，また，これはパラメーターに関するベイジアンの推論の副産物として多くの場合容易に評価され

る。

しかし，このルールがパラメーターに indifferent に与えられるということは，このルールがあまり鋭敏にモデルの preference relation を識別するものにはならないであろうことを予想させる。実際のモデル選択においては，前節の議論にも見られるように，パラメーター $\boldsymbol{\Theta}$ を適当な推定方法によって固定し，その上で例えば予測結果という observation とつき合わせ，alternative なモデルの優劣を判断するという行き方をするのが普通である。そしてパラメーターが高い精度で推定されるということは，observation が構造推定時点と予測時点とで，まったく同一の条件の下でもたらされているとするならば，prediction 自体も良い精度と適合度をもつと期待しうることが多いのである。他方われわれのルールは，$\boldsymbol{\Theta}$ のいかなる点でも事後的な degree of belief は高くないのに，パラメーター空間上のある点で高い事後的な degree of belief をもたらす他のモデルに対して，前者が preferable であるような preference relation をもたらすことを一般には排除しないのである。

そこでわれわれは次に，それぞれのモデルのパラメーター空間上にある局所領域を限定し，パラメーターがこの局所領域にあることを前提にした上でのモデルの preference を考えてみよう。

ここではこうした領域として，それぞれのモデルのパラメーター空間上の1点を限定する場合から考察をはじめる。ところでこのことは $\pi(\omega|\boldsymbol{\Theta})$ や $\pi(\omega|\boldsymbol{\Theta}, \boldsymbol{Z})$ を考えることは無意味であるから $\pi(\omega, \boldsymbol{\Theta}|\boldsymbol{Z})$ にもとずく preference を考えることを必然ならしめる。

いま $\omega_i \in \mathbf{M}$, $i=1, 2, \ldots$ のそれぞれに対応するパラメーター $\boldsymbol{\Theta}_i$, $i=1, 2, \ldots$ をそれぞれ1点に固定して，それを $\hat{\boldsymbol{\Theta}}_i$ と表わすと⑭より，

$$(17)\quad \pi(\omega_i, \hat{\boldsymbol{\Theta}}_i|\boldsymbol{Z}) \propto \pi(\omega_i, \hat{\boldsymbol{\Theta}}_i) f(\boldsymbol{Z}|\hat{\boldsymbol{\Theta}}_i, \omega_i)$$

が定められ，この比例式における比例定数は，すべての i について共通で，

$$(\sum_i \int_{\Theta_i} \pi(\omega_i, \boldsymbol{\theta}_i) f(\boldsymbol{Z} | \boldsymbol{\theta}_i, \omega_i) d\boldsymbol{\theta}_i)^{-1}$$

で与えられる。ここで Θ_i は ω_i のパラメーター空間を表わす。

さらに，

$$\pi(\omega_i, \hat{\boldsymbol{\theta}}_i) = \pi(\omega_i) p(\hat{\boldsymbol{\theta}}_i | \omega_i)$$

であるから，(17)はすべての i 対して共通の比例定数の下で，

$$(18) \quad \pi(\omega_i, \hat{\boldsymbol{\theta}}_i | \boldsymbol{Z}) \propto \pi(\omega_i) p(\hat{\boldsymbol{\theta}}_i | \omega_i) f(\boldsymbol{Z} | \hat{\boldsymbol{\theta}}_i, w_i)$$

と書ける。

さて，われわれのモデルに対する事前的な確信が，完全に $\pi(\omega_i)$ で表現されており，これと与えられた observation $\boldsymbol{Z}$ にもとずく evidence 以外にモデルを preference の上で差別するものがないとするならば，それぞれのモデルは所与の evidence の下で可能な最大の自己主張をもつと想定することは不自然ではない。いいかえるならば，特定のモデル ω_i が他のモデルに対して preference の上で，特に不利になるような $\boldsymbol{\theta}_i$ によって比較する根拠がない以上，$\pi(\omega_i, \boldsymbol{\theta}_i | \boldsymbol{Z})$ が $\boldsymbol{Z}$ の下で可能な最大の値をとりうるように $\hat{\boldsymbol{\theta}}_i$ を選ぶのが，公平な態度というべきである。ところで $\pi(\omega_i)$ は所与であるから，このためには $p(\hat{\boldsymbol{\theta}}_i | \omega_i) f(\boldsymbol{Z} | \hat{\boldsymbol{\theta}}_i, \omega_i)$ が最大になればよい。しかるに $\boldsymbol{\theta}_i$ の a posteriori probability density は，

$$(19) \quad p(\boldsymbol{\theta}_i | \boldsymbol{Z}, \omega_i) \propto p(\boldsymbol{\theta}_i | \omega_i) f(\boldsymbol{Z} | \boldsymbol{\theta}_i, \omega_i)$$

で定められるわけだから，このことは $\hat{\boldsymbol{\theta}}_i$ としては，$\boldsymbol{\theta}_i$ の事後確率分布のモード (mode) をとればよいことを意味する。そこでわれわれは，このような場合の選択ルールとして次のようなものを得る。

定義 11. （モデル選択のベイジアン・ルール 2）

二つのモデル ω_1, ω_2 において，共通の observation $\boldsymbol{Z}$ が得られたとき，それぞれのモデルの下でのパラメーター $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ の $\boldsymbol{Z}$ を与えたときの a posteriori probability density

$$p(\boldsymbol{\theta}_1 | \boldsymbol{Z}, \omega_1) \propto p(\boldsymbol{\theta}_1 | \omega_1) f(\boldsymbol{Z} | \boldsymbol{\theta}_1, \omega_1)$$
$$p(\boldsymbol{\theta}_2 | \boldsymbol{Z}, \omega_2) \propto p(\boldsymbol{\theta}_2 | \omega_2) f(\boldsymbol{Z} | \boldsymbol{\theta}_2, \omega_2)$$

をそれぞれ最大ならしめる $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ をそれぞれ $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2$ とするとき，

$$(20) \quad \pi(\omega_1, \hat{\boldsymbol{\theta}}_1 | \boldsymbol{Z}) \geqq \pi(\omega_2, \hat{\boldsymbol{\theta}}_2 | \boldsymbol{Z})$$

すなわち，

$$(21) \quad \pi(\omega_1) p(\hat{\boldsymbol{\theta}}_1 | \omega_1) f(\boldsymbol{Z} | \hat{\boldsymbol{\theta}}_1, \omega_1) \geqq \pi(\omega_2) p(\hat{\boldsymbol{\theta}}_2 | \omega_2) f(\boldsymbol{Z} | \hat{\boldsymbol{\theta}}_2, \omega_2)$$

ならば，モデル ω_1 はモデル ω_2 に対して preferable である。

ところで(18)は，比例定数を explicit に示して等式の形に書くと，

$$(22) \quad \pi(\omega_i, \hat{\boldsymbol{\theta}}_i | \boldsymbol{Z}) = \frac{\pi(\omega_i) p(\hat{\boldsymbol{\theta}}_i | \omega_i) f(\boldsymbol{Z} | \hat{\boldsymbol{\theta}}_i, \omega_i)}{\sum_i \int_{\Theta_i} \pi(\omega_i) p(\boldsymbol{\theta}_i | \omega_i) f(\boldsymbol{Z} | \boldsymbol{\theta}_i, \omega_i) d\boldsymbol{\theta}_i}$$

である。この $\pi(\omega_i, \hat{\boldsymbol{\theta}}_i | \boldsymbol{Z})$ を比較するのが，ここでのベイジアン・ルールであるが，$\hat{\boldsymbol{\theta}}_i$, $i=1, 2, \ldots$ を固定している以上，(22)の分母の値はわれわれの preference には影響を与えない。そこでいま(22)の分母の $\boldsymbol{\theta}_i$ に関する積分をやめて，

$$\frac{\pi(\omega_i) p(\hat{\boldsymbol{\theta}}_i | \omega_i) f(\boldsymbol{Z} | \hat{\boldsymbol{\theta}}_i, \omega_i)}{\sum_i \pi(\omega_i) p(\hat{\boldsymbol{\theta}}_i | \omega_i) f(\boldsymbol{Z} | \hat{\boldsymbol{\theta}}_i, \omega_i)}$$

