上級計量経済学2011年度

1　回帰分析の復習

1.1　回帰モデルとOLS推定量

　あるスカラー経済変数 y が k 次元ベクトル変数 $x=(x_1,\cdots,x_k)$ に依存して決まっており、

$$E(y|x)=\beta' x=\beta_1+\beta_2 x_2+\cdots+\beta_k x_k$$

という関係にあるとする。$\beta=(\beta_1,\cdots,\beta_k)$ は未知パラメータであり、回帰分析の主たる目的はその推定とそれに関する仮説の検定である。ここで、$x_1=1$ としており、β_1 は定数項である。$\epsilon\equiv y-\beta' x$ とおくと、

$$y=\beta' x+\epsilon$$

と書け、これを線形回帰モデルという。条件付き期待値 $E(y|x)$ は x の関数であり、それを回帰関数というが、回帰関数が x について線形であるために、線形回帰モデルと呼ぶ。y を被説明変数（従属変数）、x を説明変数（独立変数）、ϵ を誤差項（攪乱項）という。定義より、

$$E(\epsilon|x)=0$$

である。更に、

$$E(\epsilon^2|x)=\sigma^2$$

であると仮定する。この関係から無作為標本 (y_i,x_i),$i=1,\cdots,n$ が得られたとする。つまり、

$$y_i=\beta' x_i+\epsilon_i,\ i=1,\cdots,n$$

ただし $x_i=(x_{1i},\cdots,x_{ki})'$ である。$Y=(y_1,\cdots,y_n)'$、$X=(x_1,\cdots x_k)'$ とおくと、β の最小二乗（OLS）推定量 b は

$$b=argmin_\beta Q(\beta),\ \ Q(\beta)=\sum_{i=1}^n(y_i-\beta' x_i)^2=(Y-X\beta)'(Y-X\beta)$$

により定義される。その解は明示的に得られ、

$$b=(X'X)^{-1}X'Y=(\sum_{i=1}^n x_i x_i')^{-1}\sum_{i=1}^n x_i y_i$$

となる。残差を $e=(e_1,\cdots,e_n)'=Y-Xb$ とおくと、σ^2 は

$$s^2=\frac{1}{n-k}\sum_{i=1}^n(y_i-b' x_i)^2=\frac{1}{n-k}\sum_{i=1}^n e_i^2=\frac{e'e}{n-k}$$

により推定される。また、回帰のフィット（当てはまり）の良さを調べるには、決定係数

$$R^2=1-\frac{e'e}{\sum(y_i-\overline{y})^2}$$

と自由度修正済み決定係数

$$\overline{R}^2 = 1 - \frac{e'e/(n-k)}{\sum(y_i-\overline{y})^2/(n-1)}$$

が用いられる。

1.2 OLS 推定量の統計的性質ー小標本理論

(改めて)$\epsilon = (\epsilon_1, \cdots, \epsilon_n)'$ とし、以下の仮定をおく。

A1: $y_i = \beta' x_i + \epsilon_i,\ i = 1, \cdots, n$
A2: X は $n \times k$ 行列で、$rank(X) = k$ ただし、$n > k$
A3: $E(\epsilon_i | X_i) = 0$
A4: $E(\epsilon\epsilon' | X) = \sigma^2 I$
A5: $\epsilon_i | X_i \sim N(0, \sigma^2 I)$

仮定 A1-A4 の下で、
(S1): $E(b|X) = E(b) = \beta$
(S2): $Var(b|X) = \sigma^2 (X'X)^{-1}$
(S3): b は β の最小分散線形不偏推定量 (BLUE) である。
(S4): $Cov(b, e|X) = 0$
(S5): $E(s^2|X) = E(s^2) = \sigma^2$

更に A5 を付加すると、以下 X を条件づけて、
(S6): b と e は独立である。したがって、b と s^2 は独立である。
(S7): $b \sim N(\beta, \sigma^2(X'X)^{-1})$ である。
(S8): β と σ^2 の最尤推定量は

$$\hat{\beta} = b,\ \ \hat{\sigma}^2 = \frac{e'e}{n}$$

(S9): b は最小分散不偏推定量（BUE）である。すなわち、Cramer-Rao Lower bound を達成する推定量である。

(S10): $(n-k)s^2/\sigma^2 \sim \chi^2(n-k)$ である。したがって、$E(s^2) = \sigma^2$、$Var(s^2) = 2\sigma^4/(n-k)$ である。

(S11): $\frac{b_j - \beta_j}{\{s^2(X'X)^{-1}_{jj}\}^{1/2}} \sim t(n-k)$ である。ただし、$(X'X)^{-1}_{jj}$ は $(X'X)^{-1}$ の第 jj 要素である。(→ t 検定)

(S12): $\beta_2 = \cdots = \beta_k = 0$ が正しいとき、$\frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k)$ である。(→ F 検定)

1.3 OLS 推定量の統計的性質ー大標本理論

前節の仮定 A1, A2, A4 および、以下を仮定する。

A3L: $E(\epsilon_i X_i) = 0$
A6: $\frac{1}{n}X'X \xrightarrow{p} \Sigma_{xx}$、ただし Σ_{xx} は正値定符号行列である。

そのとき、
(L1): $b \xrightarrow{p} \beta$
(L2): $\sqrt{n}(b-\beta) \xrightarrow{d} N(0, \sigma^2\Sigma_{xx}^{-1})$
(L3): $g : R^k \to R$ を β で微分可能な関数、$C = \partial g(u)/\partial u|_{u=\beta}$ とすると、

$$\sqrt{n}\{g(b) - g(\beta)\} \xrightarrow{d} N(0, \sigma^2 C' \Sigma_{xx}^{-1} C)$$

(L4): $s^2 \xrightarrow{p} \sigma^2$

(L5): $\frac{b_j-\beta_j}{\{s^2(X'X)^{-1}_{jj}\}^{1/2}} \xrightarrow{d} N(0,1)$ である。

(L6): $\beta_2=\cdots=\beta_k=0$ が正しいとき、$\frac{R^2/(k-1)}{(1-R^2)/(n-k)} \xrightarrow{d} \chi^2(k-1)$ である。

1.4　回帰係数の検定

小標本検定（正規性の仮定が正しい時）

β の各要素について検定したいときは、(S11) を用いて t 検定を行う。定数項を除くすべての係数について有意性検定を行いたいときは、(S12) を用いて F 検定を行う。より一般的に β に関する J 本の線形制約 $R\beta=r$ について検定を行いたいときは

$$\frac{(Rb-r)'[R(X'X)^{-1}R']^{-1}(Rb-r)}{s^2J} \sim F(J,n-k)$$

を用いて F 検定を行う。

大標本検定

正規性の仮定 A5 は必ずしも正しいとは限らない。その場合、標本数が大きければ、代わりに大標本理論に基づく結果を用いて近似的に検定（や区間推定）を行うことが可能である。β の各要素について検定したいときは、(L5) を用いて検定を行う。もちろん棄却域は N(0,1) 表から構成する。定数項を除くすべての係数について有意性検定を行いたいときは、(L6) を用いて検定を行う。より一般的に β に関する J 本の線形制約 $R\beta=r$ について検定を行いたいときは

$$\frac{(Rb-r)'[R(X'X)^{-1}R']^{-1}(Rb-r)}{s^2} \xrightarrow{d} \chi^2(J)$$

を用いて検定を行う。また、β に関する非線形制約の検定を行いたいときは (L3) を用いて検定を行う。

1.5　仮定の拡張

上の結果は証明を簡便にするための仮定を含んでいる。特に、経済データの特徴に鑑みて、無作為標本の仮定、A3(A3L)、A4、A6 は強すぎる場合が多い。A4 が満たされないのは、誤差項に分散不均一 and/or 自己相関がある場合である。通常、A4 が満たされなくても OLS の不偏性は成立する。しかし、推定量の分散について (S2) は成立しなくなる。そのため、(S11)、(S12)、(L5)、(L6) を用いた検定等では不具合を生ずる。その場合は、一般化最小二乗法（GLS）に基づく推定、検定を行うことにより解決できる。

A4': $E(\epsilon\epsilon'|X)=\Sigma$

を仮定すると、GLS 推定量は

$$b_{GLS}=argmin_\beta Q_{GLS}(\beta),\ \ Q_{GLS}(\beta)=(Y-X\beta)'\Sigma^{-1}(Y-X\beta)$$

により定義され、その解は

$$b_{GLS}=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \tag{1}$$

となる。A1, A2, A3,A4' の下で、

(S1GLS): $E(b_{GLS}|X)=E(b_{GLS})=\beta$

(S2GLS): $Var(b_{BLS}|X)=(X'\Sigma^{-1}X)^{-1}$

また、A1, A2, A3L,A4' の下で、

(L1GLS): $b_{GLS} \xrightarrow{p} \beta$

(L2GLS): $\sqrt{n}(b_{GLS} - \beta) \xrightarrow{d} N(0, \Omega^{-1}), \quad \Omega = plim\, \frac{1}{n} X'\Sigma^{-1}X$

が成立する。一般に Σ は未知である場合が多いが、通常少ない個数のパラメータ θ で $\Sigma = \Sigma(\theta)$ と表されるモデルを考える。何らかの手法で θ を一致推定して、(1) の Σ を $\hat{\Sigma} = \Sigma(\hat{\theta})$ で置き換えて、$b_{FGLS} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y$ で feasible な推定量を構成し、それを Feasible GLS 推定量という。

A6 は、たとえば時系列データで説明変数に時間が含まれる場合、満たされない。しかし、A6 は以降の結果を導くための一つの（強い）十分条件にすぎず、実際には多くの場合問題にはならない（より弱い十分条件は Grenander conditions、cf：Econometric Analysis (W.H.Greene)）。また、上の結果は比較的初歩的な漸近理論を用いることで証明可能であるが、時系列データへの応用の際には少し難易度の高い漸近理論が必要である。

A3 が満たされないことは OLS、GLS 推定量にとっては致命的な問題である。なぜなら、その場合は (S1)、(L1) が成立せず、何を推定していることになっているかわからないからである。これに対応するための推定法が操作変数法 (IV)、一般化モーメント法 (GMM) である。

2　操作変数法（IV 法）

2.1　仮定 A3（$E(\epsilon_i|X_i)=0$）、A3L($E(\epsilon_i X_i)=0$) と OLS の不偏性、一致性

A3(A3L) が成立しない場合、次に示す通り OLS は一般に不偏性 (一致性) を失うことになり、推定法としては無意味になる。$E(\epsilon_i|x_i)=g(x_i)\neq 0$ とする。$b=\beta+(X'X)^{-1}X'\epsilon$ から、

$$E(b|X)=\beta+(X'X)^{-1}X'E(\epsilon|X)=\beta+(X'X)^{-1}X'g$$

よって $g'=(g(x_1),\cdots,g(x_n))$ として、たまたま $E\{(X'X)^{-1}X'g\}=0$ が成立していない限り

$$E(b)=\beta+E\{(X'X)^{-1}X'g\}\neq\beta$$

従って、不偏性が満たされない。同様に、$E(\epsilon_i X_i)=\gamma\neq 0$ で A3L が成立していなければ

$$b=\beta+(\frac{1}{n}X'X)^{-1}\frac{1}{n}X'\epsilon$$

で、

$$\frac{1}{n}X'\epsilon=\frac{1}{n}\sum x_i\epsilon_i\xrightarrow{p}E(x_i\epsilon_i)=\gamma\neq 0$$

なので、

$$b=\beta+(\frac{1}{n}X'X)^{-1}\frac{1}{n}X'\epsilon\xrightarrow{p}\beta+\Sigma_{xx}^{-1}\gamma\neq\beta$$

