構造パラメータの推定に関する考察

菊　地　　進

はじめに

計量経済学がその方法上の一大到達点として自負するものは，同時連立方程式体系によって経済過程を説明するという方法である。それは，この方法が自然科学における統計的方法の単なる機械的な適用ではなく，計量経済学に独自のモデル作成方式（＝同時方程式モデル方式）に基づく方法となっているからである。しかしながら，計量経済学者のこうした自負にもかかわらず，この方法においては，パラメータの推定方法が複数個存在し，実際に計量モデルの計測にたずさわる計量経済学者にとっては，それらのうちいずれを選択すべきであるかの判断が下せないといった問題が存在している。そのため，個々の計量経済学者は，各種方法による推定結果を並置して発表することを余儀なくされてきているのである。

筆者は，先に，「同時方程式モデルとその計測方法の展開について」（『立教経済学研究』第36巻第2号，1982年，12月）において，それら各種推定方法の開発理由をそれが開発された当初の事情にさかのぼって明らかにしておいた。というのは，計量経済学の通常のテキストにおいては，各種推定方法が何の脈絡もなく並列的に掲げられており，各方法が他の方法のどのような欠陥をどのようにカバーするために開発されているかという点が，本質的なところとしては何ひとつ明らかにされていない。そのため，それらのうちいずれを選択すべきであるかの指針も与えられなければ，そもそも相異なる複数個の推定方法が何故存在するのかという点も全く明ら

かになってこない。そこで，計量経済学者を悩ませているこうした事態が何故生じてくるのか，すなわち，推定方法の複数性の問題の本質は何かという点を考えるためには，各種推定方法の開発理由をそれが開発された当初の事情にさかのぼって明らかにしておくということが，その検討作業上の第一のステップとしてどうしても必要となってくるのである。筆者の前稿――とくにその二・三節――は，このような問題意識からまとめたものであった。

そこで，まず，筆者の前稿における結論を概括しておくことにしたい。同時方程式モデルのパラメータ推定方法が複数個存在するというのは，直接的には，パラメータの推定基準として一致性という推定特性――標本数を無限大にするならばその推定値は真値にほぼ一致するという性質――が採用されていることに帰因するものである。この一致性という特性は，元来は，最大尤度法とよばれる推定方法を客観的方法とみなすための一手段として推測統計学に導入された概念であった。

最大尤度法とは，未知パラメータを変数とみなした尤度関数が最大となるようにパラメータの推定値を定めるという方法で，自然科学の分野では前世紀の初頭以来実際に用いられてきていた。しかしながら，このように古い歴史をもつとはいえ，これは，方法論的には，主観的信頼度である尤度という概念を用いねばならないという難点をかかえていた。そこで，1930年代に，推測統計学の生みの親の一人であるＲ・Ａ・フィッシャーは，最尤推定量が一致性という推定特性をもつことを明らかにすることによって，最大尤度法の客観的方法としての根拠づけを行ったのである。

同時方程式モデルのパラメータ推定方法としてはじめに採用されたのは，実は，この最大尤度法であった。そして，その際，計量経済学においても，推測統計学の考え方にならい，パラメータの推定基準として一致性という推定特性が導入された。ところが，この一致性であれば，多元回帰モデル（＝単一方程式モデル）のパラメータ推定方法である最小二乗法であっても，その特性をもつ推定値を得ることは十分可能となるのである。ただし，この場合には，その適用の仕方に一定の工夫が必要ではあるが，ともかく，こうして，最小二乗法が工夫された方法が開発され，同時方程式モデルのパラメータ推定方法としては，相異なる結果を生む複数個の方法が存在することになったのである。

しかも，最大尤度法，最小二乗法，このそれぞれについて，さらに一定のより簡便な推定方法が開発され，同時方程式モデルのパラメータ推定方法としてはいかなる方法が最適であるかを確定することが実践上いよいよ重要な課題となってくるのである。そこで，計量経済学においては，1960年代から70年代にかけて，モンテ・カルロ法，推定量の漸近展開等により最適推定方法確定のための研究が積み上げられてきた。しかしながらこれまでのところ，これらの研究は一定の限定された条件下での考察にとどまり，そこでの結論を一般化することは到底不可能なものとなっている。こうして，冒頭でふれたように，個々の計量経済学者は，特定の推定方法のみに依拠するわけにいかず，各種方法による推定結果を並置して発表することを余儀なくされてきているのである。

そうすると，ここで考えてみなければならないのは，各方法はそもそもどのようなところにその独自性があるのかという点である。より簡便な推定方法が開発されたとはいえ，推定方法の選択問題は，基本的には，最大尤度法に基づく方法を選ぶか，最小二乗法に基づく方法を選ぶかという問題に集約されることになる。では，この両者それぞれの独自の意義はどのようなところにあるのであろうか。

まず，最大尤度法であるが，上述のように，これは計測材料としての統計データが実現値として最大確率をもつようにパラメータの推定値を定めるという方法である。すなわち，この方法は，正しく作成されたモデルの下でそのパラメータを最も精度よく推定しようとするところにその力点がある。ところが，一般にモデルが正しく作成されているという保証はない。そのため，例えば，多元回帰モデルの計測方法の場合であれば，推定結果に基づきモデルの変数選択の適否を吟味する機能が計測方法そのものに組み込まれているのである。しかしながら，最大尤度法による同時方程式モデルの計測の場合には，これまでのところこのような吟味は実際上不可能である。

そこで，同時方程式モデルの実際の計測にあたっては，上述の多元回帰モデルの変数選択の適否の吟味方法を援用して個々の方程式の適切性を検討し，その上で，最大尤度法に基づきパラメータの推定を行うという手順がふまれることになるのである。ところが，そうすると，推定結果についても，多元回帰モデルの計測方法である最小二乗法の適用結果に出来るだけ近いものを生みださねばならないように思

われてくる。なぜならば，そうでなければ，個々の方程式の適切性についての事前の検討が生きてこないように思われるからである。同時方程式モデルに適用可能なように最小二乗法が工夫されたというのは，実は，このような要請に応えるためであった。

しかしながら，そうした新しい方法が開発されたとはいえ，厳密には，多元回帰モデルの変数選択の適否を吟味する方法が同時方程式モデルの変数選択の適否を吟味する方法となりうるわけではない。そのため，この点をとくに強く念頭におく場合には，モデルの適切性の検証の問題についてはとりあえず別個の次元で考えることにし，推定方法それ自体としては正しく作成されたモデルのパラメータを最も精度よく推定できると見なされている最大尤度法に固執せざるをえないことにもなるのである。

同時方程式モデル方式における推定方法の複数性の問題は，明らかに，このような事情から生まれてきているということができる。したがって，推定方法の選択をめぐる混乱は，根本的には，モデルの適切性を明らかにすることにまつわる困難に帰因するものであると見なければならないのである。これが，パラメータ推定問題に関する筆者の前稿の結論であった。

では，同時方程式モデルの適切性を吟味する機能が組み込まれた計測方法を開発することは全く不可能なのであろうか。従来の方法にこのような機能を組み込むことはできないのであろうか。また，最適推定方法の選択をめぐるこれまでの研究は，こうした課題に応える方向に発展させることはできないのであろうか。もしそれが望みえないとするならば，その根本原因はどのようなところにあるのであろうか。

前稿では，あくまでも各種推定方法の歴史的成立プロセスを明らかにするところにその力点があるため，各種推定方法の理論構造に立ち入ることは必要な限りにとどめざるをえなかった。そのため，そこでは，上記の諸点の検討を行うまでにはいたっていない。そこで本稿では，同時方程式モデルの基本的仮定ならびに各種推定方法の理論構造，そしてまた，最適推定方法の選択をめぐる従来の研究について詳細にあとづけ，筆者にとって残された課題となっている以上の諸点についての検討を試みることにしたい。

I　構造パラメータの推定基準

計量経済モデルの計測は次のような単一方程式モデルとしての多元回帰モデル multiple regression model の計測から開始された。

$$y_t=\beta_0+\beta_1x_{1t}+\beta_2x_{2t}+\cdots\cdots+\beta_px_{pt}+u_t \qquad t=1,\ 2,\cdots\cdots,\ n$$

すなわち，

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u} \tag{1}$$

ただし，$\boldsymbol{y}'=(y_1,\ y_2,\ \cdots\cdots,\ y_n)$

$$\boldsymbol{X}=\begin{pmatrix} 1 & x_{11} & x_{21}\cdots\cdots\cdots & x_{p1} \\ 1 & x_{12} & \cdots\cdots\cdots\cdots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots\cdots\cdots\cdots & x_{pn} \end{pmatrix}$$

$\boldsymbol{\beta}'=(\beta_0,\ \beta_1,\ \cdots\cdots\cdots,\ \beta_p)$

$\boldsymbol{u}'=(u_1,\ u_2,\ \cdots\cdots\cdots,\ u_n)$

これは，経済変数 y の時系列的変動を他の p 個の変数 $x_1,\ x_2,\ \cdots\cdots,\ x_p$ によって説明しようとする方程式である。統計データが，

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots\cdots\beta_px_p$$

という関係を完全に満たすことはありえないため，上記のモデルにおいては攪乱項 u が明示的に導入されている。ところで，計量経済学の主たる仕事は実際の統計データから係数パラメータ $\boldsymbol{\beta}$ を推定するところにあるが，その際，通常，次のような仮定が設けられる。

〔I〕（i）　p 個の説明変数 x_i ($i=1,\ \cdots\cdots,\ p$) は指定変数——確率変数ではないということ——である。

（ii）　行列 $\boldsymbol{X}$ のランク（階数）は $1+p$ に等しい。

$$\rho(\boldsymbol{X})=1+p$$

（iii）　攪乱項 u_t ($t=1,\ 2,\ \cdots\cdots,\ n$) は，平均0，分散一定の確率分布に従う。

$$E(u_t)=0 \qquad E(u_t^2)=\sigma^2$$

（iv）　攪乱項 u_t ($t=1,\ 2,\ \cdots\cdots,\ n$) は系列的に無相関である。

$$E(u_i u_j)=0 \qquad i \neq j \qquad i,\ j=1,\ 2,\ \cdots\cdots,\ n$$

（v）攪乱項 $\boldsymbol{u}$ は多変量正規分布に従う。

$$\boldsymbol{u} \sim N(\boldsymbol{0},\ \sigma^2 \boldsymbol{I})$$

ただし，$\boldsymbol{I}$; 単位行列

仮定〔I〕（ii）は多重共線性の問題（後述）に関わるものである。また，仮定〔I〕（iii）より，被説明変数も確率変数となり，その期待値，分散は次のようになる。

$$E(\boldsymbol{y})=\boldsymbol{X\beta} \qquad V(\boldsymbol{y})=\sigma^2\boldsymbol{I} \qquad (2)$$