を考えてみると，$\boldsymbol{Z}$ および $\hat{\boldsymbol{\theta}}_i$, $i=1, 2, \ldots$ が固定されている以上，これは明らかに $\boldsymbol{M}$ の上の確率分布を与えている。いまこれを $\pi_{\hat{\theta}}(\omega_i | \boldsymbol{Z})$ と書いてみよう。すなわち，

$$(23) \quad \pi_{\hat{\theta}}(\omega_i | \boldsymbol{Z}) = \frac{\pi(\omega_i) p(\hat{\boldsymbol{\theta}}_i | \omega_i) f(\boldsymbol{Z} | \hat{\boldsymbol{\theta}}_i, \omega_i)}{\sum_i \pi(\omega_i) p(\hat{\boldsymbol{\theta}}_i | \omega_i) f(\boldsymbol{Z} | \hat{\boldsymbol{\theta}}_i, \omega_i)}$$

すると，

$$(24) \quad \frac{\pi(\omega_i, \hat{\boldsymbol{\theta}}_i | \boldsymbol{Z})}{\pi_{\hat{\theta}}(\omega_i | \boldsymbol{Z})} = p(\bigcup_i \hat{\boldsymbol{\theta}}_i | \boldsymbol{Z})$$

が導かれる。すなわち(22)，(23)より，

$$(25)\quad \frac{\pi(\omega_i, \boldsymbol{\theta}_i \mid \boldsymbol{Z})}{\pi_{\hat{\Theta}}(\omega_i \mid \boldsymbol{Z})} = \frac{\sum_i \pi(\omega_i) p(\hat{\boldsymbol{\theta}}_i \mid \omega_i) f(\boldsymbol{Z} \mid \hat{\boldsymbol{\theta}}_1, \omega_i)}{\sum_i \pi(\omega_i) \int_{\Theta_i} p(\boldsymbol{\theta}_i \mid \omega_i) f(\boldsymbol{Z} \mid \boldsymbol{\theta}_i, \omega_i) d\boldsymbol{\theta}_i}$$

であるが，

$$f(\boldsymbol{Z} \mid \omega_i) = \int_{\Theta_i} p(\boldsymbol{\theta}_i \mid \omega_i) f(\boldsymbol{Z} \mid \boldsymbol{\theta}_i, \omega_i) d\boldsymbol{\theta}_i$$

とおくと，$f(\boldsymbol{Z} \mid \omega_i)$ は ω_i の下における $\boldsymbol{\theta}_i$ と $\boldsymbol{Z}$ との同時確率密度関数を $\boldsymbol{\theta}_i$ について積分したもの，つまり $\boldsymbol{Z}$ の周辺分布を表わしており，これを用いれば(25)は，

$$\begin{aligned}
\frac{\pi(\omega_i, \hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z})}{\pi_{\hat{\Theta}}(\omega_i \mid \boldsymbol{Z})} &= \frac{\sum_i \pi(\omega_i) f(\boldsymbol{Z} \mid \omega_i) p(\hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z}, \omega_i)}{\sum_i \pi(\omega_i) f(\boldsymbol{Z} \mid \omega_i)} \\
&= \frac{\sum_i \pi(\omega_i, \boldsymbol{Z}) p(\hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z}, \omega_1)}{\sum_i \pi(\omega_i, \boldsymbol{Z})} \\
&= \frac{\sum_i \pi(\omega_i, \boldsymbol{Z}, \hat{\boldsymbol{\theta}}_i)}{\sum_i \pi(w_i, \boldsymbol{Z})} \\
&= \frac{p(\boldsymbol{Z}, \bigcup_i \hat{\boldsymbol{\theta}}_i)}{f(\boldsymbol{Z})} \\
&= p(\bigcup_i \hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z})
\end{aligned}$$

となるからである。但し，ここで $\pi(\omega_i, \boldsymbol{Z}, \boldsymbol{\theta}_i)$ は $\omega_i, \boldsymbol{Z}, \boldsymbol{\theta}_i$ の同時確率密度関数であり，$\pi(\omega_i, \boldsymbol{Z})$ はこれから導かれた ω_i と $\boldsymbol{Z}$ との周辺同時確率密度関数，$f(\boldsymbol{Z})$ は同じく $\boldsymbol{Z}$ の周辺確率密度関数である。また $p(\boldsymbol{Z}, \bigcup_i \hat{\boldsymbol{\theta}}_i)$ は $\boldsymbol{Z}$ と $\bigcup_i \hat{\boldsymbol{\theta}}_i$ との同時確率密度であり，

$$\sum_i \pi(\omega_i, \boldsymbol{Z}, \hat{\boldsymbol{\theta}}_i) = p(\boldsymbol{Z}, \bigcup_i \boldsymbol{\theta}_i)$$

は，$\omega_i \cap \hat{\boldsymbol{\theta}}_i$ と $\omega_j \cap \hat{\boldsymbol{\theta}}_j$ が $i \neq j$ ならば互いに排反であることによる。

ところでこのことは，$\omega_i \cap (\bigcup_i \hat{\boldsymbol{\theta}}_i) = \omega_i \cap \hat{\boldsymbol{\theta}}_i$ であることから

$$\pi(\omega_i, \bigcup_i \hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z}) = \pi(\omega_i, \hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z})$$

が成り立つから，

$$\frac{\pi(\omega_i,\ \bigcup_i \hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z})}{\pi_{\hat{\Theta}}(\omega_i \mid \boldsymbol{Z})} = p(\bigcup_i \hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z})$$

を意味し，したがって，

$$\pi_{\hat{\Theta}}(\omega_i \mid \boldsymbol{Z}) = \frac{\pi(\omega_i,\ \bigcup_i \hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z})}{p(\bigcup_i \hat{\boldsymbol{\theta}}_i \mid \boldsymbol{Z})}$$

であり，これは $\boldsymbol{Z}$ および $\bigcup_i \hat{\boldsymbol{\theta}}_i$ を与えたときの ω_i の条件つき確率 $\pi(\omega_i \mid \boldsymbol{Z},\ \bigcup_i \hat{\boldsymbol{\theta}}_i)$ を表わしている。すなわち，

$$\pi_{\hat{\Theta}}(\omega_i \mid \boldsymbol{Z}) = \pi(\omega_i \mid \boldsymbol{Z},\ \bigcup_i \hat{\boldsymbol{\theta}}_i)$$

である。

それゆえ，結局われわれの選択ルールは $\boldsymbol{Z}$ という observation が与えられたとき，それぞれのモデルに対して $\hat{\boldsymbol{\theta}}_i$, $i=1, 2, \ldots$ が定まり，このうちのいずれかが真であることは確実だという convention の下で，ω_i が真であることの確率を比較するということに帰着するのである。

ところでモデル選択にあたって，それぞれのモデルについてパラメーターを1点に固定するということは，必ずしも必要でない。すなわち，われわれの選択ルールが結局 $\pi(\omega_i \mid \boldsymbol{Z},\ \bigcup_i \hat{\boldsymbol{\theta}}_i)$ なる条件つき確率の比較に帰着する以上，これを拡張して，それぞれのパラメーター空間上にそれぞれ適当な領域 $\mathrm{S}_i \subseteq \Theta_i$, $i=1, 2, \ldots$ をとって，$\pi(\omega_i \mid \boldsymbol{Z},\ \bigcup_i \mathrm{S}_i)$ を比較することによってもモデル選択を行なうことができるであろう。

しかし，問題はこのような S_i としてどのようなものをとるべきかということである。いままでの議論に照らして考えると，まず S_i は $\boldsymbol{Z}$ に depend して決定される領域であるべきである。また S_i が広ければ広いほど，そのモデルのパラメーター $\boldsymbol{\theta}_i$ が S_i に入る事後確率は増加するから，S_i はそれぞれのモデルの比較にあたって同等とみなしうるなんらかの規準によって，その広さが限定されなければならない。さらに S_i は，こうして広さを限定した上で，同じ広さの領域の中では，$\boldsymbol{\theta}_i$ が S_i に入る事後確率が最大になる