となり、一致性が満たされない。A3(A3L) が成立しない時は GLS も一般に不偏性 (一致性) を失う。

経済データにおいては、モデルの構成上 A3(L) が成立しなくなることがある。
<u>1. 同時方程式モデル（消費関数）</u>　I と G は非確率的な変数とする。

$$\begin{aligned} C &= \alpha_0+\alpha_1 Y+\epsilon \quad (2)\\ Y &= C+I+G \end{aligned}$$

より、C と Y について解くと、$Y=\frac{1}{1-\alpha_1}(\alpha_0+I+G+\epsilon)$ なので、

$$E(\epsilon|Y)=(1-\alpha_1)Y-(\alpha_0+I+G)$$

となり、(2) 式に関して、説明変数 Y を条件にした誤差項の条件付き期待値は 0 でない。
<u>2. 観測（測定）誤差を含む説明変数</u>

$$y=\beta'x^*+\epsilon \quad (3)$$

においては A1-A4 が満たされているが、x^* は直接観測されず、誤差 v を含んだ変数

$$x=x^*+v \quad (4)$$

が観測されているとする。ただし、単純化のために v は x^*、ϵ とは独立で $E(v)=0$ とする。そのとき、x^* の代理変数として x を用いて

$$y=\beta'x+u \quad (5)$$

に対して OLS 推定を行うことが考えられる。しかし、(3)、(4) から x^* を消去すると

$$\begin{aligned} y &= \beta'x^*+\epsilon=\beta'(x-v)+\epsilon\\ &= \beta'x+(\epsilon-v'\beta) \end{aligned}$$

となり、(5) の誤差項 u は $u = \epsilon - \beta' v$ であることがわかる。$\beta \neq 0$ として

$$E(xu) = E\{(x^* + v)(\epsilon - v'\beta)\} = -Var(v)\beta \neq 0$$

なので、A3 は満たされない。(注：$E(xu) \neq 0$ ならば $E(u|x) \neq 0$)

3. Dynamic regression + autocorrelated errors

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 y_{t-1} + \beta' x_t + \epsilon_t \\ \epsilon_t &= \rho \epsilon_{t-1} + u_t \end{aligned}$$

という構造の時系列モデルを考えると、説明変数 y_{t-1} は ϵ_{t-1} に依存しており、また誤差項 ϵ_t も ϵ_{t-1} に依存している。そのため、誤差項 ϵ_t と説明変数 y_{t-1} の間に相関が生じ、A3(L) が満たされない。

2.2 IV(instrumental variables) 法 ··· A3L が満たされない場合の一致推定

後に定める変数 $z_i = (1, z_{2i}, \cdots, z_{ki})'$ があり、無作為標本 (y_i, x_i, z_i) が得られたとする。また、z のデータ行列を $Z = (z_1, \cdots, z_n)'$ とする。以下を仮定する。

A1: $y_i = \beta' x_i + \epsilon_i,\ i = 1, \cdots, n$
A2': X、Z は $n \times k$ 行列で、$rank(X) = rank(Z) = k$ ただし、$n > k$
A3': $E(\epsilon_i z_i) = 0$
A4': $E(\epsilon\epsilon'|Z) = \sigma^2 I$
A7: $E(z_1 x_1') = \Sigma_{zx}$、$E(z_1 z_1') = \Sigma_{zz}$、$E(x_1 x_1') = \Sigma_{xx}$ で、それらは正則、有界である。

A3'、A4' を満たすような変数 z を操作変数といい、そのような Z に対して

$$b_{IV} = (Z'X)^{-1} Z'Y$$

を IV 推定量（操作変数推定量）という。

Theorem 1. 仮定 *A1,A2',A3',A4',A7* の下で、

$$\begin{aligned} &(i) && b_{IV} \xrightarrow{p} \beta \\ &(ii) && \sqrt{n}(b_{IV} - \beta) \xrightarrow{d} N(0, \sigma^2 \Sigma_{zx}^{-1} \Sigma_{zz} (\Sigma_{zx}^{-1})') \end{aligned}$$

（証明）

(i) 一致性

$$b_{IV} = (Z'X)^{-1} Z'(X\beta + \epsilon) = \beta + (\frac{1}{n} Z'X)^{-1} \frac{1}{n} Z'\epsilon$$

と書ける。(y_i, x_i, z_i) が i.i.d.（無作為標本）であるため、A7、A3' と大数の法則より

$$\frac{1}{n} Z'X \xrightarrow{p} E(z_1 x_1') = \Sigma_{zx} \tag{6}$$

$$\frac{1}{n} Z'\epsilon \xrightarrow{p} E(z_1 \epsilon_1) = 0$$

である。したがって、

$$b_{IV} \xrightarrow{p} \beta.$$

(ii) 漸近正規性

$$\sqrt{n}(b_{IV} - \beta) = (\frac{1}{n} Z'X)^{-1} \frac{1}{\sqrt{n}} Z'\epsilon \tag{7}$$

A3',A4' より $E(z_i\epsilon_i)=0$、$V(z_i\epsilon_i)=E[z_iE(\epsilon_i^2|z_i)z_i']=\sigma^2\Sigma_{zz}$ であるから、Lindeberg=Levi の中心極限定理より

$$\frac{1}{\sqrt{n}}Z'\epsilon=\frac{1}{\sqrt{n}}\sum z_i\epsilon_i \xrightarrow{d} N(0,\sigma^2\Sigma_{zz}) \tag{8}$$

したがって、(6)、(7)、(8) より (ii) が証明される。（証明終）

（※1） $z_i=x_i$ とおくと IV 推定量は OLS 推定量に一致するので、OLS は IV の特殊ケースと解釈できる。

（※2） 一部の説明変数が誤差項と相関をもたない場合、その説明変数に対しては操作変数を使わなくてもよい。例えば $E(\epsilon_i|x_{2i})=0$ なら、$(z_{3i},\cdots,z_{ki})$ を用意して、$(1,x_{2i},z_{3i},\cdots,z_{ki})$ を操作変数ベクトルと考えて IV 推定を行えばよい。

（※3） 実は A3(L) が成立している場合（OLS が機能する場合）に、IV 法を用いて推定するとどうなるか？(次節で A3(L) が成立しているかどうかを調べる検定を紹介する）

・A3', A4' が満たされる限り、定理 1 は成立し、IV 推定量は CAN である。

・OLS と IV 推定量の漸近分散を比較すると OLS の方が漸近効率が高い。

$$A.Var(b)=\sigma^2Q^{-1}\le\sigma^2\Sigma_{zx}^{-1}\Sigma_{zz}(\Sigma_{zx}^{-1})'=A.Var(b_{IV})$$

（※4） 弱操作変数（Weak instruments）問題: IV 法では、説明変数 x に対して、A3',A7 を満たす操作変数 z を用意する。説明変数がスカラーの場合、x と z の共分散が 0 でないなら A7 が満たされる。しかし、相関が非常に低い時、A7 が満たされていたとしても Σ_{zx}^{-1} が非常に不安定になり、OLS における多重共線性と同様の問題を引き起こし、実証上問題となる。A7 はスカラーで共分散が 0 でない場合の多変量への拡張である。

2.3 A3L が満たされているかどうかの検定（Wu-Hausman 検定）

2.1 節では、モデル上で説明変数と誤差項に相関が生ずる場合を例示した。しかし、実際に相関があるかどうかは実証的な問題であり、それを確かめる方法が必要である。モデルは

$$Y=X\beta+\epsilon$$

で、検定する仮説は

$$\begin{cases} H_0: & E(\epsilon_i x_i)=0 \\ H_1: & E(\epsilon_i x_i)\neq 0 \end{cases}$$

である。(y_i,x_i,z_i) は無作為標本、b、b_{IV} をそれぞれ β の OLS, IV 推定量とし、$s^2=\frac{1}{n-k}\sum(y_i-x_i'b)^2$ とする。

仮定 A4 を以下の仮定でおきかえる。

A4": $E(\epsilon\epsilon'|X,Z)=\sigma^2I$

Theorem 2. *A1,A2',A3',A4",A7* の下で、

$$(b-b_{IV})'[s^2\{(Z'X)^{-1}(Z'Z)(X'Z)^{-1}-(X'X)^{-1}\}]^{-1}(b-b_{IV})\begin{cases}\xrightarrow{d} & \chi^2(k)\ under\ \ H_0\\ \xrightarrow{p} & \infty\ under\ \ H_1\end{cases} \tag{9}$$

（証明）

$$b-b_{IV}=\{(X'X)^{-1}X'-(Z'X)^{-1}Z'\}\epsilon$$

より

$$\sqrt{n}(b-b_{IV})=[(\frac{1}{n}X'X)^{-1}\quad -(\frac{1}{n}Z'X)^{-1}]\frac{1}{\sqrt{n}}\begin{bmatrix}X'\\Z'\end{bmatrix}\epsilon$$

である。まず帰無仮説 H_0 が正しい場合を考える。Lindeberg=Leviの中心極限定理より

$$\frac{1}{\sqrt{n}}\begin{pmatrix} X' \\ Z' \end{pmatrix}\epsilon \xrightarrow{d} N(0,\sigma^2\begin{pmatrix} \Sigma_{xx} & \Sigma'_{zx} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix})$$

であり、また、A6、A7、大数の法則より

$$[(\frac{1}{n}X'X)^{-1} \quad -(\frac{1}{n}Z'X)^{-1}] \xrightarrow{p} [\Sigma_{xx}^{-1} \quad -\Sigma_{zx}^{-1}]$$

である。これらを用いて

$$\sqrt{n}(b-b_{IV}) \xrightarrow{d} N(0,\sigma^2(\Sigma_{zx}^{-1}\Sigma_{zz}\Sigma_{zx}^{-1'}-\Sigma_{xx}^{-1}))$$

となる（この漸近分散は $A.Var(b)-A.Var(b_{IV})$ となっていることに注意)。従って、

$$n(b-b_{IV})'[\sigma^2(\Sigma_{zx}^{-1}\Sigma_{zz}\Sigma_{zx}^{-1'}-\Sigma_{xx}^{-1})]^{-1}(b-b_{IV}) \xrightarrow{d} \chi^2(k) \tag{10}$$

を得る。H_0 が正しいので、(L4)、A6、A7、大数の法則より

$$s^2 n\{(Z'X)^{-1}(Z'Z)(X'Z)^{-1}-(X'X)^{-1}\} \xrightarrow{p} \sigma^2(\Sigma_{zx}^{-1}\Sigma_{zz}\Sigma_{zx}^{-1'}-\Sigma_{xx}^{-1}) \tag{11}$$

が得られ、(10) と (11) より (9) の上段が導出される。

次に、H_1 が正しくて $E(x\epsilon)=\Gamma\neq 0$ であるとき、Lindeberg-Leviの中心極限定理より

$$\frac{1}{\sqrt{n}}\begin{pmatrix} X' \\ Z' \end{pmatrix}\epsilon-\begin{pmatrix} \sqrt{n}\Gamma \\ 0 \end{pmatrix}=\frac{1}{\sqrt{n}}\begin{pmatrix} \sum(x_i\epsilon_i-\Gamma) \\ Z'\epsilon \end{pmatrix} \xrightarrow{d} N(\begin{pmatrix} 0 \\ 0 \end{pmatrix},\Omega), \quad \Omega=Var(\begin{bmatrix} x_1\epsilon_1 \\ z_1\epsilon_1 \end{bmatrix})$$