さて，以上のような諸仮定の下での係数パラメータ $\boldsymbol{\beta}$ の推定であるが，これは通常最小二乗法 least squares によって行われる。すなわち，まず，$\boldsymbol{\beta}$ に関する次のような二次形式を求める。

$$Q=(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})$$

次に，これを最小にするような $\boldsymbol{\beta}$ を求めるために，Q を $\boldsymbol{\beta}$ で徴分して $\boldsymbol{0}$ とおき，次のような連立方程式を導く。これは正規方程式 normal equations とよばれている。

$$(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}}=\boldsymbol{X}'\boldsymbol{y}$$

仮定〔I〕（ii）の下で，行列 $\boldsymbol{X}'\boldsymbol{X}$ の逆行列が存在するから*，未知パラメータ $\boldsymbol{\beta}$ の推定量は上式を解いて次のように求められる。

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \qquad (3)$$

こうして得られた $\hat{\boldsymbol{\beta}}$ は係数パラメータ $\boldsymbol{\beta}$ の最小二乗推定量とよばれている。

* 仮定〔I〕(ii) が満たされず $\rho(\boldsymbol{X})<1+p$ となる場合，すなわち，説明変数 $x_1, x_2, \cdots\cdots x_p$ の全部または一部の間に厳密な線形関係が成り立つ場合には，$|\boldsymbol{X}'\boldsymbol{X}|=0$ となり，$(\boldsymbol{X}'\boldsymbol{X})^{-1}$ が存在しないため，上記 $\hat{\boldsymbol{\beta}}$ を求めることは不可能となる。また，$|\boldsymbol{X}'\boldsymbol{X}|$ が厳密に0にならなくとも，それに近い場合には，$\hat{\boldsymbol{\beta}}$ の分散が著しく大きくなる。このような場合には説明変数間に多重共線性 multicollinearity があるという。したがって，こうした多重共線性を発生させるような変数をあらかじめ取り除いておくということが，計量モデル作成の条件のひとつとなるのである。cf. Frisch, R., *Statistical confluence analysis by means of complete regression systems*, 1934.

上記 $\hat{\boldsymbol{\beta}}$ が係数パラメータ $\boldsymbol{\beta}$ の最も望ましい推定量であると見なされる理由は，$\hat{\boldsymbol{\beta}}$ がその期待値をとると $\boldsymbol{\beta}$ に一致する不偏推定量でるという点と，それが $\boldsymbol{\beta}$ のあらゆる推定量のうち分散が最も小さい推定量であるという点の二つに集約される。

すなわち，この推定量に基づく推定を無限回くり返した場合，それらの推定値の平均が未知パラメータに一致し，かつまた，それらの推定値の真値からの偏差平方和が他のあらゆる推定量による場合よりも小さくなるということである。このような特性をもつ推定量は最良不偏推定量 best unbiased estimator とよばれている*。

* x_1, x_2, ……, x_n ($n \geqq 2$) を未知パラメータ θ をもつ密度関数 $f(x; \theta)$ からの無作為標本とする。ただし，この場合の各 x_i は実際の標本抽出の結果＝標本値を表わすのではなく，実際に抽出が行われる以前の標本変量——密度関数 $f(x; \theta)$ の確率変数——を表わすとする。また，t, t_1, t_2, ……, t_{r-1} をこの標本変量 x_i によって構成された関数的に独立な r 個の統計量（＝標本変量の関数）とする。ただし，$2 \leqq r \leqq n$ である。

この時，t が与えられた時の t_1, t_2, ……, t_{r-1} の条件付期待値が未知パラメータ θ に依存しないならば，t は θ に関する情報をすべて引き出しているという意味で，t を θ にたいする十分統計量sufficient statistic という。

一定の未知パラメータにたいする推定量 estimator（＝推定に用いる場合の統計量)は，この十分統計量の中から選ばれる。ただし，十分性をもつ推定量は無数に存在するため，より良い推定量としてはさらに一定の条件が必要となる。推定量がさらにもつべき性質は，通常，固定標本特性と漸近標本特性との二つに大別される。漸近標本特性については後述するとして，ここでは固定標本特性について説明しておくことにしたい。固定標本特性とは標本数の大きさを一定にして構成された推定量のもつ性質である。

上記の十分統計量 t を θ の推定量として用いるために $\hat{\theta}$ と書き表すことにする。$\hat{\theta}$ は，標本変量（＝確率変数）の関数であるから，それ自体一定の確率分布に従っている。この時，$E(\hat{\theta})=\theta$ が成り立つならば，$\hat{\theta}$ を θ の不偏推定量とよぶ。ただし，期待値が未知パラメーターに一致したからといって，その推定量の分散が極めて大きいものであるならば，その推定量に基づく推定値 estimate を求めてみても，それが未知パラメータ θ に一致する確率は極めて低いものとなる。そこで，無数に存在する不偏推定量のうち，その分布の分散が最小となるような推定量をもって最良推定量とするのである。

不偏推定量の分散は一般に次の不等式を満たす。

$$V(\hat{\theta}) \geqq \frac{1}{nE\left\{\frac{\partial}{\partial\theta}\log f(x; \theta)\right\}}$$

したがって，不偏推定量の分散が上式の右辺（クラーメル・ラオの境界 Cramér-Rao bound）に一致しているならば，それは最小分散の不偏推定量であるということがわかるのである。仮定〔I〕の下で最小乗推定量 $\hat{\beta}$ が最良不偏推定量になるということは，ガウス・マルコフの定理としてまとめられている。

なお，最小分散の不偏推定量は有効推定量 efficient estimator とよばれることもある。

ところで，最小二乗推定量が最良不偏推定量になるということは，仮定〔I〕(i)～(iv)の下で一般的に得られる結論であるが，通常，この4つの仮定の他に，さら

に仮定〔I〕（v）が設けられる。これは，パラメータの信頼区間の構成，したがってまた，パラメータの有意性検定と関わりをもつものである。すなわち，$\boldsymbol{u}$ が多変量正規分布に従えば，(1)式より $\boldsymbol{y}$ も多変量正規分布に従う。そうすると，さらに，(3)式より，最小二乗推定量 $\hat{\boldsymbol{\beta}}$ も多変量正規分布に従うことになる。他方，(2)，(3)式より，$\hat{\boldsymbol{\beta}}$ の期待値，分散は次式で示されるようになる。

$$\hat{\boldsymbol{\beta}} \sim N(\beta,\ \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$$

また，$\hat{\boldsymbol{u}}=\boldsymbol{y}-\hat{\boldsymbol{y}}=\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}$ とすれば，σ^2 の不偏推定量は次のようになる。

$$\hat{\sigma}^2=\frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-p-1}$$

ただし，この時，$\hat{\boldsymbol{u}}$ は多変量正規分布に従い，その期待値，分散は次式で示されるようになる。

$$\hat{\boldsymbol{u}} \sim N(\boldsymbol{0},\ \sigma^2\boldsymbol{A})$$

　ただし，$\boldsymbol{A}=\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

したがって，これより，$\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}/\sigma^2$ は自由度 $n-p-1$ のカイ二乗分布に従うことがわかる。

$$\frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{\sigma^2}=\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p-1)$$

ここで，$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{A}=\boldsymbol{0}$ となることから，$\hat{\boldsymbol{\beta}}$ と $\hat{\sigma}^2$ が独立であることがわかり，次のような検定統計量を構成することが可能になる。まず個別の係数パラメータ β_i については，

$$\frac{\hat{\beta}_i-\beta_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim t(n-p-1) \tag{4}$$

　ただし，$\hat{\sigma}_{\hat{\beta}_i}$ は $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ の第 i 行対角要素の平方根である。

また，β_0，β_1，……β_p に関しては，

$$\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})/(p+1)}{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}/(n-p-1)} \sim F(p+1,\ n-p-1) \tag{5}$$

こうして，仮定〔I〕（v）を設けることによって，(4)，(5)式に基づいて係数パラメータの信頼区間の構成を行うことが可能となり，したがってまた，その有意性検定を行うことが可能となるのである。

たとえば，$\beta_i{}^*$ を β_i の特定の値として，仮説 hypothesis H_0; $\beta_i=\beta_i{}^*$ を対立仮

説 alternative hypothesis H_1; $\beta_i \neq \beta_i^*$ にたいして検定するには，$t=(\beta_i-\beta_i^*)/\hat{\sigma}_{\hat{\beta}_i}$ が自由度 $n-p-1$ の t 分布に従うという性質が利用される。すなわち，t_α を t の $\alpha\times100\%$水準として，$|t|\leqq t_\alpha$ となれば仮説 H_0 は有意水準 significance level α により棄却されることになり，また，$|t|<t_\alpha$ となれば対立仮説 H_1 が有意水準 α により棄却されることになるわけである。ただし，有意性検定の考え方においては，証明しようとする仮説を対立仮説として設定することが通例である。したがって，上記の検定においても，$\beta_i^*=0$ とされ，仮説（帰無仮説 null hypothesis）H_0; $\beta_i=0$ を対立仮説 H_1; $\beta_i\neq 0$ にたいして検定するという試みとして行われることになる。すなわち，H_0; $\beta_i=0$ を一定の有意水準で棄却することによって，H_1; $\beta_i\neq 0$ の採択を図るのである。しかしながら，この場合，$\beta_i\neq 0$ の採択といっても，その実質的な意味は，$\beta_i\neq 0$ が成立しないとはいえない——すなわち，そのパラメータを係数としてもつ変数をモデル内に明示的に取り入れてならない理由はないということ——という極めて消極的な結論を導くに過ぎないという点に注意しなければならない。こうした事情は(5)式の F 検定の場合においても全く同様である。

こうして，仮定〔I〕(i)〜(v)の下で，係数パラメータの有意性検定を行うことにより，当該パラメータを係数としてもつ変数をモデル内に明示的に取り入れてならない理由はないとの判断が下され，先の最小二乗推定量に一定の統計データがあてはめられた最小二乗推定値が，係数パラメータの最良の推定値として採用されていくことになるのである。ただし，この具体的な推定値においては，その適切性の根拠は，パラメータの有意性検定によって説かれるのではけっしてなく，それは，あくまでも，一定の統計データがあてはめられる以前の推定量としての特性——この場合は不偏性および最小分散性——から説かれるものとなっているということに注意しておかねばならない。

さて，以上においては仮定〔I〕(i)〜(v)が成立するものとして議論をすすめてきたが，むろん，これは保証の限りではない。そこで，計量経済学においては，これらの諸仮定のいくつかについても，その有意性検定が試みられることになる。たとえば，仮定〔I〕(iii)の不均一分散の検定であれば，$\sigma_1^2=\sigma_2^2=\cdots\cdots=\sigma^2_n$ にたいする帰無仮説の検定が行われる。また，仮定〔I〕(iv)の系列相関の検定であれば，$u_t=\rho u_{t-1}+\varepsilon_t$ として，$\rho=0$ なる帰無仮説の検定が行なわれる。ただし，こう