ように選ばれるべきである。

多くの場合，$\boldsymbol{\theta}_i$ の事後確率分布が単峰型のなめらかな分布であると仮定することは，それほど非現実的なことではない。とすると S_i は $\boldsymbol{Z}$ によって定まる $\boldsymbol{\theta}_i$ の事後確率分布のモード $\hat{\boldsymbol{\theta}}_i$ を含んだ連続的な領域をとるのが適当であろう。問題はむしろ S_i の広さを限定するルールである。$\boldsymbol{\theta}_i$ はモデルごとにまったく違った意味をもち，共通の尺度では比較できない成分をもっている。加えてその次数もモデルごとに一般に異なっている。こうした $\boldsymbol{\theta}_i$ が属するある領域 S_i を，その広さがモデル選択に影響することのないように定めることが，いまの場合必要なわけである。

いま $\boldsymbol{M}$ に属するすべてのモデルは，§2 で定義した意味で consistent であるとしょう。するとそれぞれのモデルについて $\hat{\boldsymbol{\theta}}_i$ を定め，かつ observation $\boldsymbol{Z}$ によって，外生変数 $\boldsymbol{X}_i$ もすべてのモデルについて定められたとすると，そのときこれらのモデルに共通する内生変数 $\boldsymbol{Y}$ の $\hat{\boldsymbol{\theta}}_i, \boldsymbol{X}_i,\ i=1, 2, \dots$ におけるすべての成分の値が決定される。こうして決定された $\boldsymbol{Y}$ を $\hat{\boldsymbol{Y}}_i$ で表わし，その各成分を $\hat{y}_{ij},\ j=1, 2, \dots, \mathrm{H}$ で表わそう。

もし，われわれが内生変数の prediction あるいは $\hat{\boldsymbol{Y}}_i$ に対する $\boldsymbol{Y}$ の observation の fit に主要な関心を寄せる立場をとるならば，$\hat{\boldsymbol{Y}}_i$ を含む $\underline{\boldsymbol{Y}}_i \leqq \hat{\boldsymbol{Y}}_i \leqq \overline{\boldsymbol{Y}}_i$ によって与えられる領域 $S_{Y_i}=[\underline{\boldsymbol{Y}}_i, \overline{\boldsymbol{Y}}_i]$ をすべてモデルについて定めることができるであろう。なお $[\underline{\boldsymbol{Y}}_i, \overline{\boldsymbol{Y}}_i]$ は $\boldsymbol{Y}_i$ の各成分を $\underline{y}_{i1}, \underline{y}_{i2}, \dots, \underline{y}_{i\mathrm{H}}$, $\overline{\boldsymbol{Y}}_i$ の各成分を $\overline{y}_{i1}, \overline{y}_{i2}, \dots, \overline{y}_{i\mathrm{H}}$ とするとき，$\underline{y}_{i1} \leqq y_1 \leqq \overline{y}_{i1}$, $\underline{y}_{i2} \leqq y_2 \leqq \overline{y}_{i2}, \dots$, $\underline{y}_{i\mathrm{H}} \leqq y_{\mathrm{H}} \leqq \overline{y}_{i\mathrm{H}}$ なる H 個の区間の直積によって与えられる R^{H} 上の領域を表わしている。また $\hat{\boldsymbol{Y}}_i, \overline{\boldsymbol{Y}}_i, \underline{\boldsymbol{Y}}_i$ と添字 i をつけたのは，これらの成分の値は，それぞれのモデルについて与えられた $\hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{X}}_i$ の下で，モデルごとに異なったものになりうるからである。

さて S_{Y_i} 自体は R^{H} 上の異なった部分区間を表わしているが，すべてのモデルに対して S_{Y_i} を与える同一内生変数に関する区間の幅，すなわち $\overline{y}_{ij}-\underline{y}_{ij}$ を一定にし，さらに $\overline{y}_{ij}-\hat{y}_i$, $\hat{y}_i-\underline{y}_{ij}$ も一定になるように S_{Y_i} を定

めることができる。

いま S_{Y_i} をこのように定めたとすると，これに対応して，それぞれのモデルについて $\hat{\boldsymbol{\theta}}_i$ を含むパラメーター空間上の領域 S_{θ_i} を定めることができる。すなわちモデル w_i を，

$$\mathfrak{G}_i(\boldsymbol{Y}, \boldsymbol{X}_i, \boldsymbol{\theta}_i)=\mathbf{0}$$

と表わしてみると，consistent なモデルならば，これから，

$$\boldsymbol{Y}=\mathfrak{F}_i(\boldsymbol{X}_i, \boldsymbol{\theta}_i)$$

なる方程式体系を導くことができる。いわゆる誘導形である。これは，w_i における外生変数の個数を K_i，パラメーターの個数を L_i とすると，$R^{L_i+K_i}$ から R^H への mapping $\mathfrak{F}_i$ を与えるものであるが，ここで $\boldsymbol{X}_i$ を所与とすれば，R^{L_i} から R^H への mapping と考えることができる。$\boldsymbol{X}_i$ を所与とするこの mapping を $\mathfrak{F}_{\boldsymbol{X}_i}$ と表わそう。すると S_{θ_i} は，

$$S_{\theta_i}=\mathfrak{F}_{\boldsymbol{X}_i}^{-1}(S_{Y_i})$$

によって決定されることになる。

$\hat{\boldsymbol{Y}}_i=\mathfrak{F}_{\boldsymbol{X}_i}(\hat{\boldsymbol{\theta}}_i)$ であるから，$\mathfrak{F}_{\boldsymbol{X}_i}^{-1}(\hat{\boldsymbol{Y}}_i)$ は $\hat{\boldsymbol{\theta}}_i$ を含む集合となり，$\hat{\boldsymbol{Y}}_i \in S_{Y_i}$ である以上，$\mathfrak{F}_{\boldsymbol{X}_i}^{-1}(S_{Y_i})$ は $\hat{\boldsymbol{\theta}}_i$ をもちろん含んでいる。

さらにこのような手続きによって得られる S_{θ_i} は S_{Y_i} に要求されている諸条件を充す諸領域のそれぞれに対応しているわけだが，こうした諸領域の中から $\boldsymbol{\theta}_i$ が S_{θ_i} に入る事後確率を最大ならしめるものを選ぶことができる。すなわち同じ条件を充たす S_{θ_i} の中で，事後確率を最大にするものを定めることができる。

さて，こうして定められた S_{θ_i} はそれぞれのモデルの下での内生変数の prediction の精度，あるいは $\boldsymbol{Y}$ の $\hat{\boldsymbol{Y}}_i$ に対する適合度という基準に関しては，モデル選択の上で各モデルについて同等とみなしうる領域である。この領域にもとずいてわれわれは，パラメーター空間上に局所領域を固定して行なうモデル選択のルールの一つとして，次のようなものを採用することができる。それは prediction に関して，同等の精度を基準とする選択ルールであ

るという意味で，prediction に関する Baysian model preference principle と呼んでもよいであろう。

定義 12.（モデル選択のベイジアン・ルール 3）

observation $\boldsymbol{Z}$ が与えられたとき，それによってそれぞれのモデルのパラメーターの事後確率分布のモード $\hat{\boldsymbol{\theta}}_i$, $i=1, 2, \dots$ と外生変数 $\boldsymbol{X}_i$ が決定され，また，この $\hat{\boldsymbol{\theta}}_i, \boldsymbol{X}_i$ によって内生変数 $\boldsymbol{Y}$ が一義的に $\hat{\boldsymbol{Y}}_i=\mathfrak{F}_{\boldsymbol{X}_i}(\hat{\boldsymbol{\theta}}_i)$ とそれぞれのモデルにおいて決定されるとき，

$$\boldsymbol{D}_1=\bar{\boldsymbol{Y}}_i-\hat{\boldsymbol{Y}}_i$$

$$\boldsymbol{D}_2=\hat{\boldsymbol{Y}}_i-\underline{\boldsymbol{Y}}_i$$

をすべてのモデルについて共通ならしめる領域，

$$\mathrm{S}_{\boldsymbol{Y}_i}=[\underline{\boldsymbol{Y}}_i,\ \bar{\boldsymbol{Y}}_i]$$