なので $n^{-1/2}X'\epsilon\approx\sqrt{n}\Gamma+O_p(1)$ であり、(9) の下段が得られる。（証明終）

（※1） 細かい点であるが、σ^2 の推定の候補には $s_{IV}^2=\frac{1}{n-k}\sum(y_i-x_i'b_{IV})^2$ と $s^2=\frac{1}{n-k}\sum(y_i-x_i'b)^2$ が考えられる。実際、定理２はどちらを用いても成立する。しかし、後者を用いるのには理由がある。H_0 が正しい時、両者とも σ^2 の一致推定量であるので、どちらを使っても漸近的には大きな差はない。H_1 が正しい時、

$$\begin{aligned} s^2 &= \frac{1}{n-k}\sum(y_i-x_i'b)^2=\frac{1}{n-k}\{\epsilon'\epsilon-\epsilon'X(X'X)^{-1}X'\epsilon\} \\ &\xrightarrow{p} \sigma^2-\Gamma'\Sigma_{xx}^{-1}\Gamma<\sigma^2 \end{aligned}$$

となり、他方

$$\begin{aligned} s_{IV}^2 &= \frac{1}{n-k}\sum(y_i-x_i'b_{IV})^2=\frac{1}{n-k}(Y-Xb_{IV})'(Y-Xb_{IV}) \\ &\xrightarrow{p} \sigma^2 \end{aligned}$$

である。そのため、H_1 が正しい場合は s^2 を用いる方が s_{IV}^2 を用いるよりも検定統計量の値が大きくなりやすく、帰無仮説を（正しく）棄却しやすくなるというメリットがある。

（※2） 一般に、二つの推定量 $\hat{\theta}_1,\hat{\theta}_2$ があって、

- $\hat{\theta}_1$ は H_0 と H_1 のどちらが正しい時もCANであるが、効率性はない。
- $\hat{\theta}_2$ は H_0 が正しい時にはCANで効率的であるが、H_1 が正しい時には一致性がない

という場合、上と同様の形で $\hat{\theta}_2-\hat{\theta}_1$ の二次形式によって検定統計量を構成し、H_0 vs H_1 の検定を行うことができる。このタイプの統計量を総称してHausman検定ということもある。

3　定常時系列モデル

経済データには、時系列で観測されるものも多い。そのようなデータの解析のために用いられるのが時系列モデルである。1980 年代以降、経済データには非定常なものが多いという指摘がなされ、非定常時系列モデルの分析が進められてきたが、定常時系列モデルはその基礎になるものである。

3.1　分散、自己共分散と定常性

$\{y_t\}$ を時間と共に観測される確率変数列とする。まず、時系列分析で決定的に重要な以下の用語を定義する。

Definition 1. 強定常
　任意の整数 $n, h, t_1, \cdots, t_n$ について、$y_{t_1}, \cdots, y_{t_n}$ と $y_{t_1+h}, \cdots, y_{t_n+h}$ の同時分布が等しいとき $\{y_t\}$ は強定常であるという。

　２次モーメントが存在するものとして、それより少し弱い定常性は以下のように定義される。

Definition 2. 弱定常（共分散定常、２次定常）
　以下の (a), (b) が成り立つとき、$\{y_t\}$ は弱定常であるという。
　(a) 期待値 $E(y_t) = \mu$ が t に依存しない。
　(b) 共分散 $E\{(y_t - \mu)(y_{t-s} - \mu)\}$ が s には依存するが t に依存しない。

　$\gamma_s = E\{(y_t - \mu)(y_{t-s} - \mu)\}$ を $\{y_t\}$ の s 次の自己共分散という。$s = 0$ なら、γ_0 は y_t の分散である。なお、$\gamma_s = \gamma_{-s}$ が成り立つ。また、$\rho_s = \gamma_s/\gamma_0$ を $\{y_t\}$ の s 次の自己相関係数という。
　　　ベクトル確率過程についても、同様に定常性が定義される。強定常については、全く同様である。期待値が t に依存しない定数で、自己共分散行列 $\Gamma_s = E\{(Y_t - \mu)(Y_{t-s} - \mu)'\}$ が s には依存するが t に依存しないとき、ベクトル確率過程 $\{Y_t\}$ は弱定常であるという。なお、ベクトル過程の自己共分散はスカラーの場合と違って $\Gamma_s = \Gamma'_{-s}$ となる。

Definition 3. エルゴード性
　任意の非負の整数 h、任意の $h+1$ 次元ベクトルの集合 A に対し、強定常過程 $\{y_t\}$ が

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I\{(y_t, y_{t+1}, \cdots, y_{t+h}) \in A\} = P\{(y_0, y_1, \cdots, y_h) \in A\}$$

を満たすとき、$\{y_t\}$ はエルゴード性をもつという。

　(※ 1) エルゴード性は、(※ 3) に述べるように時系列データについて大数の法則が働くための必要十分条件になっている。そのため、時系列データを用いて意味のある計量経済分析を行うためには本質的な性質である。

　(※ 2) Definition 1 は、確率過程論における通常の定義とは異なる。しかし、エルゴード性については種々の必要十分条件が調べられており（例えば「A First Course in Stochastic Processes」(Karlin and Taylor, 1975, p.487-488、「Probability」(Shiryaev, 1996, Chapter V, Section 2 を参照)、上の定義はその一つである。

　(※ 3) 計量経済学において重要なエルゴード性の必要十分条件は、$E|\phi(y_t, \cdots, y_{t+m})| < \infty$ を満たす任意の関数 ϕ について

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \phi(y_t, \cdots, y_{t+m}) \overset{a.s.}{\to} E\{\phi(y_0, \cdots, y_m)\}$$

が成立する、というものである。

Definition 4. ホワイトノイズ、強ホワイトノイズ

次の (a), (b) を満たす確率変数をホワイトノイズという。

(a) $E(\epsilon_t) = 0$

(b) $E(\epsilon_t \epsilon_s) = \begin{cases} \sigma^2 & \text{when } t = s \\ 0 & \text{otherwise} \end{cases}$

また、(a), (b) に加えて $\{\epsilon_t\}$ が互いに独立のとき、強ホワイトノイズという。ベクトルの場合は、(b) を $E(\epsilon_t \epsilon_t') = \Omega(> 0)$, $E(\epsilon_t \epsilon_s') = 0 (t \neq s)$ で置き換える。

3.2 MA(moving average:移動平均), AR(autoregression:自己回帰), ARMA モデル

計量経済学において有効な時系列モデルとして、MA、AR、ARMA モデルがある。$\{\epsilon_t\}$ をホワイトノイズとして、$\{y_t\}$ が $\epsilon_t, \cdots, \epsilon_{t-q}$ の一次結合

$$y_t = c + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \cdots + \psi_q \epsilon_{t-q}, \ \psi_q \neq 0 \tag{12}$$

で生成されるとき、$\{y_t\}$ は MA(q) モデルに従うという。また、$\{y_t\}$ が $y_{t-1}, \cdots, y_{t-p}$ と ϵ_t の一次結合

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t, \ \phi_p \neq 0 \tag{13}$$

で生成されるとき、$\{y_t\}$ は AR(p) モデルに従うという。それらを組み合わせて $\{y_t\}$ が

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \cdots + \psi_q \epsilon_{t-q} \tag{14}$$

で生成されるとき、$\{y_t\}$ は ARMA(p,q) モデルに従うという。ただし $\psi_q \neq 0, \phi_p \neq 0$ とする。

k 次元ベクトル過程を考える場合は、ϵ_t も k 次元ベクトルになり、係数は $k \times k$ 行列になる。

3.3 MA(∞)

計量経済モデルとしてはともかく、MA,AR,ARMA モデルの理論の展開のためには、MA(q) で q を ∞ に近づけた収束先を考えることが極めて有用であるため、まずその解説をする。$q \to \infty$ とするときに、MA(q) がうまく定義のできる確率変数に収束するとき、それを MA(∞) と書く。そのための一つの条件が係数列 $\{\psi_j\}$ の absolute summability、

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

である。もちろん、これが成立するためには $j \to \infty$ のとき $\psi_j \to 0$ でなければならない。なお、k 次元ベクトル過程では、ψ_j が $k \times k$ 行列になるが、その場合の absolute summability は以下のようになる。ψ_{jkl} を ψ_j の第 k, l 要素として、

$$\sum_{j=0}^{\infty} |\psi_{jkl}| < \infty \text{ for all} \, k, l = 1, \cdots, p$$

Theorem 3. *(Fuller (1976), Introduction to Time Series Analysis, Theorem 2.2.1, p.29)*

数列 $\{\psi_j\}$ が $\sum_{j=0}^{\infty} |\psi_j| < \infty$ を満たし、$E(Z_t^2) < K < \infty$ であるとき、以下をみたす確率変数列 X_t が存在する。

$$E\left\{X_t-\sum_{j=0}^{n}\psi_j Z_{t-j}\right\}^2 \rightarrow 0 \; as \; n\rightarrow\infty$$
$$E(X_t^2)<\infty$$

（証明）

任意の t, j について $E|Z_t||Z_j| \le \{E(Z_t^2)+E(Z_j^2)\}/2 < K$ が成立する。$j\rightarrow\infty$ のとき $\psi_j\rightarrow 0$ なので、任意の $\epsilon>0$ に対して、$\sum_{j=N}^{\infty}|\psi_j| < \sqrt{\epsilon/K}$ となるように N を選ぶことができる。t を固定して、$n>m>N$ に対して、

$$\begin{aligned}
E\left(\sum_{j=0}^{n}\psi_j Z_{t-j}-\sum_{j=0}^{m}\psi_j Z_{t-j}\right)^2 &= E\left(\sum_{j=m+1}^{n}\psi_j Z_{t-j}\right)^2\\
&\le \sum_{j=m+1}^{n}\psi_j^2 E(Z_{t-j}^2)+\sum_{j=m+1}^{n}\sum_{t=m+1}^{n}|\psi_j||\psi_t|E(|Z_t||Z_j|)\\
&\le \left(\sum_{j=m+1}^{n}\psi_j^2+\sum_{j=m+1}^{n}\sum_{t=m+1}^{n}|\psi_j||\psi_t|\right)K\\
&\le \left(\sum_{j=N}^{\infty}|\psi_j|\right)^2 K<\epsilon
\end{aligned}$$

従って、$\sum_{j=0}^{n}\psi_j Z_{t-j}$ は平均二乗の意味でコーシー列である。よって、その意味で収束先 X_t が存在する。(証明終)

（※１） 平均二乗の意味で収束するため、確率収束の意味でも収束する。

（※２） X_t が確率０の事象を除いて一意であることも証明される。

（※３） Z_t が互いに独立な確率変数列なら、almost sure の収束も証明される。

（※４） この定理は Z_t がホワイトノイズであることは不必要である。

（※５） $j\rightarrow\infty$ のとき $\psi_j\rightarrow 0$ なので、遠い過去の Z_j から X_t への影響は時間とともに小さくなっていく。

Proposition 1. *(Hayashi (2000), Proposition 6.1, p367)*

$\{\epsilon_t\}$ はホワイトノイズ、$\{\psi_j\}$ は *absolutely summable* な係数列とする。そのとき、

$$y_t=\mu+\sum_{j=0}^{\infty}\psi_j\epsilon_{t-j} \tag{15}$$

が定義され、

(a) $\{y_t\}$ は平均二乗収束し（*Theorem 3*)、弱定常である（*MA(∞)* と呼ばれる)。

(b) y_t の期待値と自己共分散 $\gamma_j=Cov(y_t,y_{t-j}),\ j=0,1,2,\cdots$ は以下で与えられる。

$$\begin{aligned}
E(y_t) &= \mu\\
\gamma_j &= \sigma^2\sum_{k=0}^{\infty}\psi_{j+k}\psi_k
\end{aligned}$$