した検定において導かれる結論も，上記の t 検定の場合と同様，そうした仮定が成り立たないとする理由はないというものに過ぎないことはいうまでもない。こうして，徹頭徹尾，「そうした仮定が成り立たないとする理由はない」という論法によって，種々の仮定が積み上げられ，変数が定められ，先の最小二乗法によって係数パラメータが推定されていくのである。これが単一方程式モデルの計測方法の基本的な論理構造である。

計量経済学の通常のテキストにおいては，以上概観してきた単一方程式モデルの計測方法についての解説がたいがいその冒頭部分において行われている。しかしながら，実際には，計量経済モデルが単一方程式モデルとして作成されるということは今日では極めて稀であって，ほとんどが複数の関係式を含む連立方程式モデルとして作成されたものとなっている。これは，経済現象の複雑さにたいする考慮から，先の(1)式のみでなく，(1)式の右辺の説明変数のいくつかを被説明変数とする方程式の計測をもあわせて行うようにしなければならないとする見解が支配的になっていることを意味している。そして，計量経済モデルをこのように連立方程式モデルとして作成すべき理由としては，通常，経済諸量が相互依存的関係にあるという点があげられている。

すなわち，先の単一方程式モデルでは，説明変数の被説明変数にたいする一方的依存関係しか表現しえないのにたいして，連立方程式モデルであれば，一定変数を，ある方程式では説明変数の位置に，また，他の方程式では被説明変数の位置におくことが可能となり，したがって，このモデルであれば経済諸量が相互依存的関係にあることを表現することが可能になるというわけである。このような経済諸量の相互依存性をとらえるべく作成される連立方程式モデルは，経済の構造を記述したモデルであるという意味で構造モデル structural model とよばれることが多い。この用語法にしたがえば，このモデルの各方程式は構造方程式，また，その係数パラメータは構造パラメータということになる。

ところで，モデルをこのように連立方程式モデルとして作成してみると，先の仮定〔I〕(i)の成立可能性に疑問が生じ，このモデルの計測に前述の最小二乗法を適用しうるのかどうかということが問題になってくる。そこで，次に，連立方程式モデルのパラメータ推定問題について見てみることにしたい。

連立方程式モデルは一般的には次のように設定される。

$$\boldsymbol{y}_t'\boldsymbol{B}+\boldsymbol{X}_t'\boldsymbol{\Gamma}=\boldsymbol{u}_t' \quad (t=1,\ 2,\ \cdots\cdots,\ n) \tag{6}$$

ただし，$\boldsymbol{y}_t'=(y_{1t},\ y_{2t},\ \cdots\cdots,\ y_{Gt})$

$$\boldsymbol{X}_t'=(x_{1t},\ x_{2t},\ \cdots\cdots,\ x_{Kt})$$

$$\boldsymbol{B}=\begin{pmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{G1} \\ \beta_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \beta_{1G} & \cdots & \cdots & \beta_{GG} \end{pmatrix}$$

$$\boldsymbol{\Gamma}=\begin{pmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{G1} \\ \gamma_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma_{1K} & \cdots & \cdots & \gamma_{GK} \end{pmatrix}$$

$$\boldsymbol{u}_t'=(u_{1t},\ u_{2t},\ \cdots\cdots\cdots,\ u_{Gt})$$

これは，ベクトル $\boldsymbol{y}_t$ で表わされる G 個の変数 $y_{it}(i=1,\ 2,\ \cdots\cdots,\ G)$ に関する G 本の連立方程式を意味している。すなわち，このモデルにおいては，G 個の経済変数 y_{it} の各々のデータ点を G 本の連立方程式の交点として同時的に説明することが試みられるわけである。その意味で，このモデルは同時連立方程式モデル simultaneous equation model（略して同時方程式モデル）とよばれている。また，この時の経済変数 y_{it} は，その動きがこのモデルによって内生的に決定されるという意味で内生変数 endogenous variable とよばれている。

上記モデルにおいては，この内生変数の他に，その動きがモデル外の要因によって決定されることを仮定した K 個の変数 $x_{it}(i=1,\ 2,\ \cdots\cdots,\ K)$が含まれている。上記の内生変数にたいして，これは外生変数 exogenous variable とよばれている。この x_{it} には，内生変数の数期前の値を表わす変数（＝ラグ付内生変数 lagged endogenous variable）が含まれることがある。この場合には，当期の内生変数のみである y_{it} は同時従属変数 jointly dependent variable，また，外生変数とラグ付内生変数を含む x_{it} は先決変数 predetermind variable とよばれることになる。ここでは，G 個の同時従属変数と K 個の先決変数を含んだモデルが n 期間にわたる統計データによって計測される場合を想定しておくことにする。むろん，それらの統計データが設定された，方程式を完全に満たすということにありえないため，

(1)式と同様，(6)式にも攪乱項が導入され，合計 $G\times n$ 本の確率方程式として(6)式は表わされている。

計量経済学の仕事は，いうまでもなく，(6)式の係数パラメータ $\boldsymbol{B}$ および $\boldsymbol{\Gamma}$ を実際の統計データから推定することであるが，この場合には，通常，次のような仮定が設けられる。

〔Ⅱ〕(ｉ)　攪乱項ベクトル $\boldsymbol{u}_t$ $(t=1,\ 2,\ \cdots\cdots,\ n)$ は，時間 t に関してたがいに独立であり，同一分布に従う。

(ii)　攪乱項ベクトル $\boldsymbol{u}_t$ の期待値および分散共分散行列は次のようになる。

$$E(\boldsymbol{u}_t)=\boldsymbol{0}$$

$$E(\boldsymbol{u}_t'\boldsymbol{u}_t)=\boldsymbol{\Phi}=\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots\cdots & \sigma_{1G} \\ \sigma_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{G1} & \cdots\cdots & \cdots & \sigma_{GG} \end{pmatrix}$$

(iii)　分散共分散行列 $\boldsymbol{\Phi}$ は正値定符号である。

(iv)　攪乱項ベクトル $\boldsymbol{u}_t$ は多変量正規分布に従う。

$$\boldsymbol{u}_t\sim N(\boldsymbol{0},\ \boldsymbol{\Phi})$$

以上の諸仮定の下で，まず，第1番目の構造方程式の係数パラメータの推定を考える。(6)式の第1番目の方程式を次のように表わす。

$$\boldsymbol{y}_t'\boldsymbol{\beta}_1+\boldsymbol{X}_t'\boldsymbol{\gamma}_1=u_{1t}$$

$$\boldsymbol{\beta}_1'=(\beta_{11},\ \beta_{12},\ \cdots\cdots,\ \beta_{1G})\cdots\cdots\boldsymbol{B}\text{ の第1列}$$

$$\boldsymbol{\gamma}_1'=(\gamma_{11},\ \gamma_{12},\ \cdots\cdots,\ \gamma_{1K})\cdots\cdots\boldsymbol{\Gamma}\text{ の第1列}$$

さらに，$\boldsymbol{y}_t'=[y_{1t},\ \boldsymbol{y}_{0t}']$ とし，それに対応して $\boldsymbol{\beta}_1'=[-1,\ \boldsymbol{\beta}_0']$ とする。また，攪乱項の符号を変える。この時，上式は次のようになる。

$$y_{1t}=\boldsymbol{y}_{0t}'\boldsymbol{\beta}_0+\boldsymbol{X}_t'\boldsymbol{\gamma}_1+u_{1t} \qquad (7)$$

ここで，ベクトル $\boldsymbol{u}_t'=(u_{1t},\ u_{2t},\ \cdots\cdots,\ u_{Gt})$ が結合分布に従い，また，$\boldsymbol{u}_t$ の要素 $u_{1t},\ u_{2t},\ \cdots\cdots,\ u_{Gt}$ がそれぞれ $y_{1t},\ y_{2t},\ \cdots\cdots y_{Gt}$ に攪乱効果を与えることから，上記 $\boldsymbol{y}_{0t}$ と u_{1t} とはたがいに非独立となることがわかる。そうすると，先の仮定〔Ⅰ〕(ｉ)が成り立たなくなるため，(7)式の係数パラメータの最小二乗推定量を求

めてみても，それが最小分散の不偏推定量である保証は全くなくなる。

すなわち，まず，(7)式を各 t に関してそれぞれ別記し，それを行列表示してみる。

$$
\begin{aligned}
\boldsymbol{y}_1 &= \boldsymbol{Y}_0\boldsymbol{\beta}_0 + \boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{u}_1 \\
&= [\boldsymbol{Y}_0\ \boldsymbol{X}_1]\begin{bmatrix}\boldsymbol{\beta}_0\\ \boldsymbol{\gamma}_1\end{bmatrix} + \boldsymbol{u}_1
\end{aligned}
\qquad (8)
$$

ただし，$\boldsymbol{y}_1{}' = (y_{11},\ y_{12},\ \cdots\cdots,\ y_{1n})$

$$
\boldsymbol{Y}_0 = \begin{pmatrix} y_{21} & y_{31} & \cdots & y_{G1} \\ y_{22} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ y_{2n} & \cdots & \cdots & y_{Gn} \end{pmatrix}
$$

$$
\boldsymbol{X}_1 = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{K1} \\ x_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{1n} & \cdots & \cdots & x_{Kn} \end{pmatrix}
$$

$$
\boldsymbol{u}_1{}' = (u_{11},\ u_{12},\ \cdots\cdots,\ u_{1n})
$$

この時，(3)式より，$\boldsymbol{\beta}_0$，$\boldsymbol{\gamma}$ にたいする最小二乗推定量は次のようになる。

$$
\begin{bmatrix}\hat{\boldsymbol{\beta}}_0\\ \hat{\boldsymbol{\gamma}}_1\end{bmatrix} = \begin{bmatrix}\boldsymbol{Y}_0{}'\boldsymbol{Y}_0 & \boldsymbol{Y}_0{}'\boldsymbol{X}_1\\ \boldsymbol{X}_1{}'\boldsymbol{Y}_0 & \boldsymbol{X}_1{}'\boldsymbol{X}_1\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{Y}_0{}'\\ \boldsymbol{X}_1{}'\end{bmatrix}\boldsymbol{y}_1
$$

この式に(8)式を代入すると次のようになる。

$$
\begin{bmatrix}\hat{\boldsymbol{\beta}}_0\\ \hat{\boldsymbol{\gamma}}_1\end{bmatrix} = \begin{bmatrix}\boldsymbol{\beta}_0\\ \boldsymbol{\gamma}_1\end{bmatrix} + \begin{bmatrix}\boldsymbol{Y}_0{}'\boldsymbol{Y}_0 & \boldsymbol{Y}_0{}'\boldsymbol{X}_1\\ \boldsymbol{X}_1{}'\boldsymbol{Y}_0 & \boldsymbol{X}_1{}'\boldsymbol{X}_1\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{Y}_0{}'\\ \boldsymbol{X}_1{}'\end{bmatrix}\boldsymbol{u}_1 \qquad (9)
$$