の mapping $\mathfrak{F}_{\boldsymbol{X}_i}$ による原像，

$$\mathrm{S}_{\theta_i}=\mathfrak{F}_{\boldsymbol{X}_i}^{-1}(\mathrm{S}_{\boldsymbol{Y}_i})$$

が，$\hat{\boldsymbol{Y}}_i$ を含み，かつ $\boldsymbol{D}_1+\boldsymbol{D}_2$ を一定とする $\mathrm{S}_{\boldsymbol{Y}_i}$ のうち $\boldsymbol{\theta}_i$ が S_{θ_i} に入る事後確率を最大にするものによって与えられているならば，二つの異なるモデル ω_1, ω_2 について，

$$\pi_{\mathrm{S}_\theta}(\omega_1\,|\,\boldsymbol{Z})=\pi(\omega_1\,|\,\boldsymbol{Z}, \bigcup_i \mathrm{S}_{\theta_i})$$

$$\pi_{\mathrm{S}_\theta}(\omega_2\,|\,\boldsymbol{Z})=\pi(\omega_2\,|\,\boldsymbol{Z}, \bigcup_i \mathrm{S}_{\theta_i})$$

をつくるとき，

$$\pi_{\mathrm{S}_\theta}(\omega_1\,|\,\boldsymbol{Z})\geqq\pi_{\mathrm{S}_\theta}(\omega_2\,|\,\boldsymbol{Z})$$

が充されるならば，ω_1 は ω_2 に対して preferable である。

なお $\pi_{\mathrm{S}_\theta}(\omega_i\,|\,\boldsymbol{Z})$ はくわしく書けば，

$$\pi_{\mathrm{S}_\theta}(\omega_i\,|\,\boldsymbol{Z})=\frac{\pi(\omega_i)\displaystyle\int_{\mathrm{S}_{\theta_i}} p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}\,|\,\boldsymbol{\theta}_i, \omega_i)d\boldsymbol{\theta}_i}{\displaystyle\sum_i\pi(\omega_i)\int_{\mathrm{S}_{\theta_i}} p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}\,|\,\boldsymbol{\theta}_i, \omega_i)d\boldsymbol{\theta}_i}$$

となる。但し，この表現においては，$p(\boldsymbol{\Theta}_i|\omega_i)$, $f(\boldsymbol{Z}|\boldsymbol{\Theta}_i, \omega_i)$ はいずれも S_{Θ_i} において $\boldsymbol{\Theta}_i$ の連続関数であることを前提としている。また，このルールが定義 11 によって与えられる選択ルールの拡張——$\bigcup_i \hat{\boldsymbol{\Theta}}_i$ をとる代りに $\bigcup_i S_{\Theta_i}$ をとっており，$\bigcup_i S_{\Theta_i}$ は $\bigcup_i \hat{\boldsymbol{\Theta}}_i$ を含む——ことは明かであるが，さらに $\boldsymbol{D}_1, \boldsymbol{D}_2$ を充分に広くとれば S_{Θ_i} は Θ_i と一致するに至るであろうから，定義 10 によって与えられる選択ルールをも特別の場合として含むと考えてもよいわけである。

以上によって，われわれは曲りなりにもモデル選択問題におけるベイジアン・ルールの基本的な方向づけをなしえたわけだが，ここでこのルールに関連して重要な二つの問題にふれておく必要がある。一つは逐次選択の問題であり，いま一つは構造変化がある場合のモデル選択である。

§ 4.1. モデルの逐次選択について

モデルの逐次選択というのは，第 0 時点で observation $\boldsymbol{Z}_0$ がえられ，次いで第 1 時点で observation $\boldsymbol{Z}_1$ がえられ，以下順に $\boldsymbol{Z}_2, \boldsymbol{Z}_3, \ldots$ 等の observation がえられたとき，こうした observation の累積がもたらす情報の増加につれて，逐次的にわれわれの preference を修正し，より正しい preference に接近していくことである。

パラメーターについて，その事後確率をあらためて事前確率にとり，これに新しい observation にもとずく尤度関数を乗じたものに比例するように新しい事後確率を定めるというベイジアンの逐次推測のルールはよく知られている。ここでの問題はこれと平行したアプローチをモデル選択の場合についても考え，その性格を明らかにすることである。

ベイジアンの逐次モデル選択ルールは，パラメーターの推論の場合と平行に，次のように考えることができよう。すなわち，observation $\boldsymbol{Z}_0$ によるモデル ω_i に対するわれわれの選択基準としての事後確率，

$$\pi_{S^0_\Theta}(\omega_i \mid \boldsymbol{Z}_0)=\pi(\omega_i \mid \boldsymbol{Z}_0,\ \bigcup_i S^0_{\Theta_i})$$

を observation $\boldsymbol{Z}_1$ をえた次の段階での事前確率として，最初の段階での $\pi(\omega_i)$ におきかえ，

(26) $$\pi_{S^1_{\Theta_i}}(\omega_i \mid \boldsymbol{Z}_0, \boldsymbol{Z}_1)$$

$$=\frac{\pi(\omega_i \mid \boldsymbol{Z}_0,\ \bigcup_i S^0_{\Theta_i})\int_{S^1_{\Theta_i}} p(\boldsymbol{\Theta}_i \mid \boldsymbol{Z}_0, \omega_i) f(\boldsymbol{Z}_1 \mid \boldsymbol{Z}_0, \boldsymbol{\Theta}_i, \omega_i)d\boldsymbol{\Theta}_i}{\sum \pi(\omega_i \mid \boldsymbol{Z}_0,\ \bigcup_i S^0_{\Theta_i})\int_{S^1_{\Theta_i}} p(\boldsymbol{\Theta}_i \mid \boldsymbol{Z}_0, \omega_i) f(\boldsymbol{Z}_1 \mid \boldsymbol{Z}_0, \boldsymbol{\Theta}_i, \omega_i)d\boldsymbol{\Theta}_i}$$

によって $\boldsymbol{Z}_0$ につづいて $\boldsymbol{Z}_1$ をえたときの ω_i の事後確率を評価し，preference を行なうということである。

ここで $S^0_{\Theta_i}$ は上述の手続きによって observation $\boldsymbol{Z}_0$ の下で定まる領域であり，$S^1_{\Theta_i}$ は続いてえられた $\boldsymbol{Z}_1$ の下で，$p(\boldsymbol{\Theta}_i \mid \boldsymbol{Z}_0, \omega_i) f(\boldsymbol{Z}_i \mid \boldsymbol{Z}_0, \boldsymbol{\Theta}_i, \omega_i)$ を最大にする $\hat{\boldsymbol{\Theta}}^1_i$ によって，同様の手続きによって定められる領域である。また言うまでもないことだが $\boldsymbol{Z}_0$ と $\boldsymbol{Z}_1$ が互いに独立であるならば，$f(\boldsymbol{Z}_1 \mid \boldsymbol{Z}_0, \boldsymbol{\Theta}_i, \omega_i)=f(\boldsymbol{Z}_1 \mid \boldsymbol{\Theta}_i, \omega_i)$ である。

ところでいま $\boldsymbol{Z}_0$ と $\boldsymbol{Z}_1$ が同時に与えられたとすると，われわれのモデル選択は，

(27) $$\pi_{S_\Theta}(\omega_i \mid \boldsymbol{Z}_0, \boldsymbol{Z}_1)=\frac{\pi(\omega_i)\int_{S_{\Theta_i}} p(\boldsymbol{\Theta}_i \mid \omega_i) f(\boldsymbol{Z}_0, \boldsymbol{Z}_1 \mid \boldsymbol{\Theta}_i, \omega_i)d\boldsymbol{\Theta}_i}{\sum_i \pi(\omega_i)\int_{S_{\Theta_i}} p(\boldsymbol{\Theta}_i \mid \omega_i) f(\boldsymbol{Z}_0, \boldsymbol{Z}_i \mid \boldsymbol{\Theta}_i, \omega_i)d\boldsymbol{\Theta}_i}$$