(c) $\{\gamma_j\}$ は *absolutely summable* である。

(d) もし $\{\epsilon_t\}$ が強ホワイトノイズなら、$\{y_t\}$ は強定常でエルゴード性をもつ。

（※１） y_t は未来の誤差 ϵ_{t+j} を含まないため、one-sided であるという。これは、より一般的な線形過程 $y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$ の特殊ケースである。

（※２）Theorem 3 で Z_t はホワイトノイズである必要はなかった。Proposition 1(a),(b),(c) はそれに対応して、以下の proposition のように拡張することができる。

Proposition 2. *(Hayashi (2000), Proposition 6.2, p.369)*

$\{x_t\}$ は弱定常過程、$\{h_j\}$ は *absolutely summable* とする。そのとき、

(a) 各 t に対して

$$y_t = \sum_{j=0}^{\infty} h_j x_{t-j}$$

は平均二乗収束し、$\{y_t\}$ は弱定常である。

(b) もし $\{x_t\}$ の自己共分散が *absolutely summable* なら、$\{y_t\}$ の自己共分散も *absolutely summable* である。

ϵ_t がホワイトノイズであることから、MA(∞) の期待値と自己共分散は以下のように求められる。

$$E(y_t) = E(\mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}) = \mu + \sum_{j=0}^{\infty} \psi_j E(\epsilon_{t-j}) = \mu$$

$$\gamma_0 = E(\sum_{j=0}^{\infty} \psi_j \epsilon_{t-j})^2 = \sum_{j=0}^{\infty} \psi_j^2 E(\epsilon_{t-j}^2) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$

ここで、最右辺は

$$\sum_{j=0}^{\infty} \psi_j^2 \le (\sum_{j=0}^{\infty} |\psi_j|)^2 < \infty \tag{16}$$

より有界である。更に、$s = 1, 2, \cdots$ に対して

$$\gamma_s = E(\sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \sum_{i=0}^{\infty} \psi_i \epsilon_{t+s-i}) = E(\sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \sum_{i=1}^{\infty} \psi_{j+s} \epsilon_{t-i}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+s}$$

で、その最右辺は (16) と同様に

$$\sum_{j=0}^{\infty} \psi_j \psi_{j+s} \le \sum_{j=0}^{\infty} |\psi_j||\psi_{j+s}| \le \sum_{j=0}^{\infty} |\psi_j| \sum_{j=0}^{\infty} |\psi_{j+s}| \le (\sum_{j=0}^{\infty} |\psi_j|)^2 < \infty$$

なので、自己共分散は有界であることがわかる。

MA(q) については、$\psi_{q+1} = \psi_{q+2} = \psi_{q+3} = \cdots = 0$ なので、

$$\begin{aligned} \gamma_s &= \sigma^2 \sum_{j=0}^{q-s} \psi_j \psi_{j+s} \quad s = 0, 1, 2, \cdots, q \\ &= 0, \qquad\qquad s \ge q+1 \end{aligned}$$

となって、切断がある。

y_t が k 次元ベクトル過程なら、ϵ_t を k 次元ベクトルホワイトノイズとして、自己共分散行列は

$$\gamma_s = E\{(y_t - \mu)(y_{t-s} - \mu)'\} = \sum_{j=0}^{\infty} \psi_{s+j} \Omega \psi_j'$$

3.4 ラグオペレータ、ラグ多項式、安定性

AR(p), ARMA(p,q) の理論的性質を調べるためには、ラグ多項式が便利である。
（1） ラグオペレータ

Definition 5. ラグオペレータ L
時点を１期ずらす作用素 L をラグオペレータという。すなわち、$Lx_t = x_{t-1}$ である。

ラグオペレータは以下の性質を有する。

- a を定数として、$La = a$
- $L^2 x_t = L(Lx_t) = Lx_{t-1} = x_{t-2}$。一般に j を正の整数として、$L^j x_t = L^{j-1}(Lx_t) = L^{j-1}x_{t-1} = \cdots = x_{t-j}$。

（2）ラグ多項式
ARMA(p,q) モデル (14) は、ラグオペレータを用いて、

$$y_t = c + \phi_1 L y_t + \cdots + \phi_p L^p y_t + \epsilon_t + \theta_1 L \epsilon_t + \cdots + \theta_q L^q \epsilon_t$$

と書ける。更に、形式的に以下のように書くことができる。

$$(1 - \phi_1 L - \cdots - \phi_p L^p) y_t = c + (1 + \theta_1 L + \cdots + \theta_q L^q)\epsilon_t$$

$$\Phi(L) y_t = c + \Theta(L)\epsilon_t$$

$\Phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, $\Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$ といったラグオペレータに関する多項式をラグ多項式という。
（3）ラグ多項式の積
$\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$, $\beta(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots$, とする。$L$ を通常の変数とみなして $\alpha(L) \times \beta(L)$ を展開したときの L^j の係数を δ_j とおき、$\delta(L) = \delta_0 + \delta_1 L + \delta_2 L^2 + \cdots$ とする。そのとき、

$$\delta(L) = \alpha(L)\beta(L)$$

によりラグ多項式の積を定義する。ラグ多項式の積について以下の性質がある。

- $\alpha(L)\beta(L) = \beta(L)\alpha(L)$
- $\{\alpha_j\}$, $\{\beta_j\}$ が absolutely summable のとき $\{x_t\}$ が共分散定常なら、$\alpha(L)\beta(L)x_t$ はうまく定義される確率変数で、proposition 2 より共分散定常で、

  $$\alpha(L)\beta(L) = \delta(L) \Rightarrow \alpha(L)\beta(L)x_t = \delta(L)x_t$$

  が成立する。また、$\{\delta_j\}$ も absolutely summable である。

（4）ラグ多項式の inverse
$\alpha_0 \neq 0$ として、$\alpha(L)\beta(L) = 1$ を満たす $\beta(L)$ を $\alpha(L)$ の inverse といい、$\beta(L) = \alpha(L)^{-1}$ と書く。これを用いると、$\alpha_0 \neq 0$, $\beta_0 \neq 0$ として

$$\alpha(L)\alpha(L)^{-1} = \alpha(L)^{-1}\alpha(L)$$

$$\Phi(L)\Psi(L) = \delta(L) \Leftrightarrow \Psi(L) = \Phi(L)^{-1}\delta(L) \Leftrightarrow \Phi(L) = \delta(L)\Psi(L)^{-1}$$

が成立する。$\Phi(L)=1-\phi_1 L-\cdots-\phi_p L^p$,$\Psi(L)=\Phi(L)^{-1}=\psi_0+\psi_1 L+\psi_2 L^2+\psi_3 L^3+\cdots$ とすると、

$$\begin{aligned}
constant &: \psi_0=1\\
L &: \psi_1-\phi_1\psi_0=0\\
L^2 &: \psi_2-\phi_1\psi_1-\phi_2\psi_0=0\\
L^3 &: \psi_3-\phi_1\psi_2-\phi_2\psi_1-\phi_3\psi_0=0\\
&\vdots\\
L^p &: \psi_p-\phi_1\psi_{p-1}-\phi_2\psi_{p-2}-\phi_3\psi_{p-3}-\cdots-\phi_p\psi_0=0\\
L^{p+1} &: \psi_{p+1}-\phi_1\psi_p-\phi_2\psi_{p-1}-\phi_3\psi_{p-2}-\cdots-\phi_p\psi_1=0\\
L^{p+2} &: \psi_{p+2}-\phi_1\psi_{p+1}-\phi_2\psi_p-\phi_3\psi_{p-1}-\cdots-\phi_p\psi_2=0\\
&\vdots
\end{aligned}$$

によって逐次的に $\{\psi_j\}$ が決まる。例えば、

$\alpha(L)=1-\phi L$ とすると、

$$\alpha(L)^{-1}=1+\phi L+\phi^2L^2+\phi^3L^3+\cdots$$

となる。

（５）安定性条件

AR(p), ARMA(p,q) の定常性は AR 部分のラグ多項式 $\Phi(L)$ の安定性条件と密接に関係している。安定性条件とは、k 次方程式

$$\Phi(x)=1-\phi_1x-\phi_2x^2-\cdots-\phi_px^p=0 \tag{17}$$

の解がすべて絶対値で 1 より大きいことをさす。なお、k 次元ベクトル過程の場合は $\phi_1,\cdots,\phi_p$ はそれぞれ $k\times k$ 行列で、対応する安定性条件は、$|\cdot|$ を行列式として、

$$|\Phi(x)|=|I_k-\phi_1x-\phi_2x^2-\cdots-\phi_px^p|=0$$

の解がすべて絶対値で 1 より大きいことである。

Proposition 3. *(Hayashi (2000), Proposition 6.3, p374)*

今、$\Phi(L)=1-\phi_1L-\phi_2L^2-\cdots-\phi_pL^p$ が安定性条件をみたすとする。$\Psi(L)=\Phi(L)^{-1}=1+\psi_1L+\psi_2L^2+\psi_3L^3+\cdots$ とすると、すべての $j=0,1,2,\cdots$ について

$$|\psi_j| \quad < \quad Ab^j$$

をみたす定数 $A>0$,$b\in(0,1)$ が存在する。したがって

$$\sum_{j=0}^{\infty}|\psi_j|<\sum_{j=0}^{\infty}Ab^j=\frac{A}{1-b}<\infty$$

となり、$\{\psi_j\}$ は *absolutely summable* である。

（※１） 上の $p=1$ の例については、$\psi_j=\phi^j$ なので明らかである。

（※２）次節に示すとおり、この Proposition と Proposition 1 と組み合わせて AR(p), ARMA(p,q) の性質を調べる。

3.5　AR モデルと ARMA モデル

AR や ARMA モデルは、現在の値が基本的に過去の値に依存して決まるが、そこに誤差が入ってくるというモデルであり、計量経済分析になじみやすい時系列モデルである。

3.5.1 AR(1) の MA(∞) 表現

最初に最も簡単な例として AR(1)

$$y_t = c + \phi y_{t-1} + \epsilon_t \tag{18}$$

を取り上げる。ϵ_t はホワイトノイズである。ラグ多項式で書くと

$$(1-\phi L)y_t = c + \epsilon_t$$

また、$\phi \neq 1$ なら $\mu = c/(1-\phi)$ として、

$$(1-\phi L)(y_t - \mu) = \epsilon_t$$

と書ける。

（1）$|\phi| < 1$ のとき

$$\begin{aligned} y_t - \mu &= (1-\phi L)^{-1}\epsilon_t \\ &= (1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \cdots)\epsilon_t \\ &= \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \end{aligned} \tag{19}$$

であるが、$|\phi| < 1$ より $\{1, \phi, \phi^2, \phi^3, \cdots\}$ は absolutely summable なので、Theorem 3 より右辺は平均二乗収束する。また、Proposition 1(b) より $E(y_t) = \mu$ である。

（2）$|\phi| > 1$ のとき

$$y_t - \mu = \phi^{-1}(y_{t+1} - \mu) - \phi^{-1}\epsilon_{t+1}$$

である。したがって、

$$y_t - \mu = -\sum_{j=1}^{\infty} \phi^{-j} \epsilon_{t+j} \tag{20}$$

と書ける。$|\phi|^{-1} < 1$ なので、これも平均二乗収束する。しかし、ϵ の将来の値で現在の y が決まるという形になり、少なくとも計量経済学では応用しにくい。