すでに述べたように，$\boldsymbol{u}_1$ と $\boldsymbol{Y}_0$ とは非独立であるから，(9)式の右辺第２項は一般に $\boldsymbol{0}$ とはならない。それゆえ，(6)式の第１番目の方程式に最小二乗法を適用することによって得られる最小二乗推定量 $\hat{\boldsymbol{\beta}}_0$，$\hat{\boldsymbol{\gamma}}_1$ は，一般に不偏推定量とはならないということがわかる。こうして，同時方程式モデルのパラメータの場合には，先の単一方程式のモデルのパラメータの場合と同一の推定方法を適用することは不可能となり，それに独自の推定方法が必要となってくるのである。

同時方程式モデルの最も一般的な計測方法として採用されているのは，最大尤度法 maximum likelihood とよばれる方法である。これは，次のような考えに基づく方法である。

x_1, x_2, ……, x_n を m 個の未知パラメータ θ_1, θ_2, ……, θ_m をもつ密度関数 $f(x;\ \theta_1,\ \theta_2,\ \cdots\cdots,\ \theta_m)$ からの無作為標本とする。この時，x_1, x_2, ……, x_n の結合密度関数は次のようになる。

$$f(x_1;\ \theta_1,\ \theta_2,\ \cdots\cdots\theta_m)\cdot f(x_2;\ \theta_1,\ \theta_2,\ \cdots\cdots,\ \theta_m)$$
$$\cdot\cdots\cdots\cdot f(x_n;\ \theta_1,\ \theta_2,\ \cdots\cdots,\ \theta_m) \qquad (10)$$

標本x_1, x_2, ……, x_n が特定の値をとっている状態で，(10)式を未知パラメータ θ_1, θ_2, ……, θ_m の関数とみなす時，(10)式を θ_1, θ_2, ……, θ_m に関する尤度関数 likelihood function とよぶ。また，尤度関数の値を尤度 likelihood とよぶ。そうすると尤度関数を $L(\theta_1,\ \theta_2,\ \cdots\cdots,\ \theta_m)$ で表わせば，尤度 L は次のようになる。

$$L=L(\theta_1,\ \theta_2,\ \cdots\cdots\theta_m)$$
$$=f(x_1;\ \theta_1,\ \cdots\cdots,\ \theta_m)\cdot(f(x_2;\ \theta_1,\ \cdots\cdots\theta_m)\cdot\cdots\cdots\cdot f(x_n;\ \theta_1;\cdots\cdots\theta_m)$$

最大尤度法とは，この尤度 L を最大にするような θ_1, θ_2, ……, θ_m をもって，未知パラメータ θ_1, θ_2, ……θ_m の推定値とする方法である。この推定値が最尤推定値とよばれるもので，通常，最小二乗推定値と区別して，$\tilde{\theta}_1$, $\tilde{\theta}_2$, ……, $\tilde{\theta}_m$ と表わされる。この $\tilde{\theta}_1$, $\tilde{\theta}_2$, ……, $\tilde{\theta}_m$ が標本変量（確率変数）x_1, x_2, ……x_n の関数の形で与えられる時は，これらは最尤推定量とよばれることになる。

ところで，尤度を最大にする値は次の連立方程式を解くことによって求められる。

$$\frac{\partial L}{\partial \theta_i}=0 \quad \text{または} \quad \frac{\partial \log L}{\partial \theta_i}=0 \quad (i=1,\ 2,\ \cdots\cdots,\ m)$$

そこで，こうした考え方を同時方程式モデルの場合に適用してみることにしよう。仮定〔Ⅱ〕(ⅰ)より，$\boldsymbol{u}_t$ が系列的に独立であるから，$\boldsymbol{X}_t$ が与えられた時の $\boldsymbol{y}_t$ の条件付尤度関数は次のようになる。

$$\boldsymbol{L}=\boldsymbol{P}(\boldsymbol{y}_1,\ \boldsymbol{y}_2,\ \cdots\cdots\boldsymbol{y}_n|\boldsymbol{X}_1,\ \boldsymbol{X}_2,\ \cdots\cdots,\ \boldsymbol{X}_n)$$
$$=|det\ \boldsymbol{B}|^n P(\boldsymbol{u}_1)\cdot P(\boldsymbol{u}_2)\cdot\cdots\cdots\cdot P(\boldsymbol{u}_n)$$

さらに，仮定〔Ⅱ〕(iv)より，$\boldsymbol{u}_t$ が多変量正規分布に従うから，尤度 $\boldsymbol{L}$ の対数$\boldsymbol{L}^*$は次のようになる。

$$\boldsymbol{L}^*=log\ \boldsymbol{L}$$

$$=k+n\,log|det\,\boldsymbol{B}|-\frac{n}{2}log\,det\,\boldsymbol{\Phi}\ -\frac{1}{2}\sum_{t=1}^{n}\boldsymbol{u}_t'\boldsymbol{\Phi}\boldsymbol{u}_t \qquad (11)$$

この $\boldsymbol{L}^*$ を最大にする β_{ij} を求めるには，$\partial\boldsymbol{L}^*/\partial\beta_{ij}=0$ を解けばよい。これは次のような形の式となる。

$$\frac{n}{det\,\boldsymbol{B}}\cdot\frac{\partial|det\,\boldsymbol{B}|}{\partial\beta_{ij}}-\frac{1}{2}\cdot\frac{\partial S}{\partial\beta_{ij}}=0 \quad (i,\ j=1,\ 2,\ \cdots\cdots,\ n) \quad (12)$$

ただし，$S=\sum_{t=1}^{n}\boldsymbol{u}_t'\boldsymbol{\Phi}\boldsymbol{u}_t$

同時方程式モデルの係数パラメータ $\boldsymbol{B}$ の最尤推定値は，この(12)式を解くことによって求められるのである。

では，こうして得られる最尤推定値の望ましさとは一体何であろうか。最小二乗推定値の場合には，最小二乗推定量が不偏性ならびに最小分散性という固定標本特性をもつところからその推定値の望ましさが説かれるものとなっていた。しかしながら，最尤推定量は一般にこのような固定標本特性はもちえない。そのため，最尤推定値の場合には最小二乗推定値と同様の論理でその望ましさを説くことは不可能である。

だが，最尤推定量は，固定標本特性こそもたないものの，次のような漸近標本特性はもっている。すなわち，密度関数 $f(x;\ \theta)$ からの大きさ n の無作為標本に基づく最尤推定量 $\tilde{\theta}$ は，n を大きくすると漸近的に次の正規分布に従う。

$$\tilde{\theta}\sim N\left(\theta,\ \frac{1}{nE\left\{\frac{\partial}{\partial\theta}log\,f(x;\ \theta)\right\}}\right)$$

ここで，$E(\tilde{\theta})=\theta$ より，最尤推定量 $\tilde{\theta}$ は漸近的不偏推定量であることがわかる。また，その分散はクラーメル・ラオの境界に等しいことから，最尤推定量 $\tilde{\theta}$ は漸近的に最小分散の推定量であることがわかる。さらに，最尤推定量 $\tilde{\theta}$ の分散は n を大きくすれば0に近づくことから，$\tilde{\theta}$ は θ に確率収束することがわかる。このような推定量は一致推定量 consistetency estimator とよばれている*。　一八八

* ε，δ を任意に与えた正の小さな数とする時，もし $n>N$ にたいして，

$$P(|\tilde{\theta}-\theta|<\varepsilon)>1-\delta$$

を満たす整数 N が常に存在するならば，$\tilde{\theta}$ は θ に確率収束するといい，$plim\,\tilde{\theta}=\theta$ と表わす。そして，この時の $\tilde{\theta}$ を θ の一致推定量という。

つまり，一致推定量とは，標本の大きさ n を限りなく大きくすれば未知パラメータに大体一致する推定量であるということである。

また，未知パラメータが複数個の場合には，つまり，先の $\boldsymbol{\theta}=(\theta_1,\ \theta_2,\ \cdots\cdots,\ \theta_m)$ の場合には，$\tilde{\boldsymbol{\theta}}$ は漸近的に次の多変量正規分布に従う。

$$\tilde{\boldsymbol{\theta}}\sim N(\boldsymbol{\theta},\ \boldsymbol{\Phi})$$

ただし，$\boldsymbol{\Phi}$ は $\boldsymbol{\Phi}^{-1}$ の第 i 行 j 列の要素 v_{ij} が次のようになる行列である。

$$v_{ij}=-nE\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}log f(x;\ \theta_1,\ \theta_2,\ \cdots\cdots,\ \theta_m)\right\}$$

それゆえ，この場合にも，上記の性質はそのままあてはまることがわかる。

こうして，先の(12)式を解いて得られる同時方程式モデルの係数パラメータの最尤推定量は，漸近的に最小分散の不偏推定量であり，かつまた，一致推定量であるということがわかるのである。そして，計量経済学においては，最尤推定量がこのような性質をもつことに依拠して，それに一定の統計データがあてはめられた最尤推定値が同時方程式モデルの係数パラメータの最も望ましい推定値であるとみなされていくことになるのである。

先の最小二乗推定値の場合には，標本数を一定とした上で導かれる推定特性に基づいてその望ましさが説かれていたが，この最尤推定値の場合には，標本数を無限に増した時の推定特性（＝漸近特性）に基づいてその望ましさが説かれることになるという点が特徴的である。すなわち，もし標本数を無限に多くとったならば，最推定値は未知パラメータにほぼ一致するはずであるから，こうした特性をもたない推定値よりは望ましいというわけである*。その意味で，計量経済学者が特に注目するのは一致性という推定特性であるという点に注意しなければならない。

* 同時方程式モデルの係数パラメータの最尤推定量が一致性，漸近的不偏性，漸近的最小分散性，漸近的正規性といった漸近特性をもつことについては，次で明らかにされた。

Mann, H.B. and A. Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica*, Vol. 11 No. 1, 1943.

Koopmans, T.C., H. Rubin and R.B. Leipnik, "Measuring the Equation Systems in Dynamic Economics," *Statistical Inference in Dynamic Economic Model*, edited by T.C. Koopmans, 1950.