によって決定されるであろう。ここに S_{Θ_i} は，$p(\boldsymbol{\Theta}_i \mid \omega_i) f(\boldsymbol{Z}_0, \boldsymbol{Z}_1 \mid \boldsymbol{\Theta}_i, \omega_i)$ を最大にする $\hat{\boldsymbol{\Theta}}_i$ によって前に述べた手続きによって決定される領域である。

いま $\boldsymbol{Z}_1$ を得た段階で考えてみると，すでに $\boldsymbol{Z}_0$ は得られているわけだから，われわれがもっているところの observation を通して得られる情報量は，正しいモデルが第0時点でも第1時点でも同一であり，また，この間構造変化もなく，したがって $\boldsymbol{Z}_0$, $\boldsymbol{Z}_1$ は同一条件の下でもたらされたと考えられるならば，$\boldsymbol{Z}_0, \boldsymbol{Z}_1$ を同時に得た場合と，実質的に異なるものではない。したがって，われわれがモデルに対する同一の事前確率から出発し，observation $\boldsymbol{Z}_0$ と $\boldsymbol{Z}_1$ のみからえられる情報によって事後確率を評価する以上，このような評価が consistent であるためには，(26)と(27)が等しくなることが期待さるべきであろう。実際，パラメーターに対するベイジアンの推論では，(26)に対応する逐次的推論と(27)に対応する同時的推論は一致する。このような consistency をここでは recursive consistency と呼ぶことにしょう。そこで次に，モデル選択に関するわれわれのベイジアン・ルールについて，この recursive consistency が保たれるかどうかを吟味してみよう。

いま $C_i^{-1}=\int_{\Theta_i} p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}_0\,|\,\boldsymbol{\theta}_i,\omega_i)d\boldsymbol{\theta}_i$ とおくと，

$$p(\boldsymbol{\theta}_i\,|\,\boldsymbol{Z}_0,\omega_i)=C_i\cdot p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}_0\,|\,\boldsymbol{\theta}_i,\omega_i)$$

であるから，これを(26)に代入すると，

$$\pi_{S^1_\Theta}(\omega_i\,|\,\boldsymbol{Z}_0,\boldsymbol{Z}_1)$$

$$=\frac{\pi(\omega_i\,|\,\boldsymbol{Z}_0,\bigcup_i S^0_{\Theta_i})\cdot C_i\int_{S^1_{\Theta_i}} p(\boldsymbol{\theta}_i|\omega_i)f(\boldsymbol{Z}_0|\boldsymbol{\theta}_i,\omega_i)f(\boldsymbol{Z}_1|\boldsymbol{Z}_0,\boldsymbol{\theta}_i,\omega_i)d\boldsymbol{\theta}_i}{\sum_i\pi(\omega_i\,|\,\boldsymbol{Z}_0,\bigcup_i S^0_{\Theta_i})\cdot C_i\int_{S^1_{\Theta_i}} p(\boldsymbol{\theta}_i|\omega_i)f(\boldsymbol{Z}_0|\boldsymbol{\theta}_i,\omega_i)f(\boldsymbol{Z}_1|\boldsymbol{Z}_0,\boldsymbol{\theta}_i,\omega_i)d\boldsymbol{\theta}_i}$$

$$=\frac{\pi(\omega_i\,|\,\boldsymbol{Z}_0,\bigcup_i S^0_{\Theta_i})\cdot C_i\int_{S^1_{\Theta_i}} p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}_0,\boldsymbol{Z}_1\,|\,\boldsymbol{\theta}_i,\omega_i)d\boldsymbol{\theta}_i}{\sum_i\pi(\omega_i\,|\,\boldsymbol{Z}_0,\bigcup_i S^0_{\Theta_i})\cdot C_i\int_{S_{\Theta_i}} p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}_0,\boldsymbol{Z}_1\,|\,\boldsymbol{\theta}_i,\omega_i)d\boldsymbol{\theta}_i}$$

さらに，

$$\pi(\omega_i\,|\,\boldsymbol{Z}_0,\bigcup_i S^0_{\Theta_i})=\frac{\pi(\omega_i)\int_{S^0_{\Theta_i}} p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}_0\,|\,\boldsymbol{\theta}_i,\omega_i)d\boldsymbol{\theta}_i}{\sum_i\pi(\omega_i)\int_{S^0_{\Theta_i}} p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}_0\,|\,\boldsymbol{\theta}_i,\omega_i)d\boldsymbol{\theta}_i}$$

を代入し，また $\sum_i\pi(\omega_i)\int_{S^0_{\Theta_1}} p(\boldsymbol{\theta}_i\,|\,\omega_i)f(\boldsymbol{Z}_0\,|\,\boldsymbol{\theta}_i,\omega_i)\,d\boldsymbol{\theta}_i$ が i に依存しない定数であることに注目して約分すれば，

$$(28)\quad \pi_{S^1_\Theta}(\omega_i|\boldsymbol{Z}_0,\boldsymbol{Z}_1,\cup S_{\Theta_i})=\frac{w_i\pi(\omega_i)\int_{S^1_{\Theta_i}} p(\boldsymbol{\theta}_i|\omega_i)f(\boldsymbol{Z}_0,\boldsymbol{Z}_1|\boldsymbol{\theta}_i,\omega_i)d\boldsymbol{\theta}_i}{\sum_i w_i\pi(\omega_i)\int_{S^1_{\Theta_i}} p(\boldsymbol{\theta}_i|\omega_i)f(\boldsymbol{Z}_0,\boldsymbol{Z}_1|\boldsymbol{\theta}_i,\omega_i)d\boldsymbol{\theta}_i}$$

をうる。但しここで，

$$w_i = C_i \int_{S^0_{\theta_i}} p(\boldsymbol{\theta}_i \mid \omega_i) f(\boldsymbol{Z}_0 \mid \boldsymbol{\theta}_i, \omega_i) d\boldsymbol{\theta}_i$$

で，これは $\boldsymbol{Z}_0$ を得た段階でのパラメーター $\boldsymbol{\theta}_i$ の事後確率分布において，パラメーター $\boldsymbol{\theta}_i$ が $S_{\theta^0_i}$ なる領域に入る確率を表わしている。

パラメーターの推論に関しては recursive consistency，すなわち，

$$p(\boldsymbol{\theta}_i \mid \boldsymbol{Z}_0, \omega_i) \cdot f(\boldsymbol{Z}_1 \mid \boldsymbol{Z}_0, \boldsymbol{\theta}_i, \omega_i) \propto p(\boldsymbol{\theta}_i \mid \omega_i) f(\boldsymbol{Z}_0, \boldsymbol{Z}_1 \mid \boldsymbol{\theta}_i, \omega_i)$$

がすべての ω_i について成り立っているから，$\max_{\theta_i} p(\boldsymbol{\theta}_i \mid \boldsymbol{Z}_0, \omega_i) f(\boldsymbol{Z}_1 \mid \boldsymbol{Z}_0, \boldsymbol{\theta}_i, \omega_1)$ も $\max_{\theta_i} p(\boldsymbol{\theta}_i \mid \omega_i) f(\boldsymbol{Z}_0, \boldsymbol{Z}_1 \mid \boldsymbol{\theta}_i, \omega_i)$ も同一の $\hat{\boldsymbol{\theta}}_i$ を与え，また，これから定められる $S^1_{\theta_i}$, S_{θ_i} はまったく同一の領域を与える。すなわち $S^1_{\theta_i} = S_{\theta_i}$ である。このことに留意して(27)と(28)とを較べると，両者が等しくなるための必要かつ十分な条件は，

(28)　$w_1 = w_2 = \cdots\cdots = w_i = \cdots\cdots$

がすべての i について成り立つことであることがわかる。

ところで既に指摘したように w_i はモデル ω_i において observation $\boldsymbol{Z}_0$ を与えたときのパラメーター $\boldsymbol{\theta}_i$ の事後確率分布で $\boldsymbol{\theta}_i$ が $S^0_{\theta_i}$ に入る確率になっている。したがって(29)が成り立つということは，すべてのモデルについて，この確率が等しいということであるが，一般にこれを期待することはできない。しかし特別な場合として $S^0_{\theta_i} = \Theta_i$ となっているとき，すなわち，われわれのベイジアン選択ルール 1 の下では，いうまでもなく $w_1 = w_2 = \cdots = w_i = \cdots = 1$ となっているから，recursive consistency が成り立つ。すなわち，