（3）$|\phi| = 1$ のとき

$$y_t = c + y_{t-1} + \epsilon_t$$

なので、逐次代入により、任意の自然数 j に対して

$$\begin{aligned} y_t &= c + y_{t-1} + \epsilon_t \\ &= c + (c + y_{t-2} + \epsilon_{t-1}) + \epsilon_t = 2c + y_{t-2} + \epsilon_t + \epsilon_{t-1} \\ &= 2c + (c + y_{t-3} + \epsilon_{t-2}) + \epsilon_t + \epsilon_{t-1} = 3c + y_{t-3} + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} \\ &\vdots \\ &= jc + y_{t-j} + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \cdots + \epsilon_{t-j+1} \end{aligned}$$

となる。これは弱定常ではない。

（証明）$\{y_t\}$ が弱定常であると仮定する。

(i) $c \neq 0$ のとき

$E(\epsilon_t) = 0$ より $E(y_t) = jc + E(y_{t-j})$ である。弱定常性より $E(y_t) = E(y_{t-j})$ でなので、$jc = 0$ となる。上の変形は任意の自然数 j について成立するので、$c = 0$ となり、矛盾する。

(ii) $c=0$ のとき

$$y_t - y_{t-j} = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \cdots + \epsilon_{t-j+1}$$

となる。両辺の分散を取ると

$$Var(y_t) + Var(y_{t-j}) - 2Cov(y_t, y_{t-j}) = j\sigma^2$$

弱定常性より $Var(y_t) = Var(y_{t-j}) = \gamma_0$、$Cov(y_t, y_{t-j}) = \gamma_j$ なので、

$$2(\gamma_0 - \gamma_j) = j\sigma^2$$

したがって、j 次の自己相関係数は $\rho_j = 1 - \frac{\sigma^2}{2\gamma_0} j$ となる。j は任意の自然数なので $j > \frac{4\gamma_0}{\sigma^2}$ ととってもよいが、そのとき $\rho_j < -1$ となり、相関係数の性質と矛盾する。（証明終）

通常は（1）の $|\phi| < 1$ である場合が AR(1) の定常性の条件とされる。（2）$|\phi| > 1$ の場合でも、上に示したように (18) は数学的には共分散定常な解をもつわけであるが、現実的な意味は大きく異なる。（1）は時間の流れと整合的であるが、（2）は時間が逆に流れている状況でなければ理解しにくい。（1）の状況では、$E(y_0) = 0, Var(y_0) = \sigma^2/(1-\phi^2)$ を満たすように y_0 が発生していれば、その後は毎期 ϵ_t が発生し、(18) に従って逐次的に $y_1, y_2, \cdots$ を構成すれば定常過程が得られる。しかし、（2）の場合は、先にすべての $\epsilon_t, t = 1, 2, \cdots$ を発生させて、そこから (20) を計算して「初期値」y_0 を作って初めて定常過程となる。また、(20) を見ればわかるように、y_t は $\epsilon_{t+1}, \epsilon_{t+2}, \epsilon_{t+3}, \cdots$ で決まっていて、次期の y_{t+1} はそこから ϵ_{t+1} の部分を抜いたものとして作られるわけである。その意味で、数学的にはともかく現実的には（2）の状況は不自然である。そのため、通常は（1）の場合のみ考える。別の言い方をすると、将来の誤差 ϵ に依存させずに初期値を定めて確率過程をスタートさせる場合は、（2）$|\phi| > 1$ の場合、定常にならない。

3.5.2 定常 AR(1) の自己共分散と自己相関係数

MA(∞) 表現 (19) を用いると、$\epsilon_t, t = 1, 2, \cdots$ がホワイトノイズであることから AR(1) の自己共分散を簡単に求めることができて、

$$\gamma_j = \frac{\sigma^2 \phi^j}{1 - \phi^2}$$

であることがわかる。また、自己相関係数は

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi^j$$

となる。また、別の方法として Yule=Walker 方程式を用いるやり方もある。単純化のために $\mu = 0$ とする。$y_t = \phi y_{t-1} + \epsilon_t$ の両辺に y_t, y_{t-j}, ϵ_t $(j = 1, 2, \cdots)$ をかけて期待値をとると

$$\begin{aligned} E(y_t^2) &= \phi E(y_{t-1} y_t) + E(\epsilon_t y_t) \\ E(y_t y_{t-j}) &= \phi E(y_{t-1} y_{t-j}) + E(\epsilon_t y_{t-j}) \\ E(\epsilon_t y_t) &= \phi E(\epsilon_t y_{t-1}) + E(\epsilon_t^2) \end{aligned}$$

を得る。$E(\epsilon_t^2) = \sigma^2$、$E(\epsilon_t y_{t-j}) = 0$ なので、

$$\begin{aligned} \gamma_0 &= \phi \gamma_1 + \sigma^2 \\ \gamma_j &= \phi \gamma_{j-1}, \ j = 1, 2, \cdots \end{aligned}$$

である。これを Yule=Walker 方程式という。第１式と第２式の $j = 1$ の場合から γ_0, γ_1 が得られ、第２式から逐次的に $\gamma_2, \gamma_3, \cdots$ が得られる。

3.5.3 AR(p) の MA(∞) 表現と自己共分散

3.5.1 節で例示した内容を一般の AR(p)

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

に拡張する。ラグ多項式を用いると、

$$\Phi(L) y_t = c + \epsilon_t \tag{21}$$

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$

である。

Proposition 4. *(Hayashi (2000), Proposition 6.4, p.379)*
$\Phi(L)$ が安定性条件 *(17)* を満たすとき、
(a) (21) は一意な弱定常解をもち、*MA(∞)* 表現

$$y_t = \mu + \Psi(L)\epsilon_t$$

$$\Psi(L) = \Phi(L)^{-1} = 1 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \cdots$$

される。また、すべての $j = 0, 1, 2, \cdots$ について

$$|\psi_j| \quad < \quad A_0 b_0^j$$

をみたす定数 $A_0 > 0, b_0 \in (0,1)$ が存在し、従って $\{\psi_j\}$ は *absolutely summable* である。
(b) $E(y_t) = \mu = \Phi(1)^{-1} c$ である。
(c) すべての $j = 0, 1, 2, \cdots$ について

$$|\gamma_j| \quad < \quad A_1 b_1^j$$

をみたす定数 $A_1 > 0, b_1 \in (0,1)$ が存在し、$\{\gamma_j\}$ は *absolutely summable* である。

(a) は Proposition 3 から明らかである。(b) は Proposition 1(b) から明らかである。(c) は (a) と Proposition 1(b) を組み合わせれば簡単に証明できる。AR(1) の場合と同様に MA(∞) 表現もしくは Yule=Walker 方程式から自己共分散が計算できる。MA(∞) を用いる場合は $\{\psi_j\}$ が必要であるが、3.4 節に示した方法で逐次的に求めることができる。

3.5.4 ARMA(p,q) の MA(∞) 表現と自己共分散

ARMA(p,q) モデル (14) は、ラグ多項式 $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$、$\Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$ を用いて

$$\Phi(L) y_t = c + \Theta(L)\epsilon_t$$

と書ける。また、$\Phi(1) \neq 0$ のとき、$\mu = c/\Phi(1)$ として

$$\Phi(L)(y_t - \mu) = \Theta(L)\epsilon_t$$

となる。

Proposition 5. *(Hayashi (2000), Proposition 6.5, p.381)*
$\Phi(z)$ が安定性条件を満たすとき、
(a) ARMA(p,q) は一意な共分散定常な解

$$y_t = \mu + \Psi(L)\epsilon_t$$

$$\Psi(L) = \Phi(L)^{-1}\Theta(L)$$

をもつ。$\Psi(L)$ の係数 $\{\psi_j\}$ は絶対値で幾何的に減少する定数列でおさえられ、従って *absolutely summable* である。

(b) y_t の期待値は $E(y_t) = \mu = c/\Phi(1)$ である。

(c) 自己共分散 $\{\gamma_j\}$ は絶対値で幾何的に減少する定数列でおさえられ、従って *absolutely summable* である。

証明は AR(p) の場合と同様にできる。(a) は Hayashi(2000, p381-382) 参照。 (b), (c) は Proposition 1(b)、(c) から明らかである。

ARMA(p,q) の自己共分散は Proposition 5(a) の MA(∞) 表現または、AR(p) と同様に Yule=Walker 方程式から求めることができる。$\Theta(z)$ が安定性条件を満たすとき、ARMA(p,q) は反転可能であるという。そのとき、ARMA(p,q) は AR(∞) 表現を持ち、

$$\Theta(L)^{-1}\Phi(L)y_t = \frac{c}{\Theta(1)} + \epsilon_t$$

と書ける。なお、MA(q) は ARMA(p,q) から AR 部分を取り除いたものなので、上と同様に MA(q) の反転可能性を定義する。

3.6 時系列モデルの推定

AR(p), MA(q), ARMA(p,q) のいずれかのモデルを推定するときは、自己共分散によるモーメント法を用いることができる。AR(p) は回帰モデルの形をしているために OLS が適用できる。ARMA(p,q) は回帰部分を持つが、誤差項と説明変数の間に相関が生じるために OLS を使うことはできず、IV を適用することになる。また、MA(q),ARMA(p,q) については誤差項の分布が仮定できる場合には最尤法を用いることができる。以下では、次数 p,q は既知とする。実際にはそれらは未知である。MA(q) については、自己共分散を推定し、その切断点から q を決めることができる。最尤法の枠組みを取る場合は AIC（赤池の情報量基準）が用いられる。これは真のモデルと推定モデルの近さを測る尺度で、これが小さいモデルが選ばれる。。

3.6.1 時系列の漸近理論

$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ とする。

（1）大数の法則

下に述べる大数の法則が成立する。証明のために、まず以下の Lemma を準備する。

Lemma 1. $\{S_j\}$ を定数列として、$\lim_{j\to\infty} S_j = s$ のとき、$\lim_{T\to\infty}\frac{1}{T}\sum_{j=1}^{T} S_j = s$ が成立する。

（証明）

仮定より、任意の $\epsilon > 0$ に対して十分大きい T_0 を選ぶと、すべての $j > T_0$ に対して $|S_j - s| < \epsilon/2$ となる。

$$\begin{aligned} |\frac{1}{T}\sum_{j=1}^{T} S_j - s| &\leq \frac{1}{T}\sum_{j=1}^{T_0}|S_j - s| + \frac{1}{T}\sum_{j=T_0+1}^{T}|S_j - s| \\ &\leq \frac{1}{T}\sum_{j=1}^{T_0}|S_j - s| + \frac{\epsilon}{2} \end{aligned}$$

T_0 は固定されているので T が十分大きければ $\frac{1}{T}\sum_{j=1}^{T_0}|S_j - s| < \epsilon/2$ となる。したがって、$|\frac{1}{T}\sum_{j=1}^{T} S_j - s| < \epsilon$ となる。

Lemma 2. $\lim_{T\to\infty}\sum_{j=0}^{T} a_j = A < \infty$ のとき、$\lim_{T\to\infty}\sum_{j=1}^{T}\frac{j}{T}a_j = 0$ である。

（証明） まず

$$\begin{aligned}
\sum_{j=1}^{T}\frac{j}{T}a_j &= \frac{1}{T}\{(a_1+a_2+\cdots+a_T)+(a_2+a_3+\cdots+a_T)+\cdots \\
&\quad +(a_{T-1}+a_T)+a_T\} \\
&= \frac{1}{T}\sum_{j=1}^{T}\sum_{k=j}^{T}a_k \le \frac{1}{T}\sum_{j=1}^{T}|\sum_{k=j}^{T}a_k| \\
&= \frac{1}{T}\sum_{j=1}^{T_0}|\sum_{k=j}^{T}a_k| + \frac{1}{T}\sum_{j=T_0+1}^{T}|\sum_{k=j}^{T}a_k|
\end{aligned}$$