ところで，先に，個々の構造方程式に最小二乗法を適用したのでは最小分散の不偏推定量が得られないという理由から，構造パラメータの推定に（直接）最小二乗

法を適用することは不可能なのであると説明しておいた。しかしながら，上に見たように，パラメータの推定特性を固定標本特性に限らず漸近標本特性でもよいということにすれば，個々の構造方程式に最小二乗法を適用したのでは，そうした特性は得られないのであろうかということが問題になってくる。そこで，先の(9)式に立ち戻ってみることにしよう。もし，最小二乗推定量が一致推定量であるとすれば，(9)式の右辺第2項の確率極限値はゼロベクルトとなるはずであるが，$\boldsymbol{Y}_0$ と $\boldsymbol{u}_1$ とが非独立である以上，一般的には，

$$plim\left(\begin{bmatrix}\boldsymbol{Y}_0'\boldsymbol{Y}_0 & \boldsymbol{Y}_0'\boldsymbol{X}_1\\ \boldsymbol{X}_1'\boldsymbol{Y}_0 & \boldsymbol{X}_1'\boldsymbol{X}_1\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{Y}_0'\\ \boldsymbol{X}_1'\end{bmatrix}\boldsymbol{u}_1\right)\neq\boldsymbol{0}$$

である。こうして，個々の構造方程式に最小二乗法を適用したのでは，固定標本特性はおろか，漸近標本特性すら得られないということがわかり，同時方程式モデルの場合には，先の最尤法——これは後に完全情報最尤法full information maximum likelihood とよばれることになる——がその一般的な計測方法として位置づけられていくことになるのである。

Ⅱ　推定方法の展開

構造パラメータ β_{ij} の最尤推定値を求めるためには，前節⑿式を解かねばならないが，この式の $det\ \boldsymbol{B}$ および $\boldsymbol{u}_t'\boldsymbol{\Phi}\boldsymbol{u}_t$ は非常に複雑な非線形式となり，モデルの変数が多い場合には，これを解くことは実際上不可能である。そこで，理論的には先の方法が構造パラメータの最も望ましい推定方法であるとしながらも，実際上はより簡便な推定方法に依拠せざるをえなくなるのである。現在考案されているより簡便な推定方法の多くは，次のような方程式を利用するものとなっている。

$$\boldsymbol{y}_t'=-\boldsymbol{x}_t'\boldsymbol{\Gamma}\boldsymbol{B}^{-1}+\boldsymbol{u}_t'\boldsymbol{B}^{-1}\quad(t=1,\ 2,\ \cdots\cdots,\ n)$$

これは，前節(6)式の同時従属変数を先決変数と攪乱項とに関して解くことによって得られたものであり，構造方程式体系 structural equations system から誘導された方程式体系という意味で，誘導型方程式体系 reduced form equations system とよばれている。新しい記号を用いてこれを次のように表わす。

$$\boldsymbol{y}_t'=\boldsymbol{x}_t'\boldsymbol{\Pi}+\boldsymbol{v}_t'\quad(t=1,\ 2,\ \cdots\cdots,\ n)$$

前節の仮定〔Ⅱ〕(iv)の下では，誘導型攪乱項 $\boldsymbol{v}_t$ も多変量正規分布に従うことに

なる。

$$\boldsymbol{v}_t \sim N(\boldsymbol{0},\ \boldsymbol{\Sigma})$$

ただし，$\boldsymbol{\Sigma}=\boldsymbol{B}'^{-1}\boldsymbol{\Phi}\boldsymbol{B}^{-1}$

さて，構造パラメータのより簡便な推定方法としては，まず，誘導型最尤法 reduced form maximum likelihood とよばれる方法を取り上げねばならない。というのは，この方法が以下順次取り上げる各種推定方法の考案にあたっての出発点に位置するものとなっているからである。誘導型最尤法とは，誘導型パラメータの最尤推定値を求めて，それを変数変換することによって構造パラメータの推定値を求めるという方法である*。このような手続きが可能となるのは，最尤推定量の推定特性である一致性が変数変換によってもその性質が保持される不変性 invariantness という性質をもっているからである。すなわち，誘導型パラメータの一致推定値を求め，それを変数変換するならば構造パラメータの一致推定値が得られるというわけである。

* この方法は次で提唱された。
Haavelmo, T., "The Statistical Implication of a System of Simultaneous Equations," *Econometrica*, Vol. 11 No. 1, 1943.

ただし，この方法が適用可能となるためには，あらかじめモデル設定面で一定の制約が必要となる。というのは，この方法で構造パラメータの推定値が一義的に確定——構造パラメータが適度認定 just-identified——されるためには，構造パラメータ$\boldsymbol{B}$，$\boldsymbol{\Gamma}$と誘導型パラメータ$\boldsymbol{\Pi}$との間に一意対応の関係が存在しなければならず，どのような場合でも構造パラメータの推定値が確定可能であるというわけではないからである。前者の個数が多い場合には，構造パラメータの推定値を確定することは不可能——構造パラメータが認定不能 under-identified——であるし，後者の個数の方が多い場合には，構造パラメータの推定値が複数個確定——構造パラメータが過剰認定 over-identified——されてしまう。この両パラメータの個数の間の関係は，同時従属変数の個数と先決変数の個数とに依存して決ってくるものである。そこで，計量経済学においては，両パラメータの個数の間の関係が同時従属変数の個数と先決変数の個数との間の関係に還元され，構造パラメータの推定値の確定条件が認定条件として整理されている。

上述のごとく，一般的には，構造パラメータの個数と誘導型パラメータの個数との間の関係によって，構造パラメータが適度認定であるか，認定不能であるか，あるいは過剰認定であるかが決ってくる。しかし，実際には構造パラメータのいくつかはアプリオリに0とおかれるため，認定不能モデルであっても，方程式によっては構造パラメータの推定値を一義的に確定することが可能なものもでてくることになる。そのため，構造パラメータの認定条件は方程式ごとに吟味する必要が生じてくるのである。

まず，構造方程式体系ならびに誘導型方程式体系を次のように表してみる。

$$[\mathrm{III}]\quad \begin{cases} \beta_{11}y_{1t}+\cdots\cdots+\beta_{1G}y_{Gt}+\gamma_{11}x_{1t}+\cdots\cdots+\gamma_{1K}x_{Kt}=u_{1t} \\ \vdots \\ \beta_{G1}y_{1t}+\cdots\cdots+\beta_{GG}y_{Gt}+\gamma_{G1}x_{1t}+\cdots\cdots+\gamma_{GK}x_{Kt}=u_{Gt} \end{cases}$$

$$[\mathrm{IV}]\quad \begin{cases} y_{1t}=\tilde{\Pi}_{11}x_{1t}+\cdots\cdots+\tilde{\Pi}_{1K}x_{Kt}+v_{1t} \\ \vdots \\ y_{Gt}=\tilde{\Pi}_{G1}x_{1t}+\cdots\cdots+\tilde{\Pi}_{GK}x_{Kt}+v_{Gt} \end{cases}$$

ここでは，〔Ⅲ〕の第1式の認定条件を考えてみる。通常，構造方程式のどれかひとつを単一方程式モデルの被説明変数のごとく扱い，その係数の絶対値を1とする（=基準化 normalize）ため，この場合は $\beta_{11}=1$ としておく。そうすると，構造パラメータの推定値は次の連立方程式を解くことによって求められる。

$$[\mathrm{V}]\quad \begin{cases} \tilde{\Pi}_{11}+\beta_{12}\tilde{\Pi}_{21}+\cdots\cdots+\beta_{1G}\tilde{\Pi}_{G1}=-\gamma_{11} \\ \vdots \\ \tilde{\Pi}_{1K}+\beta_{12}\tilde{\Pi}_{2K}+\cdots\cdots+\beta_{1G}\tilde{\Pi}_{GK}=-\gamma_{1K} \end{cases}$$

連立方程式〔V〕は，未知数の個数が（$K+G-1$）個で，方程式の本数が K 本である。これが一義的に解けるためには，未知数の個数がこれより（$G-1$）個少なくなければならない。そこで次のように，連立方程式〔V〕の未知数 β_{1i} の個数を $G^{\triangle\triangle}$ 個減らし，また，γ_{1j} を K^{**} 個減らすとする。

$$[\mathrm{VI}]\quad \left.\begin{cases} \tilde{\Pi}_{11}+\beta_{12}\tilde{\Pi}_{21}+\cdots\cdots+\beta_{1G^{\triangle}}\tilde{\Pi}_{G^{\triangle}1}+0\cdots\cdots+0=-\gamma_{11} \\ \vdots \\ \tilde{\Pi}_{1K^*}+\beta_{12}\tilde{\Pi}_{2K^*}+\cdots\cdots+\beta_{1G^{\triangle}}\tilde{\Pi}_{G^{\triangle}K^*}+0\cdots\cdots+0=-\gamma_{1K^*} \end{cases}\right\} K^*\text{本}$$

$$
〔\text{VII}〕\left\{
\begin{array}{l}
\tilde{\Pi}_{1K^*+1}+\beta_{12}\tilde{\Pi}_{2K^*+1}+\cdots+\beta_{1G^{\triangle}}\tilde{\Pi}_{G^{\triangle}K^*+1}+0\cdots\cdots+0=0\\
\vdots\\
\tilde{\Pi}_{1K}+\beta_{12}\tilde{\Pi}_{2K}+\cdots+\beta_{1G^{\triangle}}\tilde{\Pi}_{G^{\triangle}K}+0\cdots\cdots+0=0
\end{array}
\right\}K^{**}\text{本}
$$

$$
\underbrace{\qquad\qquad\qquad\qquad}_{G^{\triangle}\text{個}}\qquad\qquad\underbrace{\qquad\quad}_{G^{\triangle\triangle}\text{個}}
$$

ただし，$G=G^{\triangle}+G^{\triangle\triangle}$

$K=K^*+K^{**}$

そうすると，上の連立方程式が一義的に解けるためには，〔VII〕式のK^{**}本の方程式から（$G^{\triangle}-1$）個の未知パラメータ β_{1i} が求められればよいということがわかる。というのは，未知パラメータ γ_{1j} については，そこで求められた β_{1i} の値と〔VI〕式とを用いて求めることが可能となるからである。〔VII〕式が解けるためには，〔VII〕式の方程式の本数が未知パラメータの個数以上であるという条件が必要である。したがって，〔III〕の第1式の係数パラメータの推定値が得られるための必要条件は次の不等式で与えられることになる。

$K^{**}\geqq G^{\triangle}-1$　あるいは，$K^{**}+G^{\triangle\triangle}\geqq G-1$

これは，同時方程式モデル内の1本の構造方程式の係数パラメータが認定可能 identifiable であるための位数条件 order condition とよばれている。ただし，この等号が成立する場合が適度認定のための位数条件であることはいうまでもない。

ところで，この条件は一構造方程式の係数パラメータが認定可能であるためのあくまでも必要条件に過ぎないということに注意しなければならない。というのは，〔VII〕式においてもし一次従属な方程式が存在するとすれば，すべての β_{1i} を一義的に確定するということは不可能となってしまうからである。そのため，認定可能のための必要十分条件としては，〔VII〕式の係数行列のランクが $G^{\triangle}-1$ 以上であるという条件が必要となってくるのである。これは，認定可能のための階数条件 rank condition とよばれている*。

* 前稿(234ページ)でも指摘しておいたように，認定問題とは，元来は，計測結果が経済理論によって導かれた関係式であるのか否か，例えば，それが需要曲線であるのか供給曲線であるのか等を認定する問題であった（cf. Working, E. J., "What Do Statistical 'Demand Curves' Show ?", *The Quarterly Journal of Economics*, Vol. 41, 1927)。ところが認定問題についての上述の定式化（cf. Koopmans T.C., H. Rubin and R.B.