定理 1．(26)によって定義されるベイジアンの逐次選択が recursive consistency をもつための必要かつ十分な条件は，すべてのモデルについて $\int_{S^0_{\theta_i}} p(\boldsymbol{\theta}_i \mid \boldsymbol{Z}_0, \omega_i) d\boldsymbol{\theta}_i$ が等しくなっていることである。

系．モデル ω_i に対する事後確率を $\pi(\omega_i \mid \boldsymbol{Z})$ で評価するベイジアン選択ル

ール1の下での逐次選択は recursive consistency をもつ。

が得られる。

ところでこの recursive consistency は，$\boldsymbol{Z}_i$ を得た段階での事前確率として $\max_{\Theta_i} p(\boldsymbol{\theta}_i|\boldsymbol{Z}_0, \omega_i)f(\boldsymbol{Z}_1|\boldsymbol{Z}_0, \boldsymbol{\theta}_i, \omega_i)$ から定められる $S^1_{\Theta_i}$ を各モデルについて固定して，$\pi_{S^1_{\Theta_i}}(\omega_i|\boldsymbol{Z}_0)=\pi(\omega_i|\boldsymbol{Z}_0, \bigcup_i S^1_{\Theta_i})$ を用いることにしても保たれない。なぜならば，これは，

$$(30)\quad w'_i=C_i\int_{S^1_{\Theta_i}} p(\boldsymbol{\theta}_i|\omega_i)f(\boldsymbol{Z}_0|\boldsymbol{\theta}_i, \omega_i)d\boldsymbol{\theta}_i=\int_{S^1_{\Theta_i}} p(\boldsymbol{\theta}_i|\boldsymbol{Z}_0, \omega_i)d\boldsymbol{\theta}_i$$

とおけば，この場合の事後確率を表現する式は(28)の w_i を w'_i でおきかえただけのものになり，これがすべてのモデルについて等しくなるという保証は与えられないからである。すなわち，

定理 2. S_{Θ_i} をパラメーターに関する最終段階での事後確率分布によって定める逐次選択ルールが，recursive consistency をもつための必要かつ十分な条件は，すべてのモデルのパラメーターのその前の段階での事後確率分布（その段階での事前確率分布）において，パラメーター $\boldsymbol{\theta}_i$ がこの領域に入る確率が，すべてのモデルについて等しいということが，逐次選択のいかなる段階においても成り立っていることである。

が一般に成り立つのである。

それゆえ，パラメーター空間に局所領域を限定して行なわれるベイジアンのモデル選択ルールでは，一般に，recursive consistency は維持されず，したがって厳密には逐次選択ルールは採用できない。しかし筆者はこのことから実際上，逐次選択がこのベイジアン・ルールの下で許されないとは考えない。

(27)と(28)とを比較すると，前者では $\sum_i \pi(\omega_i)\int_{S_{\Theta_i}} p(\boldsymbol{\theta}_i|\omega_i)f(\boldsymbol{Z}_0, \boldsymbol{Z}_1|\boldsymbol{\theta}_i, \omega_i)$

$d\boldsymbol{\Theta}_i$ に占める $\pi(\omega_i)\cdot\int_{S_{\Theta_i}} p(\boldsymbol{\Theta}_i|\omega_i)f(\boldsymbol{Z}_0,\ \boldsymbol{Z}_1|\boldsymbol{\Theta}_i,\ \omega_i)d\boldsymbol{\Theta}_i$ の share によって ω_i の事後確率を評価しているのに対して，後者では，これに w_i なる weight を付して評価している点が異なるだけである。そしてこの w_i は，$\int_{S^0_{\Theta_i}} p(\boldsymbol{\Theta}_i|\boldsymbol{Z}_0,\ \omega_i)d\boldsymbol{\Theta}_i$ であり，モデル ω_i に対する $\boldsymbol{Z}_0$ をえた段階での事後確率の評価にあたって，パラメーター $\boldsymbol{\Theta}_i$ に対する事後確率分布の局所的な性質が，その評価の条件として利用されたために(28)の表現の中に入り込んで来たのである。筆者はこれをモデル ω_i に対する $\boldsymbol{Z}_0$ による prejudice effect と呼ぶ。これが prejudice であるというのは，第0時点と第1時点の間に構造変化がない以上，$\boldsymbol{Z}_0, \boldsymbol{Z}_0$ が逐次的に与えられても，二つの時点の間で他の特殊な情報が与えられていない限り，われわれの情報量は不変のはずであり，われわれがそれぞれのモデルに対してもつところの判断は，いちばん最初の事前分布 $\pi(\omega_i)$ と $\boldsymbol{Z}_0, \boldsymbol{Z}_1$ 等逐次的に与えられる observation のみに基ずいて行なわれてこそ公平なのであって，このためには recursive consistency は当然保たれていなければならず，w_i によって加重する根拠は利用される情報の状況から見てまったく見出しえないからである。

しかし，もし $\boldsymbol{Z}_0. \boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$ 等の observation の無限の系列が与えられたとき，これらを同時に与えられたものとして行なう preference が，asymptotic に逐次選択がもたらすところの preference と同一のものになるならば，そしてさらにこうした preference の収束がわりと早いものであるならば，逐次選択の各段階で入り込んでくる prejudice effect は無視してよいことになろう。すなわち逐次選択はその recursive inconsistency にもかかわらず採用できるわけである。

残念ながら筆者はまだこの asymptotic property を確かめえていないが，パラメーターの事後分布については，それが次第に一定の分布に収束していくことが保証されるから，一般にまたはゆるい条件の下で，この asymptotic property が示せるのではないかと予想している。

§ 4.2.　構造変化がある場合のベイジアン・ルール

構造変化がある場合のモデル選択について，一般性のあるベイジアン・ルールを見出すことは容易でないように思われる。しかし，二つの時点の間で正しいモデルが別個にあると考えないで，正しいモデルはあくまでも一つであるが，この二つの時点の間で構造パラメーターが変化しているという前提の下での選択ルールについてはは，上述の議論の延長として次のような考え方をすることが許されるであろう。

まず observation $\boldsymbol{Z}$ は $\boldsymbol{Z}=(\boldsymbol{Y}, \boldsymbol{X})$ という形をしているものとしよう。第0時点の observation を $\boldsymbol{Z}^0=(\boldsymbol{Y}^0, \boldsymbol{X}^0)$, 第1時点の observation を $\boldsymbol{Z}^1=(\boldsymbol{Y}^1, \boldsymbol{X}^1)$ とする。この場合われわれは，第0時点でそれぞれのモデルのパラメーター $\boldsymbol{\theta}_i^0$ を推定するとともに，第1時点におけるパラメーター $\boldsymbol{\theta}_i^1$ との差，$\Delta\boldsymbol{\theta}_i=\boldsymbol{\theta}_i^1-\boldsymbol{\theta}_i^0$ を構造変化として推定しなければならない。

この推論は，通常，$\boldsymbol{Z}^0$ と $\boldsymbol{Z}^1$ から作られるある適当な統計量 $\mathfrak{D}=t(\boldsymbol{Z}_0, \boldsymbol{Z}_1)$ によって，ベイジアン・ルールの下では，

$$p(\Delta\boldsymbol{\theta}_i \mid \mathfrak{D}, \omega_i) \propto p(\Delta\boldsymbol{\theta}_i \mid \omega_i) f(\mathfrak{D} \mid \Delta\boldsymbol{\theta}_i, \omega_i)$$