である。この表現は任意の T_0 について成り立つ。まず第二項を調べる。仮定から、任意の $\epsilon > 0$ に対して十分大きい T_0 を選ぶと、$T > j > T_0$ をみたす任意の j, T について $|\sum_{k=j}^{T}a_k| < \epsilon/2$ とできる。従って、

$$\frac{1}{T}\sum_{j=T_0+1}^{T}|\sum_{k=j}^{T}a_k| < \frac{\epsilon(T-T_0)}{2T} < \frac{\epsilon}{2}$$

である。第一項は次のように抑えられる。任意の T と任意の $j\in[1,T]$ について $|\sum_{k=j}^{T}a_k| < M < \infty$ となる $M > 0$ が存在する。従って、十分大きい T について、

$$\frac{1}{T}\sum_{j=1}^{T_0}|\sum_{k=j}^{T}a_k| < \frac{T_0 M}{T} < \frac{\epsilon}{2}$$

となる。

したがって、$\sum_{j=1}^{T}\frac{j}{T}a_j < \epsilon$ となる。

Theorem 4. *(Hayashi, Proposition 6.8, p401)*

$\{y_t\}$ は共分散定常で、$E(y_t)=\mu$、$Cov(y_t, y_{t-j})=\gamma_j$ とする。

(a) もし $\lim_{j\to\infty}\gamma_j = 0$ なら、$\bar{y} \overset{m.s.}{\to} \mu$ *as* $T\to\infty$

(b) もし $\{\gamma_j\}$ が *summable* なら、$\lim_{T\to\infty}Var(\sqrt{T}\bar{y}) = \sum_{j=-\infty}^{\infty}\gamma_j = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j < \infty$

（証明）

(a) 共分散定常の仮定から

$$\begin{aligned}
Var(\bar{y}) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}Cov(y_t, y_s) \\
&= \frac{2}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t}Cov(y_t, y_s) - \frac{1}{T^2}\sum_{t=1}^{T}Var(y_t) \\
&= \frac{2}{T^2}\sum_{t=1}^{T}tCov(y_t, \bar{y}_t) - \frac{1}{T^2}\sum_{t=1}^{T}Var(y_t) \\
&\le \frac{2}{T}\sum_{t=1}^{T}|Cov(y_t, \bar{y}_t)|
\end{aligned}$$

ただし、３つ目の等号は $\bar{y}_t = \frac{1}{t}\sum_{j=1}^{t} y_j$ として、

$$tCov(y_t, \bar{y}_t) = tCov(y_t, \frac{1}{t}(y_1 + \cdots + y_t)) = \sum_{s=1}^{t} Cov(y_t, y_s)$$

を用いている。また、不等号は $t/T \leq 1$、$Var(y_t) > 0$ を用いた。仮定から $\lim_{j\to\infty}\gamma_j = 0$ なので、$t \to \infty$ のとき Lemma 1 より

$$Cov(y_t, \bar{y}_t) = \frac{1}{t}\sum_{j=0}^{t-1}\gamma_j \to 0$$

が得られ、$|Cov(y_t, \bar{y}_t)| \to 0$ となる。したがって、再び Lemma 1 を用いると

$$\frac{2}{T}\sum_{t=1}^{T}|Cov(y_t, \bar{y}_t)| \to 0$$

となる。よって、$Var(\bar{y}) \to 0$ であり、$\bar{y} \overset{m.s.}{\to} \mu$ as $T \to \infty$。

(b)

$$\begin{aligned}
TVar(\bar{y}) &= \frac{1}{T}E\{(y_T - \mu) + (y_{T-1} - \mu) + \cdots + (y_1 - \mu)\}\{(y_T - \mu) + (y_{T-1} - \mu) + \cdots + (y_1 - \mu)\} \\
&= \frac{1}{T}\{T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 4\gamma_{T-2} + 2\gamma_{T-1}\} \\
&= \gamma_0 + 2\sum_{j=1}^{T-1}(1 - \frac{j}{T})\gamma_j \\
&= \gamma_0 + 2\sum_{j=1}^{T-1}\gamma_j - 2\sum_{j=1}^{T-1}\frac{j}{T}\gamma_j
\end{aligned}$$

仮定より $\{\gamma_j\}$ は summable なので、Lemma 2 より $T \to \infty$ のとき $\sum_{j=1}^{T-1}\frac{j}{T}\gamma_j \to 0$ である。従って、

$$Var(\sqrt{T}\bar{y}) = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j < \infty$$

（証明終）

なお、summability は absolute summability より少し弱い仮定である。前者は和の順番を入れ替えると結果が変わるが、後者は順番を入れ替えても収束先は変わらない。

（２）中心極限定理

次に、時系列について成立する中心極限定理を証明なしで示す。

Theorem 5. *CLT for MA(∞) (Hayashi, Proposition 6.9, p402)*

$\{y_t\}$ は *MA(∞)* 表現

$$y_t = \mu + \Psi(L)\epsilon_t$$

を持ち、$\{\epsilon_t\}$ は強ホワイトノイズであるとする。そのとき、

$$\sqrt{T}(\bar{y} - \mu) \overset{d}{\to} N(0, \sum_{j=-\infty}^{\infty}\gamma_j)$$

Gordin(1961) を用いると以下の中心極限定理が示される。

Theorem 6. *Gordin (Hayashi, Proposition 6.10, p404)*
$\{y_t\}$ はエルゴード性をもつ定常過程とする。$I_t = (y_t, y_{t-1}, y_{t-2}, \cdots)$ として、以下の条件が成立するとする。
(a) $E(y_t^2) < \infty$
(b) $E(y_t|I_{t-j}) \overset{m.s.}{\to} 0$ *as* $j \to \infty$
(c) $r_{tj} = E(y_t|I_{t-j}) - E(y_t|I_{t-j-1})$ として、

$$\sum_{j=0}^{\infty}\{E(r_{tj}^2)\}^{\frac{1}{2}} < \infty$$

そのとき、

$$\sqrt{T}(\bar{y}-\mu) \overset{d}{\to} N(0, \sum_{j=-\infty}^{\infty}\gamma_j)$$

また、martingale difference について以下の中心極限定理が成立する (Billingsley, 1961)。

Theorem 7. *Martingale CLT (Hayashi, p.106)*
$\{y_t\}$ はエルゴード性をもつ定常な *martingale difference* とする。更に $Var(y_t) = \Sigma$ は有界とする。そのとき、

$$\sqrt{T}(\bar{y}-\mu) \overset{d}{\to} N(0, \Sigma),\ \mu = 0$$

3.6.2 AR(p) モデルの推定

AR(p) の推定には、OLS, Yule-Walker 方程式（モーメント法)、最尤法が考えられる。
（1）OLS
$\{\epsilon_t\}$ が強ホワイトノイズのとき、AR(p) モデル

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t \\ &= \beta' x_t + \epsilon_t \end{aligned} \tag{22}$$

では、「説明変数」$x_t = (1, y_{t-1}, \cdots, y_{t-p})'$ は $t-1$ 時点以前の innovation$\epsilon_{t-1}, \epsilon_{t-2}, \cdots$ で構成されているためにモデルの誤差項 ϵ_t とは独立であり、名前の通り回帰モデルの形を持っている。なお、$\beta = (c, \phi_1, \cdots, \phi_p)'$ である。従って、OLS 推定により一致推定量を得ることができる。OLS 推定量を

$$\hat{\beta} = \begin{bmatrix} \hat{c} \\ \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = (\sum_{t=p+1}^{T} x_t x_t')^{-1} \sum_{t=p+1}^{T} x_t y_t$$

とする。

Proposition 6 でこの推定量の性質を述べるが、その証明に次の結果を用いる。

Remark 1. 定常エルゴード過程の関数の定常エルゴード性（Karlin and Taylor (1975） A first course in stochastic processes, 2nd ed., Remark 5.3, p.488)
$\{x_t\}$ が定常エルゴードであるとき、任意の可測関数 ϕ について、$y_t = \phi(x_t, x_{t+1}, \cdots)$ は定常エルゴードである。

証明は Stout(1974, Almost sure convergence, p.182)。ベクトルへの拡張も可能である（White(1999）, Asymptotic Theory for Econometricians, Theorem 3.35, p.44)。

Proposition 6. *(Hayashi(2000), Proposition 6.7, p393, Fuller (1976), Theorem 8.2.1, p335)*

$\{y_t\}$ は *(22)* に従う $AR(p)$ 過程で、ラグ多項式は安定性条件を満たすものとする。また、$\{\epsilon_t\}$ は期待値 *0*、分散 σ^2 の強ホワイトノイズであるとする。そのとき、

$$A = E(x_t x_t') = \begin{bmatrix} 1 & \mu & \mu & \cdots & \mu \\ \mu & \gamma_0 + \mu^2 & \gamma_1 + \mu^2 & \cdots & \gamma_{p-1} + \mu^2 \\ \mu & \gamma_1 + \mu^2 & \gamma_0 + \mu^2 & \cdots & \gamma_{p-2} + \mu^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu & \gamma_{p-1} + \mu^2 & \gamma_{p-2} + \mu^2 & \cdots & \gamma_0 + \mu^2 \end{bmatrix}$$

を正値定符号な対称行列として

(i) $\hat{\beta} \xrightarrow{p} \beta$

(ii) $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 A^{-1})$

(iii) $\hat{A} = T^{-1} \sum_{t=p+1}^{T} x_t x_t'$、$s^2 = \frac{1}{T} \sum_{t=p+1}^{T} (y_t - \hat{\beta}' x_t)^2$ として、

$$s^2 \hat{A}^{-1} \xrightarrow{p} \sigma^2 A^{-1}$$

である。

（証明）

普通のクロスセクションデータの OLS 推定量と同様に

$$\hat{\beta} = \beta + (T^{-1} \sum_{t=p+1}^{T} x_t x_t')^{-1} T^{-1} \sum_{t=p+1}^{T} x_t \epsilon_t \tag{23}$$

である。安定性条件の仮定の下で、Proposition 4(a) より y_t は係数が absolutely summable な MA(∞) 表現をもつ。更に、仮定より $\{\epsilon_t\}$ が強ホワイトノイズなので、Proposition 1(d) から y_t はエルゴード性をもつ定常過程である。したがって、

$$\frac{1}{T} \sum_{t=p+1}^{T} x_t x_t' \xrightarrow{a.s.} A \tag{24}$$

また、$x_t = \{1, y_{t-1}, \cdots, y_{t-p}\}'$ なので、$E(x_t \epsilon_t | x_{t-1}\epsilon_{t-1}, x_{t-2}\epsilon_{t-2}, \cdots) = 0$ となる。すなわち、$\{x_t \epsilon_t\}$ は martingale difference である。更に、Remark 1 から定常エルゴードである。Theorem 4(a) より、

$$T^{-1} \sum_{t=p+1}^{T} x_t \epsilon_t \xrightarrow{m.s.} 0 \tag{25}$$

である。(23), (24), (25) より、(i) が示される。

(ii) $\{x_t \epsilon_t\}$ は定常エルゴードな martingale difference で、$V(x_t \epsilon_t) = E(\epsilon_t^2 x_t x_t') = E\{E(\epsilon_t^2 | x_t) x_t x_t'\} = \sigma^2 A$ なので、Theorem 7 より、

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} x_t \epsilon_t \xrightarrow{d} N(0, \sigma^2 A) \tag{26}$$