Leipnik, *ibid.*) は，それに比べるとはるかに形式的なものとなっている。これが何故であるかということは，同時方程式モデル方式の特質を考えるにあたって決定的に重要な意味をもっているので留意されたい。この点については後述することにする。

ともかく，このように考えてくると，先の誘導型最尤法が適用可能となるのは，少なくとも，$K^{**}+G^{\triangle\triangle}=G-1$ が成り立つ場合のみであるということがわかる。しかし，そうすると常にこの条件を満たすようにモデルを作成せねばならず，したがって，計測方法という技術的条件によってモデルに含まれる変数の個数が決められることになるため，モデルの客観的妥当性にたいして疑問が生じかねないことになる。そこで，計量経済学においては，適度認定以外の場合でも構造パラメータの推定値が得られるような方法を考案しておかねばならなくなるのである。このような要請に応えるために，先の誘導型最尤法をより一般化し，過剰認定の場合でも適用可能な方法として考案されたのが情報制限最尤法 limited information maximum likelihood とよばれる方法である。そこで，次に，この方法が誘導型最尤法をどのように一般化したのかという点について見てみることにしたい。

上記構造方程式体系〔Ⅲ〕の第1番目の方程式の係数パラメータの推定を考えてみる。ただし，そこにおけるパラメータのベクトル $\boldsymbol{\beta}_1$, $\boldsymbol{\gamma}_1$ のいくつかの要素をアプリオリに0と仮定し，〔Ⅲ〕の第1式を次のように過剰認定の方程式と考える。

$$\beta_{11}y_{1t}+\cdots\cdots+\beta_{1G^{\triangle}}y_{G^{\triangle}t}+\gamma_{11}x_{1t}+\cdots\cdots+\gamma_{1K^{*}}x_{K^{*}t}=u_{1t}$$

ただし，$K^{**}>G^{\triangle}-1$

(13)式より，誘導型攪乱項 $\boldsymbol{v}_t$ も系列的に独立に多変量正規分布に従っているから，上式の $G^{\triangle}$ 個の同時従属変数についての誘導型方程式をとり，それについての尤度関数を構成することが可能となる。この尤度関数は誘導型方程式体系の係数パラメータの行列 $\boldsymbol{\Pi}$ の最初の $G^{\triangle}$ 行からなるパラメータ〔$\boldsymbol{\Pi}_{\triangle *}$ $\boldsymbol{\Pi}_{\triangle **}$〕を用いて表わされる。したがって，この尤度関数を最大とするならば，最尤推定量〔$\tilde{\boldsymbol{\Pi}}_{\triangle *}$ $\tilde{\boldsymbol{\Pi}}_{\triangle **}$〕を求めることが可能となるのである。

この場合，当該構造方程式が認定可能式であるならば，次の方程式を解くことによって構造パラメータの推定量を求めることが可能となる。

$$\begin{cases}\Pi_{\triangle *}\,\boldsymbol{\beta}_{1\triangle}=-\boldsymbol{\gamma}_1\\ \Pi_{\triangle **}\,\boldsymbol{\beta}_{1\triangle}=\boldsymbol{0}\end{cases}\qquad(14)$$

しかしながら，過剰認定（$K^{**}>G^{\vartriangle}-1$）の場合には，未知数の個数より方程式の本数の方が多くなるため，上記第2式からは $\boldsymbol{\beta}_{1\vartriangle}$ の推定量が複数個得られてしまう。

そこで，$G^{\vartriangle}$個の同時従属変数についての尤度関数を $\rho\,(\tilde{\boldsymbol{\Pi}}_{\vartriangle**})=G^{\vartriangle}-1$ という制約条件下で最大にするのである。そうすると，$\tilde{\boldsymbol{\beta}}_{1\vartriangle}$ に含まれる $G^{\vartriangle}$ 個の要素の比を一義的に定めることが可能となる。したがって，それらのうちのひとつを1とおく（＝基準化則）ならば，他の値も自動的に定ってくるのである。

すなわち，まず，構造方程式体系の第1番目の方程式を，基準化則を導入し，前節(8)式のように表わす。

$$\boldsymbol{y}_1=\boldsymbol{Y}_0\boldsymbol{\beta}_0+\boldsymbol{X}_1\boldsymbol{\gamma}_1+\boldsymbol{u}_1 \tag{15}$$

また，この方程式に含まれる同時従属変数に関する誘導型方程式を集め，それを各 t に関してそれぞれ別記し，行列を用いて次のように表わす。

$$\boldsymbol{Y}_1=\boldsymbol{X\Pi}_1+\boldsymbol{V}_1$$

この時，$\boldsymbol{Y}_1$ に基づく対数尤度関数は次のようになる。

$$\boldsymbol{L}_1{}^*=log\ \boldsymbol{L}_1=k-\frac{1}{2}log\ det\ \boldsymbol{\Sigma}_{1\vartriangle}-\frac{1}{2}\sum_{t=1}^{n}\boldsymbol{v}_{1t}{}'\boldsymbol{\Sigma}_{1\vartriangle}\boldsymbol{v}_{1t}$$

ただし，$\boldsymbol{\Sigma}_{1\vartriangle}$；$\boldsymbol{V}_1$ の分散共分散行列

$\boldsymbol{v}_{1t}$；$\boldsymbol{V}_1$ の要素ベクトル

ここで，$\boldsymbol{\lambda}$ をラグランジュ乗数を要素とするベクトルとして，次式の最大化を図る。

$$l=-\frac{1}{2}log\ \boldsymbol{\Sigma}_{1\vartriangle}-\sum_{t=1}^{n}\boldsymbol{v}_{1t}{}'\boldsymbol{\Sigma}_{1\vartriangle}\boldsymbol{v}_{1t}-\boldsymbol{\lambda}\,(\boldsymbol{\pi}^*-\boldsymbol{\Pi}_{\vartriangle**}\boldsymbol{\beta}_0)$$

ただし，$\boldsymbol{\pi}^*$ は $\boldsymbol{\Pi}_{\vartriangle**}$ の第1行である。

上式を最大にするパラメータは次式を解くことによって求られる。

$$\frac{\partial l}{\partial \boldsymbol{v}_{1t}}=\boldsymbol{0},\ \frac{\partial l}{\partial \boldsymbol{\lambda}}=\boldsymbol{0},\ \frac{\partial l}{\partial \boldsymbol{\beta}_0}=\boldsymbol{0},\ \frac{\partial l}{\partial \boldsymbol{\Sigma}_{1\vartriangle}}=\boldsymbol{0}$$

こうして，$\tilde{\boldsymbol{\beta}}_0$ が求められ，(14)第1式より $\tilde{\boldsymbol{\gamma}}_0$ を求めることが可能となるのである*。

* この方法は次で提唱された。

Anderson, T.W. and H. Rubin, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equation," *Annals of Mathematical Statistics*, 1949.

この方法の特徴は，計測しようとする１本の構造方程式に含まれる同時従属変数と体系中のすべての先決変数についての値を知りさえすれば，その構造方程式の係数パラメータが推定可能になるという点にある。逆にいえば，この方法は計測しようとする構造方程式以外の構造方程式に含まれる同時従属変数に関する情報を無視しているという弱点をかかえている。ここに，この方法が情報制限最尤法とよばれるゆえんがある。情報報制限最尤法による推定量は一般的には次のように与えられる。

$$\begin{pmatrix} \tilde{\beta}_0 \\ \tilde{\gamma} \end{pmatrix} = \begin{pmatrix} Y_0{}'Y_0 - \mu\tilde{V}_0\tilde{V}_0 & Y_0{}'X_1 \\ X_1{}'Y_0 & X_1{}'X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_0{}' - \mu\tilde{V}_0 \\ X_1{}' \end{pmatrix} y_1$$

ただし，$\tilde{V}_0$ は Y_0 の誘導型残差 $\tilde{V}_0 = Y_0 - X_0\tilde{\Pi}_0$，また，$\mu$ は行列方程式 $|Y_1{}'M_1Y_1 - \mu Y_1{}'MY_1| = 0$ の最小特性根である。

ここに，$M = I - X(X'X)^{-1}X'$

$$M_1 = I - X_1(X_1{}'X_1)^{-1}X_1{}'$$

情報制限最尤法の概要ならびにその適用結果は以上のごとくであるが，モデル全体が適度認定の場合には，これは先に述べた誘導型最尤法に帰着することになることはいうまでない。その意味で，誘導型最尤法は情報制限最尤法の特殊な形態であるということができるのである。ところで，これまでは，構造パラメータの推定方法を最尤法の適用という面からのみ見てきた。しかしながら，冒頭でもふれたように，最尤推定量の望ましさの根拠を一致性という推定特性に求めるのであれば，必ずしも最尤法の適用ということに推定方法を限定する必要はなくなってくる。最小二乗法であっても，その適用の仕方次第では構造パラメータの一致推定量を得ることは十分可能なのである。そこで，次に，一致推定量を得るために，最小二乗法の適用の仕方がどのように工夫されたのかという点について見てみることにしたい。

前節で，個々の構造方程式に直接最小二乗法を適用したのでは，その推定量は不偏性はおろか一致性すらもたないということを明らかにしておいた。その主たる原因は，当該構造方程式の説明変数に含まれる同時従属変数とその方程式の攪乱項とが非独立となるためであった。では，誘導型方程式に最小二乗法を適用する場合はどうであろうか。

誘導型パラメータの最小二乗推定量は次のようになる。

$$\hat{\boldsymbol{\Pi}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{Y}$$

上式に $\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{\Pi}+\boldsymbol{V}$ を代入すると次のようになる。

$$\hat{\boldsymbol{\Pi}}=\boldsymbol{\Pi}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{V}$$

ここで，$\boldsymbol{X}$ は先決変数であるから，$plim\frac{1}{n}\boldsymbol{X}'\boldsymbol{U}=\boldsymbol{0}$ が成立し，$\hat{\boldsymbol{\Pi}}$ の確率極限は次のようになる。

$$plim\hat{\boldsymbol{\Pi}}=\boldsymbol{\Pi}+\left[plim\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right]^{-1}plim\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{U}\right)\boldsymbol{B}^{-1}$$

$$=\boldsymbol{\Pi}$$

こうして，$\hat{\boldsymbol{\Pi}}$ が一致推定量であるということがわかり，これを変数変換するならば，構造パラメータの一致推定量を得ることが可能になるということがわかるのである。このような手続きで構造パラメータの推定を行う方法は間接最小二乗法 indirect least squares とよばれている。とくに先決変数がすべて指定変数である場合には，この方法による推定量と先の誘導型最尤法による推定量とは一致し，そのようなモデルの場合には，通常，間接最小二乗法が採用されることになるのである。

ところで，この間接最小二乗法が適用可能となるのは，誘導型最尤法と同様，適度認定のモデルの場合のみである。しかしながら，この方法を適用するために適度認定のモデルのみを作成するというのであれば，計測方法という技術的条件によってモデルに含まれる変数の個数が決められることになるため，モデルの客観的妥当性にたいして疑問が生じかねないことになる。この点は誘導型最尤法に関連して先に述べた通りである。そのため，最尤法の適用という見地からは，こうした事態にたいする配慮として，適度認定のモデルのみでなく過剰認定のモデルであっても適用可能となる情報制限最尤法というものが考案されるにいたっていた。