の形で行なわれるであろう。ここで $p(\Delta\boldsymbol{\theta}_i \mid \omega_i)$ は構造変化 $\Delta\boldsymbol{\theta}_i$ に対するモデル ω_i の下での事前確率密度を表わし，$f(\mathfrak{D} \mid \Delta\boldsymbol{\theta}_i, \omega_i)$ は $\omega_i, \mathfrak{D}$ を所与とするとき，$\Delta\boldsymbol{\theta}_i$ の尤度である。この推論で定まる $\Delta\boldsymbol{\theta}_i$ の事後確率分布 $p(\Delta\boldsymbol{\theta}_i \mid \mathfrak{D}, \omega_i)$ のモードを $\Delta\hat{\boldsymbol{\theta}}_i$ としょう。すなわち，

$$p(\Delta\hat{\boldsymbol{\theta}}_i \mid \mathfrak{D}, \omega_i)=\max_{\Delta\theta_i} p(\Delta\boldsymbol{\theta}_i \mid \mathfrak{D}, \omega_i)$$

である。

次に，われわれの選択の対象となるモデルが §2. で述べた意味で consistent であるとすると，$\boldsymbol{X}_0$ を所与として，$\hat{\boldsymbol{\theta}}_i^0$ の下での $\boldsymbol{Y}$ と $\hat{\boldsymbol{\theta}}_i^0+\Delta\hat{\boldsymbol{\theta}}_i$ の下での $\boldsymbol{Y}$ とを誘導形によって定めることができる。これをそれぞれ $\hat{\boldsymbol{Y}}_i^0, \hat{\boldsymbol{Y}}_i^1$ と表わそう。ここで $\hat{\boldsymbol{\theta}}_i^0$ は第0時点での observation $\boldsymbol{Z}_0$ によって定められたパラメーターの事後確率分布のモードである。そこで $\Delta\boldsymbol{Y}_i=\boldsymbol{Y}^0-\hat{\boldsymbol{Y}}_i^0$ より，

$$\tilde{\boldsymbol{Y}}^0_i = \hat{\boldsymbol{Y}}^1_i + \Delta \boldsymbol{Y}_i$$

をつくると，これは内生変数 $\boldsymbol{Y}$ の $\hat{\boldsymbol{\Theta}}^0_i$ の下での $\boldsymbol{Y}$ の推定ベクトルに対する observation の残差のパタンを保存しながら，$\Delta\hat{\boldsymbol{\Theta}}_i$ なるパラメーター・シフトに対応するシフトを $\boldsymbol{Y}$ の observation に与えたことになる。いま構造変化は，構造パラメーター $\boldsymbol{\Theta}_i$ についてのみ起り，random disturbance の分散等，他の確率分布パラメーターは不変であると考えることが許されるならば，こうして得られた $\tilde{\boldsymbol{Y}}^0_i$ と $\boldsymbol{X}^0$ との組み合せ，

$$\tilde{\boldsymbol{Z}}^0_i = (\tilde{\boldsymbol{Y}}^0_i, \boldsymbol{X}^0)$$

は，モデル ω_i の下で，第1時点に関して定常化された第0時点での observation とみなすことができるであろう。

そこで第1時点において，構造パラメーター $\hat{\boldsymbol{\Theta}}^0_i + \Delta\hat{\boldsymbol{\Theta}}_i$ の下で，$\boldsymbol{Z}^1_i = (\boldsymbol{Z}_1, \tilde{\boldsymbol{Z}}^0_i)$ が与えられたものとして，各モデル毎に，

$$\pi(\omega_i \mid \boldsymbol{Z}^1_i, \bigcup_i (\hat{\boldsymbol{\Theta}}^0_i + \Delta\hat{\boldsymbol{\Theta}}_i))$$

を評価し，われわれの選択ルールを適用すればよい。

この場合の選択は第1時点でパラメーター空間上の1点をそれぞれのモデルにおいて限定して行なうことになるが，この点を中心に適当な部分領域を限定する場合への拡張も，また各時点での observation $\boldsymbol{Z}^t = (\boldsymbol{Y}^t, \boldsymbol{X}^t)$ が複数個与えられ，observation が行列形式で表現される場合への拡張も容易である。

§5. 計量経済学による実証の限界と経済分析における主観確率の意義

計量経済学における inference が，所与のモデルにもとずいて行なわれる以上，それが経済理論に提示する一切の evidence は，そのモデルが確実であるという前提の下で評価されるものだという意味で，conditional である。このような conditional な evidence が，経済理論を検証するというのは，次のような論理にもとずく。

すなわち，絶対的な意味でそれが確実であるかどうかは別として，その経済理論のある部分を真であると仮定して，これをモデルとして前提するとき，いまこの理論の下で必然的に導かれねばならないある命題に対して，この命題が成立しているならば，本来稀にしかその実現が期待できないような observation をえたとすると，この observation はこの命題に対して否定的であると考えられ，この命題が所与の理論の下で必らず成立しなければならないものである以上，この observational な negation は理論自体に対する negation と考えることができるであろう。

この論理は Neyman-Pearson の統計理論——特に仮説検定論——をつらぬくものであるが，それはまさに論理実証主義の論理であり，Haavelmo がモデルを ideal な実験計画だと言ったのも，このような論理に立ってのことだったのである。そして筆者は，この論理自体は否定しがたい正当性をもつと考える。

いま少し実際に即して言うならば，モデルは不確定のパラメーターを含んでいるから，まずそのパラメーターを当の observation がえられる確率をなるべく大きくするように，例えば observation が内生変数と外生変数について与えられ，外生変数は確定変数だとすれば，各構造方程式に対してなんらかの意味で内生変数の全体的な fit がもっともよくなるような方式，例えば最小二乗法なり最尤法なりによって推定し，この推定パラメーターの下で，なおかつそれがえられる確率がきわめて小であるような observation が残っているとするならば，あるいはまたこうして推定されたパラメーターが理論の要求するところと明らかに矛盾したものであるとするならば，これをそのモデルが正しくなく，したがってそのモデルを導いた理論も正しくないことを示唆する evidence とみなすことができるであろう。推定パラメーターが理論と矛盾する場合を別にすると，これは observation に対してもっとも fit のよいパラメーターの下で，なおかつ内生変数の observational な variation が非常に大きいという形になって現われるであろうから，§3. で

示したモデル選択の諸原則を justify することになる。

だがここで注意すべきは，このような論法は確率モデルにおいては，常に negation のみが積極的な意味をもつということである。もし当の observation がえられる確率が十分に小さくないならば，このような negation ができないのはもちろんであるが，当の observation がえられる確率が，そのモデルと推定パラメーターの下で十分に大きくとも，なおかつそのモデルならびにそのモデルを導くところの理論の正当性を実証したことにはならないのである。

さらに重大なことは，observation が連続的な変量である場合を考えれば明らかであるが，このような判定にあたって当の observation がえられる確率のみを考えることは一般に無意味で，その observation がそのモデルの下では稀にしか実現されないと期待され，かつ理論に照らして異常と考えうるようなある特殊な領域に入るかどうかということで判定されなければならないが，ひとたび所与の observation にもとずいてパラメーターの推定を行なうと，この推定パラメーターの下では，もはや observation がこの領域に入ることは，この手続き自体のためにほとんどなくなってしまうのである。したがって，それにもかかわらず，この observation によってモデルおよび理論の negation を可能ならしめるためには，そのモデルおよび理論が正しいならば，推定パラメーターの下での構造方程式からもたらされる内生変数の理論値からの observation の deviation がある所与の限界をこえる確率はきわめて稀であるということが確実に根拠づけられていなければならない。このことはいわゆる random disturbance の variation について，われわれが検証しようとする理論にもとずいて必然的なある種の仮定なり，あるいは現実に根ざした information なりが与えられていなければならないことを意味する。しかし，このことを一般に期待することはほとんど不可能であると筆者は考えている。