したがって、(23)、(24), (26) より (ii) を得る。

(iii)　$\{y_t\}$ はエルゴード性を持つ定常過程なので、$\hat{A} \xrightarrow{p} A$ である。また、$y_t - \hat{\beta}' x_t = \epsilon_t - (\hat{\beta} - \beta)' x_t$ なので、

$$\begin{aligned} s^2 &= \frac{1}{T} \sum_{t=p+1}^{T} (y_t - \hat{\beta}' x_t)^2 \\ &= \frac{1}{T} \sum_{t=p+1}^{T} \epsilon_t^2 - 2(\hat{\beta} - \beta)' \frac{1}{T} \sum_{t=p+1}^{T} x_t \epsilon_t + (\hat{\beta} - \beta)' \frac{1}{T} \sum_{t=p+1}^{T} x_t x_t' (\hat{\beta} - \beta) \end{aligned}$$

となる。強ホワイトノイズに関する大数の法則から $\frac{1}{T} \sum_{t=p+1}^{T} \epsilon_t^2 \xrightarrow{p} \sigma^2$ なので、(i)、(24)、(25) より

$$s^2 \xrightarrow{p} \sigma^2$$

である。（証明終）

（２）Yule=Walker 方程式からのモーメント法

モーメント法によって AR(p) のパラメータを推定することも可能である。定数項については、定常性と

$$E(y_t) = c + \phi_1 E(y_{t-1}) + \cdots + \phi_p E(y_{t-p})$$

より、

$$(1 - \phi_1 - \cdots - \phi_p) E(y_t) = c$$

というモーメント条件を得る。また、AR(p) の Yule=Walker 方程式は

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}$$

である。これらから、

$$\begin{aligned} \bar{y} &= \frac{1}{T} \sum_{t=1}^{T} y_t \\ \hat{\gamma}_j &= \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})(y_{t-j} - \bar{y}) \end{aligned}$$

として、

$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{p-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{p-1} & \hat{\gamma}_{p-2} & \cdots & \hat{\gamma}_0 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_p \end{pmatrix}$$

$$\hat{c} = (1 - \hat{\phi}_1 - \cdots - \hat{\phi}_p) \bar{y}$$

によりモーメント法推定量が求められる。この推定量の性質は、以下の標本モーメントの漸近的性質を用いて導くことができる。

Theorem 8. 標本モーメントの漸近的性質

$\{\epsilon_t\}$ は期待値 *0*、分散 σ^2 の強ホワイトノイズで、$\{y_t\}$ は *(22)* に従う *AR(p)* 過程で、ラグ多項式は安定性条件を満たすものとする。また、$Cov(y_t, y_{t-j}) = \gamma_j$ とする。そのとき、以下が成立する。

(i) $\bar{y} \xrightarrow{p} \mu = c/\Phi(1)$
(ii) $\hat{\gamma}_j \xrightarrow{p} \gamma_j$
(iii) $\sqrt{T}(\bar{y}-\mu) \xrightarrow{d} N(0, \sum_{j=-\infty}^{\infty}\gamma_j)$
更に、$E(\epsilon_t^4) = \eta\sigma^4 < \infty$ として、
(iv) $\sqrt{T}\begin{pmatrix} \hat{\gamma}_1 - \gamma_1 \\ \vdots \\ \hat{\gamma}_p - \gamma_p \end{pmatrix} \xrightarrow{d} N(0, V)$
ただし、V の i, j 要素は

$$V_{ij} = (\eta - 3)\gamma_i\gamma_j + \sum_{p=-\infty}^{\infty}(\gamma_p\gamma_{p-i+j} + \gamma_{p+j}\gamma_{p-i})$$

である。

（証明の概略）
仮定より $\{y_t\}$ が定常エルゴードなので (i), (ii) は簡単である。(iii) は仮定の下で Theorem 5 から明らかである。(iv) は煩雑であるが、Fuller (1976), Thoerem 6.3.5, p.254 に証明が掲載されている。
（３）最尤法
AR(1) の尤度関数は

$$\begin{aligned} logL(c,\phi,\sigma^2|y_1,\cdots,y_T) &= -\frac{T}{2}log(2\pi) - \frac{T}{2}log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - c - \phi y_{t-1})^2 \\ &\quad -\frac{1}{2}log(2\pi) - \frac{1}{2}log\frac{\sigma^2}{1-\phi^2} - \frac{1}{2\sigma^2}(y_1 - \frac{c}{1-\phi})^2 \end{aligned}$$

である。推定量の漸近的性質は、OLS 推定量の性質 Proposition 6 と同じである。詳細は、例えば Hamilton (1994, Time Series Analysis, Section 5.1-5.3, p.118-126)、山本拓「経済時系列の分析」４章を参照。

以上３種類の推定法を紹介したが、$\phi_1, \cdots, \phi_p$ の推定量の漸近的性質はすべて同じである (Proposition 6)。

3.6.3 MA(q) モデルの推定

（１）モーメント法
MA(q) は説明変数を持たないため、AR(p) のような推定はできない。自己共分散を用いたモーメント法を用いることができる。単純化のために期待値が 0 の MA(1) モデル $y_t = \epsilon_t + \theta\epsilon_{t-1}$ の推定を考える。

$$\begin{aligned} \gamma_0 &= \sigma^2(1+\theta^2) \\ \gamma_1 &= \sigma^2\theta \end{aligned}$$

であり、$\rho_1 = \gamma_0/\gamma_1$ は $\theta = 1$ のとき最大値 0.5 をとり、$\theta = -1$ のとき最小値-0.5 をとる。反転可能性を仮定すると、$|\theta| < 1$ より、

$$\theta = \frac{1 - \sqrt{1 - 4\rho_1^2}}{2\rho_1}$$

を得る。ρ_1 を標本自己相関係数 $\hat{\rho}_1 = \hat{\gamma}_1/\hat{\gamma}_0$ でおきかえて

$$\hat{\theta} = \frac{1-\sqrt{1-4\hat{\rho}_1^2}}{2\hat{\rho}_1} \tag{27}$$

により θ を推定することができる。

Theorem 9. *MA(1) のモーメント法推定*

$\{\epsilon_t\}$ は期待値 *0*、分散 σ^2 の強ホワイトノイズであるとする。$y_t = \epsilon_t + \theta\epsilon_{t-1}$ が反転可能な *MA(1)* 過程である（すなわち $|\theta| < 1$）とき、推定量 *(27)* について

$$\begin{array}{ll} (i) & \hat{\theta} \xrightarrow{p} \theta \\ (ii) & \sqrt{T}(\hat{\theta}-\theta) \xrightarrow{d} N(0, 1-\theta^2) \end{array}$$

が成立する。

同様に MA(q) のモーメント法推定も可能で、その性質は Fuller (1976, 8.3 節) を参照のこと。

（２）最尤推定

$\epsilon_t \sim iidN(0,\sigma^2)$ が仮定できる場合は、最尤推定を行うことも可能である。尤度関数は、$Y = (y_1, \cdots, y_T)'$ として、

$$logL(\theta,\sigma^2|Y) = -\frac{T}{2}log(2\pi) - \frac{1}{2}log|\Omega| - \frac{1}{2}Y'\Omega^{-1}Y$$

ただし

$$\Omega = \sigma^2 \begin{pmatrix} 1+\theta^2 & \theta & 0 & \cdots & 0 \\ \theta & 1+\theta^2 & \theta & \cdots & 0 \\ 0 & \theta & 1+\theta^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\theta^2 \end{pmatrix}$$

である。非線形の尤度関数なので、最適化には適当なアルゴリズムを用いる必要がある。MA(1) については、$\sqrt{T}(\hat{\theta}-\theta) \xrightarrow{d} N(0,1-\theta^2)$ となり、モーメント法と同じ漸近分布を持つことがわかる。色々な計算法や簡略化は Fuller (1976, Section 8.3) や Hamilton (1994, Time Series Analysis, Section 5.4-5.5, p.127-131) を参照。その推定量は CAN である。漸近的性質は Brockwell and Davis (1991, Time Series: Theory and Methods, 8 章）を参照。

3.6.4 ARMA(p,q) モデルの推定

（１）IV 推定

ARMA(p,q) 過程は AR 部分をもつために OLS の適用が考えられるが、誤差項（MA 部分）と説明変数の間に相関が生ずるために、一致性をもたない。一つの解決法は IV 法である。$(y_{t-q-1}, y_{t-q-2}, \cdots)$ は MA 部分と相関をもたないので、これを操作変数として IV 推定をすることによって一致推定ができる。例えば ARMA(1,1)

$$y_t = c + \phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$$

の AR 部分のパラメータの IV 推定量は

$$\begin{pmatrix} \hat{c} \\ \hat{\phi} \end{pmatrix} = \left\{ \sum_{t=3}^{T} \begin{pmatrix} 1 & y_{t-1} \\ y_{t-2} & y_{t-2}y_{t-1} \end{pmatrix} \right\}^{-1} \sum_{t=3}^{T} \begin{pmatrix} 1 \\ y_{t-2} \end{pmatrix} y_t$$

である。

Theorem 10. *ARMA(1,1)* の AR パートの IV 推定量の漸近理論

$\{\epsilon_t\}$ は強ホワイトノイズで、AR 部分のラグ多項式が安定性条件を満たすとする。また、

$$A = \begin{pmatrix} (1+\theta)^2 & (1+\theta)^2\mu \\ (1+\theta)^2\mu & (1+\theta^2)(\gamma_0+\mu^2)+2\theta(\gamma_1+\mu^2) \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & \mu \\ \mu & \gamma_1+\mu^2 \end{pmatrix},\ \mu = \frac{c}{1-\phi}$$

とおき、B は正値定符号な対称行列とする。そのとき、

(i) $\begin{pmatrix} \hat{c} \\ \hat{\phi} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} c \\ \phi \end{pmatrix}$

(ii) $\sqrt{T}\begin{pmatrix} \hat{c}-c \\ \hat{\phi}-\phi \end{pmatrix} \xrightarrow{d} N\left(0, \sigma^2 B^{-1}AB^{-1}\right)$

が成立する。

（証明）

通常の IV の場合と同様に

$$\begin{pmatrix} \hat{c} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} c \\ \phi \end{pmatrix} + \left\{\frac{1}{T}\sum_{t=3}^{T}\begin{pmatrix} 1 & y_{t-1} \\ y_{t-2} & y_{t-2}y_{t-1} \end{pmatrix}\right\}^{-1}\frac{1}{T}\sum_{t=3}^{T}\begin{pmatrix} 1 \\ y_{t-2} \end{pmatrix}(\epsilon_t+\theta\epsilon_{t-1}) \tag{28}$$

と書ける。

(i) 条件より、$\{y_t\}$ は MA(∞) 表現をもち、また ϵ_t は強ホワイトノイズである。したがって Proposition 1(d) より、$\{y_t\}$ は強定常でエルゴード性を持つ。従って、大数の法則が成立し、

$$\frac{1}{T}\sum_{t=3}^{T}\begin{pmatrix} 1 & y_{t-1} \\ y_{t-2} & y_{t-2}y_{t-1} \end{pmatrix} \xrightarrow{p} E\begin{pmatrix} 1 & y_{t-1} \\ y_{t-2} & y_{t-2}y_{t-1} \end{pmatrix} = B$$