ところが，この方法はモデルの変数選択問題にたいしては全く無関与である。そこでは，過剰認定のモデルをあくまでも所与として，すなわち，そのようなモデルの客観的妥当性をあくまでも前提として，計測材料である統計データが実現値として最大確率をもつようにパラメータの推定値を定めるということが試みられていたに過ぎない。では，仮に情報制限最尤法を用いるとしても，変数選択が適切であるようなモデルは一体いかにしたら作成しうるのであろうか。

「理論に基づく計測」という計量経済学の自己規定からすれば計量経済モデル作成のためのさし当りの手掛りとなるのは，一定の経済理論によって導かれた経済関係式である。その代表的なものとして，需要曲線，供給曲線，消費関数，投資関数，生産関数等があるが，これらは通常の計量経済モデルに比べて説明変数が少数個であることが特徴的である。これは，経済理論においては，種々の要因を一定とした上でいくつかの経済諸量間の関係を説明するという方法がとられていることと関係している。

このように，計測対象としての経済関係式はあくまでも同質的な条件を前提して導かれた関数式である。これにたいして計測材料である統計データ＝時系データは，異なる時点における調査結果であり，異なる条件のもとでの生起結果であるという意味で非同質的な経済資料である。したがって，このままでは両者を結びつけることは不可能である。そこで，計量経済学においては，経済理論で一定とされた種々の要因を追加的説明変数として陽表的にモデルに組み入れ，経済理論によって導びかれた経済関係式よりも多くの変数を含んだ関係式としてモデルが設定されることになるのである。こうして，この追加的説明変数の選択が，モデルの計測に先だつ，計量経済学に独自の仕事となってくるのである。

単一方程式モデルとしての多元回帰モデルの場合に変数選択がどのように行われるかということは前節で見た通りである。では，同時方程式モデルの場合には，各構造方程式の追加説明変数は一体どのようにして選択されることになるのであろうか。上述の通り，最尤法はこの要請に応えることはできない。そのため，実は，実際にモデルを作成する個々の計量経済学者は，同時方程式モデルの追加説明変数――これは先決変数ばかりでなく同時従属変数である場合もある――の選択基準として，単一方程式モデルとしての多元回帰モデルの説明変数の選択基準を援用することを余儀なくされてきているのである。というのは，追加的説明変数が経済理論によって導き出されない以上，その選択は経験的基準に委ねる以外なく，そのようなものとしては，単一方程式モデルとしての多元回帰モデルの説明変数の選択基準以外に依拠すべき実際的方法が存在しないからである。

多元回帰モデルの説明変数の選択基準とは，前節で述べた，パラメータの t 検定および F 検定に他ならない。個々の計量経済学者は，あらかじめこのような方法で

構造方程式の追加的説明変数の選択を行うことによって，モデルの変数選択が基本的には適切であると暗黙裏に前提するのである。そして，その上で，先の情報制限最尤法などの適用を図るのである。しかしながら，モデルの変数選択を実際にこのような方法で行うとすれば，他面，パラメータの推定方法もそれに見合ったものとしなければならなくなるというのもまた事実である。前節で述べた t 検定量ならびに F 検定量は，最小二乗原理に基づく基準であることが特徴的である。したがって，こうした基準で同時方程式モデルの変数選択を行うとすれば，個々の構造方程式を直接最小二乗法で計測することが必要となるはずである*。しかしながら，前節で述べたように，モデルが同時方程式モデルであることをつらぬけば，このような試みは不可能となってしまう。

* この側面を強調し，構造パラメータの推定に前節で見た直接最小二乗法を適用すべきであるとする意見も少なからず存在している。ただし，この場合には，推定精度面での基準の充足性を後方に退け，モデルによる事後的予測成績の良さといった別の基準からこのような主張を行わねばならなくなることが特徴的である。

cf. Waugh, F.V., "The Place of Least Squares in Econometrics," *Econometrica*, Vol. 29 No. 3, 1961.

こうして，計量経済学は，モデルが同時方程式モデルであることをつらぬけば追加的説明変数の選択基準を設けることができなくなり，また，追加的説明変数の選択基準として上記の基準を採用するならば同時方程式モデルという枠組を崩さねばならなくなるという二律背反に悩まされることになるのである。このような二律背反を克服するためには，同時方程式モデルという枠組を崩さずに個々の構造方程式を実質的に最小二乗法で計測するという方法がぜひとも必要となってくる。このような要請に応えるために考案されたのが二段階最小二乗法 two-stage least squares とよばれる方法である。

この方法は，一般化最小二乗法 generalized least squares を同時方程式モデルのパラメータ推定に応用したものに他ならない。そこで，二段階最小二乗法の概要に入る前に，その準備として，一般化最小二乗法について見ておくことにしたい。前節で，多元回帰モデルの係数パラメータの最小二乗推定量 $\hat{\beta}$ は，最小分散の不偏推定量であるということを，また，本節では，それが一致推定量であるということを明らかにしておいた。ただし，これはあくまでも，前節の仮定〔I〕がすべて成

立する限りにおいてそうなのであって，そのうちの一つでも不成立であるならば，最小二乗推定量$\hat{\boldsymbol{\beta}}$がそのような特性をもつということはなくなる。しかし，仮定〔I〕(iii)の分散一定の仮定については例外であって，仮にそれが不成立でも，分散共分散行列が既知でありさえすれば，上記の特性をもつ最小二乗推定量$\hat{\boldsymbol{\beta}}$を求めることは可能となるのである。

すなわち，前節(1)式で表わされたモデルの分散共分散行列を$\boldsymbol{\Omega}$とすると，$\boldsymbol{\Omega}$は正値定符号行列であるから，$\boldsymbol{P}'\boldsymbol{\Omega}\boldsymbol{P}=\boldsymbol{I}$ かつ $\boldsymbol{P}'\boldsymbol{P}=\boldsymbol{\Omega}^{-1}$ となるようなn次の非特異行列$\boldsymbol{P}$が存在する。そこで，(1)式の左からこの$\boldsymbol{P}$をかけ，次のように書き換える。

$$\boldsymbol{y}^*=\boldsymbol{X}^*\boldsymbol{\beta}+\boldsymbol{u}^* \qquad (16)$$

ただし，$\boldsymbol{y}^*=\boldsymbol{P}\boldsymbol{y}$

$$\boldsymbol{X}^*=\boldsymbol{P}\boldsymbol{X}$$

$$\boldsymbol{u}^*=\boldsymbol{P}\boldsymbol{u}$$

この時，$\boldsymbol{u}^*$の期待値ならびに分散共分散行列は次のようになる。

$$E(\boldsymbol{u}^*)=\boldsymbol{P}E(\boldsymbol{u})=\boldsymbol{0}$$

$$E(\boldsymbol{u}^{*\prime}\boldsymbol{u}^*)=E(\boldsymbol{P}'\boldsymbol{u}'\boldsymbol{u}\boldsymbol{P})=\boldsymbol{P}'E(\boldsymbol{u}'\boldsymbol{u})\boldsymbol{P}$$

$$=\boldsymbol{P}'\sigma^2\boldsymbol{I}\boldsymbol{P}=\sigma^2\boldsymbol{I}$$

したがって，(16)式は前節仮定〔I〕の全条件を満たすことになるため，(16)式の係数パラメータ$\boldsymbol{\beta}$の最小二乗推定量$\hat{\boldsymbol{\beta}}$は，最小分散の不偏推定量であり，かつまた一致推定量になるということがわかる。

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{y}^*$$

$$=(\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{y}$$

$$=(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$$

このようにして適用される最小二乗法が，一般化最小二乗法*とよばれているものである。では，二段階最小二乗法はこの方法をどのように応用したのであろうか。 一七六

* Aitken, A.C., *On Least Squares and Linear Combination of Observation,* 1935.

情報制限最尤法と同様に，先の(15)式のパラメータの推定を考えてみる。(15)式に前節で見た最小二乗法を直接適用することができないのは，$\boldsymbol{Y}_0$と$\boldsymbol{u}_1$とが非独立となるためであった。そこで，$\boldsymbol{Y}_0$に関するすべての誘導型方程式を集め，それに最小二

乗法を適用することによって得られる誘導型パラメータの一致推定量に基づく $\boldsymbol{Y}_0$ の推定量 $\hat{\boldsymbol{Y}}_0$ を求める。そうすると，$\hat{\boldsymbol{Y}}_0$ と $\boldsymbol{u}_1$ とは独立となるから，$\boldsymbol{Y}_0$ を $\hat{\boldsymbol{Y}}_0$ におき換えて，(15)式に直接最小二乗法を適用すると，係数パラメータの一致推定量が得られるはずである。このような手順で，過剰認定式の構造パラメータの一致推定量を得ようとするのが，二段階最小二乗法の基本的方針に他ならない。

すなわち，$\boldsymbol{Y}_0=\boldsymbol{X\Pi}_0+\boldsymbol{V}_0$ とし，このパラメータ $\boldsymbol{\Pi}_0$ の最小二乗推定量を $\hat{\boldsymbol{\Pi}}_0$ とする。そうすると，$\hat{\boldsymbol{\Pi}}_0$，$\hat{\boldsymbol{Y}}_0$ は次のようになる。

$$\hat{\boldsymbol{\Pi}}_0=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Y}_0$$

$$\hat{\boldsymbol{Y}}_0=\boldsymbol{X}_0\hat{\boldsymbol{\Pi}}_0=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Y}_0$$

これを(15)式に代入すると次のようになる。

$$\boldsymbol{y}_1=\hat{\boldsymbol{Y}}_0\boldsymbol{\beta}_0+\boldsymbol{X}_1\boldsymbol{\gamma}_1+\boldsymbol{u}_1 \tag{17}$$

ここで，$\hat{\boldsymbol{Y}}_0=\boldsymbol{Y}_0-\hat{\boldsymbol{V}}_0$ であるから，上式は次のようになる。

$$\begin{aligned}\boldsymbol{y}_1&=(\boldsymbol{Y}_0-\hat{\boldsymbol{V}}_0)\boldsymbol{\beta}_0+\boldsymbol{X}_1\boldsymbol{\gamma}_1+(\boldsymbol{u}_1-\hat{\boldsymbol{V}}_0\boldsymbol{\beta}_0)\\&=[\boldsymbol{Y}_0-\hat{\boldsymbol{V}}_0\ \ \boldsymbol{X}_1]\begin{bmatrix}\boldsymbol{\beta}_0\\\boldsymbol{\gamma}_1\end{bmatrix}+(\boldsymbol{u}_1-\hat{\boldsymbol{V}}_0\boldsymbol{\beta}_0)\end{aligned} \tag{18}$$