推定パラメーターが理論と矛盾する場合についても，こうした事情は本質

的には変わらない。この場合，理論はパラメーターが動く範囲について，ある限定を課すことを必然ならしめる論理的根拠を含むものでなければならないが，こうした限定はあらかじめモデルの中に折り込んで，例えば制約条件つき最小二乗法等によって，理論に矛盾しない推定パラメーターを求めることもできるから，問題はやはり推定パラメーターの下での構造方程式に対する observation の fit の問題に帰着するのである。

random disturbance の中に，本来 stochastic とは考えられないが，把握困難な諸要因の働きを便宜上含めて考えているところから，この variation が大きいということはたしかに，モデルが現象を完全に説明するものにはなっていないということを意味するかも知れないが——そしてこの意味でなら，そのモデルなり理論なりが正しくないということもできようが——例えば生産における要素投入の配分を企業家が限界原理に従って行なっているというメカニズムを表わす構造方程式の攪乱項，つまりなんらかの偶然的な事情なり，簡単にその要因が把握できないその他の事情によって実際の配分が限界原理によって strictly に決定される点からそれて決定されていることが観察された場合の両者の間の deviation が大きいということは，限界原理がこうした deviation をもたらす諸要因と結合しながらも働いているということ——このことを当の構造方程式は表わしている——自体を否定することにはならない。もし，われわれが random disturbance に含めていた諸要因の中からある少数のものをとり出して，それによって格段に fit のよいモデルなり理論なりをつくることができれば，もちろんそうすることが望ましい。そしてそれは理論に対する一つの発見となり，理論の充実をもたらすことになるだろう。しかし，多くの場合これは簡単にはできない。また，仮りにできたとしても，最初の構造方程式で十分に説明できなかった部分を補充したにとどまるのであって，最初のモデルなり理論なりが表現していた経済メカニズムが存在しないということになるのではない。

したがって，論理的には可能と思われる理論の negative な実証すらも，

実際の計量経済学の分析手続きに照らして考えてみると一般には困難なのである。

それゆえ筆者は，計量経済学は経済理論に対して，その正否を疑いの余地なく明らかにするほどに決定的な evidence を与えるものとしてではなく，経済理論をさらに深めていくための素材をその conditional evidence の提供を通じて与えるものとして評価すべきであると考えている。

モデル選択におけるベイジアン・アプローチは，このような素材を適切に利用していく上での一つの手がかりを与えるものと言えよう。われわれがモデルに対して決定できるのは，そのモデルが絶対的な意味で正しいかどうかではなく，そのモデルが evidence に照らして他のモデルに対して preferable であるかどうかということであり，さらにわれわれは当の evidence 以外のもろもろの a priori information をもっているのが通常であるから，こうした a priori information にもとずいて形づくられたそれぞれの理論なりモデルなりに対するわれわれの degree of belief が evidence に照らしていかに修正され，われわれの認識の深まりの各段階で，どのようなモデルがもっとも preferable と考えられるかということを決定しうるにとどまるのである。

このことはまた，実験が許されない経済学では「抽象」が基本的な方法とならねばならないという Marx [4] の指摘を想起させる。すなわち，われわれが選択の対象とするところのモデルならびにそれを導く理論は，まずなによりも経済学者のすぐれた洞察力，つまり何を本質的なものとして抽象し，何を捨象するかという直観的な決定の能力によってもたらされなければならないのはもちろんであるが，さらに今日の発達した計量経済学の手法をもってしても，実証的見地からどの理論が正しく，どの理論が誤まりであるかを疑いの余地なく決定することはできないのであって，それはやはり究極的には経済学者の本質洞察力に支えられた確信によってのみ決定される他はないのである。

だが経済学者が，彼に提供されるいかなる observation によっても彼の確

信を修正することをかたくなに拒むならば，その理論はドグマと化そう。こうした observation によって a priori な彼の確信を絶えず修正し，より確かな理論を追求し続けてこそ経済理論の進歩は生み出され，その実証性も高められるのである。計量経済学が提出する evidence は，それ自体こうした observation の一つとしての意味を経済理論に対して有しているのであって，経済学者がより実証性の高い理論をつくり出し，さらに本質的な認識に進んでいくための素材なのである。

§4. で示したモデル選択のベイジアン・ルールは，計量経済学が提示する evidence によって経済学者の確信を systematic に修正する一つの方式を与えるものである。それはあくまで可能な方法の一つであるに過ぎない。ましてこの方法の implication が十分に展開し切れていない現在では，それは一つの提案にとどまっている。しかしそれは，計量経済学によって提供される evidence にもとずいて理論を深めていこうとする際に，われわれが直面する本質的な状況に即したアプローチであると筆者は考える。

ベイジアン・アプローチについてしばしば指摘される難点は，事前確率をいかにして consistent に定めうるかということである。しかも §4. における事前確率が，文字通り研究者の degree of belief, すなわち主観確率として考えられている点について，およそ科学がもつべき客観性を損ねるのではないかという非難が付加されるかも知れない。

筆者は後者の批判については，モデル選択という問題の性格から言って，この場合あてはまらないと思う。実験を許されない経済学にあっては，理論の実証性と客観性は，逐次提出される evidence を理論に feed back するプロセスを通じて徐々に獲得されるのであって，このプロセスでは，研究者の主観的確信を客観的 evidence によっていかに修正するかということこそが問題の核心なのであって，この確信の強さを定量的に表現する主観確率を explicit に介在させることこそが，正しい，そしてむしろ客観的なアプローチと言うべきなのである。

ただ，この主観確率の分布をそれぞれの研究者が，自らの確信の強さに正確に照応するものとして，かつまた consistent に提示できるとか，明確に意識しているというふうに考えることは非現実的であろう。しかし，合理的な思考能力と自己を客観的に省察する能力とを備えた科学者であるならば，彼が可能だと考えるいくつかの alternative な理論ならびにそれから導かれるモデルについて，彼の degree of belief を consistent に定めることができないとは筆者は考えない。むしろ自己の理論に客観性と実証性を付与したいと真剣に考える研究者であるならば，まさにそのための前提として，彼の degree of belief の分布を明確化する労を払うべきであると考えたいのである。

計量経済学による分析結果が往々にして経済学上のわれわれの常識と明らかに相反する推定パラメーターを与えることがある。論理実証主義の立場からは，observation の正しさが疑いのないものだとすれば，われわれの常識は，したがってまたその土台となっている経済理論は廃棄されねばならないはずである。だが多くの場合，研究者はこのような革命的な主張を積極的に提示することを差控える。それはわれわれがもっている統計データの信頼度がこのような断定をあえて行なうほどに高いものではないという理由にもとずくものであるが，それでいて統計データのどのような欠陥がこのような誤まった evidence をもたらしたかということは，必らずしも明らかにされるとは限らないのである。それにもかかわらず，多くの研究者がこのような慎重な態度を示し，またそれが justify されるのは，筆者の考えるところでは，われわれにとってそれが常識であるほどにその経済理論の正しさに対するわれわれの degree of belief は高く，observation がこの degree of belief を低めるとしても，なおかつこの observation にもとずいた事後確率の意味での degree of belief が，その理論を否定するある alternative な理論に対する事後的な degree of belief を上回っているからなのである。

しかし，それが常識だからと言って，その常識にとって不利な observation

は observation そのものが誤まっているのだと常に考えるとすれば，それはおよそ非科学的な態度と言わねばなるまい。実際，こうした常識に反する observation が数多く提出されると研究者は自らの常識を疑い始める。そしてそれが新しい理論への道を開いて来たし，これからも開いていくだろう。このようなアプローチは常套化しているものであるが，どれだけの evidence の蓄積によっていままでの常識化されていた理論を廃棄するという決断に導くかということを，なんら formal な手続きを経ることなく，まったく研究者の恣意にゆだねるだけで十分であろうか。

こうした際に予想される不用な紛糾を避けるためにも，このような認識過程を explicit に formulate し，モデル選択の基本的な rule を合理的に定めることは有益であろう。

引 用 文 献

[1] Christ, C. F.: *Econometric Models and Methods*, New York, John Wiley & Sons, 1966.

[2] 福地崇生：計量経済学入門，東京，東洋経済新報社，昭和37年 (1962).

[3] Haavelmo, T.: *The Probability Approach in Econometrics*, Econometrica, Vol. 12 (1944), Supplement (July), 118 pp.

[4] カール・マルクス：資本論，第1版への序言，長谷部訳 第1巻第1分冊，日本評論社，112頁.

[5] Samuelson, P. A., T. C. Koopmans, and J. R. N. Stone: *Report of the Evaluative Committee for Econometrica*, Econometrica, Vol. 22 (1954), No. 2 (April), pp. 141–146.