(17)式の $\boldsymbol{Y}_0$ と $\boldsymbol{u}_1$ とは独立となるから，(18)式においても，説明変数と攪乱項とは独立となる。そうすると，(17)式に直接最小二乗法を適用しうる可能性が生まれてくるわけだが，その攪乱項については，分散共分散行列が $\sigma^2\boldsymbol{I}$ となる保証は全くない。そこで，この分散共分散行列については既知であると想定し，また，先の $\boldsymbol{P}$ が満たすべき条件を $\boldsymbol{Y}_0-\hat{\boldsymbol{V}}_0$ が満たしていると想定し，(18)式に一般化最小二乗法を適用するのである。そうすると，次のような推定量が得られる。

$$\begin{bmatrix}\hat{\boldsymbol{\beta}}_0\\\hat{\boldsymbol{\gamma}}_1\end{bmatrix}=(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{y}_1$$

ただし，$\boldsymbol{A}=\boldsymbol{PX}_0=(\boldsymbol{Y}_0-\hat{\boldsymbol{V}}_0)\ \boldsymbol{X}_0$

これを書き換えれば次のようになる。

$$\begin{bmatrix}\hat{\boldsymbol{\beta}}_0\\\hat{\boldsymbol{\gamma}}_1\end{bmatrix}=\begin{bmatrix}\boldsymbol{Y}_0'\boldsymbol{Y}_0-\hat{\boldsymbol{V}}_0'\hat{\boldsymbol{V}}_0 & \boldsymbol{Y}_0'\boldsymbol{X}_1\\\boldsymbol{X}_1'\boldsymbol{Y}_0 & \boldsymbol{X}_1'\boldsymbol{X}_1\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{Y}_0'-\hat{\boldsymbol{V}}_0'\\\boldsymbol{X}_1'\end{bmatrix}\boldsymbol{y}_1$$

これが二段階最小二乗法によって求められる構造パラメータの推定量である*。

* Theil, H., Economic Forecasting and Policy, 1958. 〈邦訳〉『経済の予測と政策』

岡本哲治訳，創文社，昭和39年。

むろん，上記 $\boldsymbol{Y}_0-\hat{\boldsymbol{V}}_0$ が先の $\boldsymbol{P}$ が満たすべき条件を完全に満たすかどうかは不明である。しかしながら，構造方程式の変数選択基準として多元回帰モデルの変数選択基準を採用していることを強く念頭におく場合には，構造パラメータの推定値を出来るだけ多元回帰モデルの推定方法の適用結果に近づけておくということが至上命令となり，一致推定量が得られる以上，このような方法も構造パラメータの推定方法としては有効であると見なされることになるのである。

この方法は，(15)式の $\boldsymbol{Y}_0$ と $\boldsymbol{u}_1$ とが独立となるように $\boldsymbol{Y}_0$ を $\hat{\boldsymbol{Y}}_0$ におき換えていることが特徴的である。その意味では，形式的には，この方法は，独立性の条件を満たすために，$\boldsymbol{Y}_0$ 部分に他の変数を導入していると見ることができる。このような，別の変数を導入することによって推定条件を満たそうとする方法は操作変数法 instrumental varable method とよばれている。したがって，この範疇からすれば，二段階最小二乗法は操作変数法の一種であるということになる。

さて，以上が二段階最小二乗法であるが，この方法も，情報制限最尤法と同様，計測しようとする1本の構造方程式に含まれる同時従属変数と体系中のすべての先決変数についての値を知りさえすれば，その構造方程式の係数パラメータが推定可能になるという点にその特徴がある。逆にいえば，この方法も，計測しようとする構造方程式以外の構造方程式に含まれる同時従属変数についての情報を無視しているという弱点をかかえている。そこで，最小二乗法の適用という観点からすれば，全情報最尤法に匹敵する，全変数を考慮に入れた計測方法を考案し，その簡便法として上記の二段階最小二乗法を位置付けておくということが必要となってくるのである。このような要請に応えるために考案されているのが三段階最小二乗法 three-stage least squares とよばれている方法である。そこで，最後に，この方法の概要を見ておくことにしたい。

先の(15)式の計測を考えてみる。まず，(15)式を次のように書き改める。

$$\boldsymbol{y}_1=\boldsymbol{Z}_1\boldsymbol{\delta}_1+\boldsymbol{u}_1$$

ただし，$\boldsymbol{Z}_1=[\boldsymbol{Y}_0 \quad \boldsymbol{X}_1]$

$$\boldsymbol{\delta}_1'=[\boldsymbol{\beta}_0' \quad \boldsymbol{\gamma}']$$

次に，上式に左から $\boldsymbol{X}$ を掛ける。

$$\boldsymbol{X}'\boldsymbol{y}_1=\boldsymbol{X}'\boldsymbol{Z}_1\boldsymbol{\delta}_1+\boldsymbol{X}'\boldsymbol{u}_1 \qquad (19)$$

この式の攪乱項の分散共分散行列は次のようになる。

$$V(\boldsymbol{X}'\boldsymbol{u}_1)=E(\boldsymbol{X}'\boldsymbol{u}_1\boldsymbol{u}_1'\boldsymbol{X})=\sigma_{11}(\boldsymbol{X}'\boldsymbol{X})$$

これは，前節仮定〔I〕(iii)が成り立たないことを意味している。そこで，(19)式の係数パラメータの一致推定量を得るためには，通常の最小二乗法ではなく，上に見た一般化最小二乗法を適用しなければならなくなる。(19)式に一般化最小二乗法を適用してみると，次のような推定量が得られる。

$$\hat{\boldsymbol{\delta}}_1=[\boldsymbol{Z}_1'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{Z}_1]^{-1}\boldsymbol{Z}_1'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}_1$$

これは実は二段階最小二乗法による推定量に他ならない。この時，σ_{11} にたいする一致推定量は次のように定められる*。

$$s_{11}=n^{-1}(\boldsymbol{y}_1-\boldsymbol{Z}_1\hat{\boldsymbol{\delta}}_1)'(\boldsymbol{y}_1-\boldsymbol{Z}_1\hat{\boldsymbol{\delta}}_1)$$

*　これより，十分大きな $\boldsymbol{n}$ にたいして，$\hat{\boldsymbol{\delta}}_1$ の漸近的共分散の推定値として次式を用いることが可能となる。

$$s_{11}[\boldsymbol{Z}_1'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}_1]^{-1}$$

そうすると，この行列の要素 a_{ii} を用いることによって，$\boldsymbol{\delta}_1$ の第 $\boldsymbol{i}$ 要素 δ_{1i} に関する有意性検定が可能となる。すなわち，帰無仮説H_0; $\delta_{1i}=0$，対立仮説H_1; $\delta_{1i}\fallingdotseq 0$ として，$\boldsymbol{H_0}$ が真なる時，$\delta_{1i}/\sqrt{a_{ii}}$ が $N(0,\ 1)$ に従うという性質を利用するのである。

ただし，この検定はあくまでも標本数が十分大きくとれるということを前提とした検定であって，計量経済モデルの計測の場合のように利用しうる統計データが比較的少数個である場合には，仮にこの検定を行うにしても，それは推測統計学上の条件を無視した検定とならざるをえないのである。しかも，この検定は計算が非常に複雑である。そのため，実際には，やはり先に述べた単一方程式モデルとしての多元回帰モデルの説明変数の選択基準が採用されることになるのである。

三段階最小二乗法の基本的な方針は，モデルに含まれる全構造方程式を(19)式のような形に改め，それらの方程式全体に一般化最小二乗法を適用することによって，全ての未知パラメータの一致推定量を同時に求めようとするものである。すなわち，次式に一般化最小二乗法を適用するのである。

$$\begin{pmatrix}\boldsymbol{X}'\boldsymbol{y}_1\\ \boldsymbol{X}'\boldsymbol{y}_2\\ \vdots\\ \boldsymbol{X}'\boldsymbol{y}_G\end{pmatrix}=\begin{pmatrix}\boldsymbol{X}'\boldsymbol{Z}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{X}'\boldsymbol{Z}_2 & & \vdots\\ \vdots & & \ddots & \\ \boldsymbol{0} & \cdots & & \boldsymbol{X}'\boldsymbol{Z}_G\end{pmatrix}\begin{pmatrix}\boldsymbol{\delta}_1\\ \boldsymbol{\delta}_2\\ \vdots\\ \boldsymbol{\delta}_G\end{pmatrix}+\begin{pmatrix}\boldsymbol{X}'\boldsymbol{u}_1\\ \boldsymbol{X}'\boldsymbol{u}_2\\ \vdots\\ \boldsymbol{X}'\boldsymbol{u}_G\end{pmatrix}$$

上式を新しい記号を用いて次のように表わす。

$$\boldsymbol{A}=\boldsymbol{B\delta}+\boldsymbol{V} \tag{20}$$

(20)式の攪乱項の分散共分散行列は次のようになる。

$$var\begin{pmatrix} \boldsymbol{X}'\boldsymbol{u}_1 \\ \boldsymbol{X}'\boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{X}'\boldsymbol{u}_G \end{pmatrix}=\begin{pmatrix} \sigma_{11}\boldsymbol{X}'\boldsymbol{X} & \sigma_{12}\boldsymbol{X}'\boldsymbol{X} & \cdots & \sigma_{1G}\boldsymbol{X}'\boldsymbol{X} \\ \sigma_{21}\boldsymbol{X}'\boldsymbol{X} & \sigma_{22}\boldsymbol{X}'\boldsymbol{X} & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{G1}\boldsymbol{X}'\boldsymbol{X} & \sigma_{G2}\boldsymbol{X}'\boldsymbol{X} & \cdots & \sigma_{GG}\boldsymbol{X}'\boldsymbol{X} \end{pmatrix}$$

ただし，$\sigma_{ij}=E(\boldsymbol{u}_i\boldsymbol{u}_j)=\sigma_{ij}\boldsymbol{I}$

そこで，(20)式に一般化最小二乗法を適用すると，$\boldsymbol{\delta}$ の推定量は次のようになる。

$$\hat{\boldsymbol{\delta}}=(\boldsymbol{B}'\boldsymbol{V}^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{V}^{-1}\boldsymbol{A} \tag{21}$$

ただし，(21)式における $\boldsymbol{V}$ は未知の共分散を含んでいる。そこで，これらを二段階最小二乗法による共分散の推定量 s_{ij} でおきかえることによって，(21)式を実際に推定量として操作可能にするのである。このおきかえがなされた(21)式が三段階最小二乗法による推定量に他ならない*。こうした方法であれば，その推定にあたって，モデルに含まれる全変数に関する情報を考慮に入れることが可能となり，その意味で，この方法は理論的には前節で述べた完全情報最尤法に匹敵する極めて重要な推定方法であると位置づけられることになるのである。

* Theil, H. and A. Zellner, "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," *Econometrica*, Vol. 30 No. 1, 1962.

以上，構造パラメータの各種推定方法について概観してきた。そこで，次に，節を改めて，これら各種方法のうちいずれが最適であるのかを判断するために，計量経済学がどのような研究を積み上げてきたのかという点を見てみることにしたい。