である。また、

$$\frac{1}{T}\sum_{t=3}^{T}\begin{pmatrix} 1 \\ y_{t-2} \end{pmatrix}(\epsilon_t+\theta\epsilon_{t-1}) = \frac{1}{T}\left\{\begin{pmatrix} 1 \\ y_{T-2} \end{pmatrix}\epsilon_T + \theta\begin{pmatrix} 1 \\ y_1 \end{pmatrix}\epsilon_2\right\} + \frac{1}{T}\sum_{t=3}^{T-1}\begin{pmatrix} 1+\theta \\ y_{t-2}+\theta y_{t-1} \end{pmatrix}\epsilon_t \tag{29}$$

である。右辺第一項は $T\to\infty$ の時に平均二乗の意味で 0 に収束する。また、右辺第二項は Remark 1 より定常でエルゴード性をもつ martingale difference の和になっているため、Theorem 4(a) より

$$\frac{1}{T}\sum_{t=3}^{T-1}\begin{pmatrix} 1+\theta \\ y_{t-2}+\theta y_{t-1} \end{pmatrix}\epsilon_t \xrightarrow{p} 0$$

となる。したがって、(i) が成立する。

(ii) (28), (29) より、

$$\sqrt{T}\begin{pmatrix} \hat{c}-c \\ \hat{\phi}-\phi \end{pmatrix} \approx \left\{\frac{1}{T}\sum_{t=3}^{T}\begin{pmatrix} 1 & y_{t-1} \\ y_{t-2} & y_{t-2}y_{t-1} \end{pmatrix}\right\}^{-1}\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\begin{pmatrix} 1+\theta \\ y_{t-2}+\theta y_{t-1} \end{pmatrix}\epsilon_t$$

である。厳密な等号が成立しないのは、(29) の右辺第一項を落としているためであるが、これは T でなく $\sqrt{T}$ で割っても 0 に収束するため無視できる。$(1+\theta, y_{t-2}+\theta y_{t-1})'\epsilon_t$ は Remark 1 より定常エルゴードな MD で、以下の分散をもつ。

$$Var\left\{\begin{pmatrix} 1+\theta \\ y_{t-2}+\theta y_{t-1} \end{pmatrix}\epsilon_t\right\} = \sigma^2 E\begin{pmatrix} (1+\theta)^2 & (1+\theta)(y_{t-2}+\theta y_{t-1}) \\ (1+\theta)(y_{t-2}+\theta y_{t-1}) & (y_{t-2}+\theta y_{t-1})^2 \end{pmatrix} = A$$

よって、Theorem 7 より

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\left(\begin{array}{c} 1+\theta \\ y_{t-2}+\theta y_{t-1} \end{array}\right)\epsilon_t \xrightarrow{d} N(0,A)$$

従って、(ii) が成立する。（証明終）

　また、MA(q) と同様のモーメント法と最尤法による推定も可能である。
（２）最尤法
　色々な計算法や簡略化は Fuller (1976, Section 8.4) や Hamilton (1994, Time Series Analysis, Section 5.4-5.5, p.127-131) を参照。最尤推定量は CAN であるが、漸近的性質は Brockwell and Davis (1991, Time Series: Theory and Methods, 8 章）を参照。

4　一般化モーメント法（GMM 法）… 導入

4.1　GMM（generalized method of moments）法 … 定義

経済データにおいては、仮定 A1、A3(L)、A4 といった最小二乗法を用いる際の前提条件が満たされていないことも多い。その場合は、それらが満たされていなくても必要な未知パラメータを推定する方法が必要である。それを目指して Hansen(1982, Econometrica) によって提案されたのが GMM 推定法である。これは名前の通りモーメント法を拡張したものであるが、また 2.2 節の IV 推定法の拡張、または minimum distance 推定法の一例と見ることもできる。

4.1.1　モーメント法：復習

モーメント法とは、母集団モーメントと標本モーメントが等しくなるようにパラメータの値を決める推定法である。ある既知の k 次元ベクトル値関数 $m(\cdot,\cdot)$ に対して確率変数 W と k 次元未知パラメータ θ_0 がモーメント条件

$$E\{m(W;\theta_0)\} = 0 \tag{30}$$

を満たしているとする。標本 $\{W_1,\cdots,W_n\}$ が得られたとき、θ_0 のモーメント法推定量 $\hat{\theta}$ は

$$\frac{1}{n}\sum_{i=1}^{n} m(W_i;\hat{\theta}) = 0 \tag{31}$$

により定義される。

<u>例：　IV 推定量</u>

$$y_i = x_i'\beta + \epsilon_i \tag{32}$$

において、モーメント条件

$$E(z_i\epsilon_i) = 0 \tag{33}$$

が成立している時、(32)、(33) から ϵ_i を消去して

$$E\{z_i(y_i - x_i'\beta\} = 0$$

これに対応する標本モーメントと母集団モーメントが等しいとおくと、

$$\frac{1}{n}\sum_{i=1}^{n}\{z_i(y_i - x_i'\beta\} = E\{z_i(y_i - x_i'\beta\} = 0$$

を得る。したがって、

$$b_{IV} = \{\sum_{i=1}^{n} z_i x_i'\}^{-1}\sum_{i=1}^{n} z_i y_i = (Z'X)^{-1}Z'Y$$

となる。

4.1.2　Minimum distance (MD) 推定量

モーメント法は原則としてパラメータの数 $(= k)$ と同数のモーメント条件を用意して推定を行う。パラメータの数より多くのモーメント条件 $(= J > k)$ がモデルから得られることもある。$g(\cdot;\cdot)$ を J 次元の既知のベクトル関数として、

$$E\{g(W;\theta_0)\} = 0$$

というモーメント条件が与えられるとき、これを用いたモーメント法推定量として

$$\frac{1}{n}\sum_{i=1}^{n} g(W_i;\hat{\theta}) = 0$$

を満たす推定量と定義することが考えられる。しかし、この方程式は k 個のパラメータに対して $J(>k)$ 本の方程式があるため、一般には解は存在しない。そこで、できるだけ全てのモーメント条件を満たすように、W を $J\times J$ の正値定符号行列として

$$\hat{\theta}_{MD} = arg\min_{\theta} Q(\theta),\ \ Q(\theta) = \{\frac{1}{n}\sum_{i=1}^{n} g(W_i;\theta)\}'W\{\frac{1}{n}\sum_{i=1}^{n} g(W_i;\theta)\}$$

によって定める推定量を MD 推定量という。

4.1.3 GMM 推定量 ··· 単一方程式の場合

GMM 法は、OLS が正当化されるための仮定 A1,A3(L),A4 が満たされない時に、IV 法、MD 法のアプローチを用いてパラメータを推定する方法である。以下のような状況を考える。既知のスカラー関数 $m(\cdot,\cdot;\cdot)$ に対して、経済変数 y_t, x_t が

$$m(y_t, x_t;\theta) = \epsilon_t$$

という関係にあるとする。θ は k 次元の未知パラメータである。例えば、$m(y_t,x_t;\beta) = y_t - x_t'\beta$ なら、仮定 A1 のモデルと同じになる。IV 法と同様に $E(\epsilon_t|z_t) = 0$ を満たす操作変数 z があるが、z の次元が $J(>k)$ であるとする。そのとき、モーメント条件として J 本の方程式

$$E(z_t\epsilon_t) = 0$$

があり、対応する標本モーメントの条件は

$$\frac{1}{n}\sum_{t=1}^{n} z_t\epsilon_t = \frac{1}{n}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta) = 0$$

となる。MD 法と同様に、J 本すべての式を満たす θ は一般に存在しないので、

$$\hat{\theta}_{GMM} = arg\min_{\theta}\{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta)\}'\hat{W}\{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta)\}$$

によって推定量を定義することが考えられる。$\hat{W}$ は任意の正値定符号な $J\times J$ 行列でよいが、うまく選択することによって効率性を上げることができる。なお、$\hat{W}$ はデータに依存してもよく、その意味でハットがつけられている。

4.1.4 GMM 推定量 ··· 複数本の方程式の場合

前節と同様であるが、既知の p 次元ベクトル値関数 $m(\cdot,\cdot;\cdot) = (m_1(\cdot,\cdot;\cdot),\cdots,m_p(\cdot,\cdot;\cdot))'$ に対して、経済変数 y, x と操作変数 z が

$$m(y_t, x_t;\theta) = \epsilon_t$$

$$E(\epsilon_t|z_t) = 0$$

という関係を満たしているとする。もちろん誤差項も p 次元で、$\epsilon_t = (\epsilon_{1t}, \cdots, \epsilon_{pt})'$ である。このとき、単一方程式の場合と同様に考えて、以下のように GMM 推定量を定義する。各式ごとに含むパラメータが違う可能性を考慮し、$\theta = \theta_1 \cup \theta_2 \cup \cdots \cup \theta_p$、

$$g(y_t, x_t, z_t; \theta) = \begin{pmatrix} z_t m_1(y_t, x_t; \theta_1) \\ z_t m_2(y_t, x_t; \theta_2) \\ \vdots \\ z_t m_p(y_t, x_t; \theta_p) \end{pmatrix}$$

として、

$$\hat{\theta}_{GMM} = arg \min_{\theta} \{\frac{1}{n} \sum_{t=1}^{n} g(y_t, x_t, z_t; \theta)\}' \hat{W} \{\frac{1}{n} \sum_{t=1}^{n} g(y_t, x_t, z_t; \theta)\}$$

ここで、$z_t m_i(y_t, x_t; \theta), i = 1, \cdots, p$ はそれぞれ J 次元ベクトルであり、結果として $g(y_t, x_t, z_t; \theta)$ は Jp 次元ベクトルである。また、$\hat{W}$ は正値定符号な $Jp \times Jp$ 行列である。表現を単純にするために、すべての $m_i(y_t, x_t; \theta)$ に対して同じ z_t が操作変数である場合を考えているが、$m_i(y_t, x_t; \theta)$ と $m_j(y_t, x_t; \theta)$ が異なる操作変数を持つ場合への拡張は容易である。

5 Extremum estimator の漸近理論

GMM 推定量の漸近的性質を調べるために、まず extremum estimator と呼ばれる広いクラスの推定量の大標本特性の一般理論を解説する。GMM 推定量は extremum estimator の特別な場合なので、その漸近論は GMM 推定量にもあてはまる。

5.1 Extremum estimator

データ $\{W_1, \cdots, W_n\}$ に依存する関数 $Q_n(W_1, \cdots, W_n; \theta)$ の最大化（最小化）問題の解として定義される推定量を Extremum estimator という。すなわち、

$$\hat{\theta}_{EE} = \arg\max_{\theta} Q_n(W_1, \cdots, W_n; \theta) \tag{34}$$

例： 以下の推定量は extremum estimator である。

- OLS 推定量

$$Q_n(W_1, \cdots, W_n; \beta) = -\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2$$

- ML 推定量　$f(w;\theta)$ を W_i の密度関数として、

$$Q_n(W_1, \cdots, W_n; \theta) = \frac{1}{n}\sum_{i=1}^{n}\log f(W_i;\theta)$$

- M 推定量　ある既知の関数 $m(\cdot;\cdot)$ に対して、

$$Q_n(W_1, \cdots, W_n; \theta) = \frac{1}{n}\sum_{i=1}^{n} m(W_i;\theta)$$

  ※ OLS, ML は M 推定量の特殊ケースであることは明らかである。

- GMM 推定量

$$Q_n(W_1, \cdots, W_n; \theta) = -\{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\}' W \{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\}$$

以下、表現の簡単化のために $Q_n(W_1, \cdots, W_n; \theta)$ を $Q_n(\theta)$ と書く。

5.2 Extremum estimator の一致性

A. パラメータ空間 Θ がコンパクトなケース
　以下を仮定する。

(i) Θ は R^k のコンパクト集合である。
(ii) $Q_n(\theta)$ は任意のデータ $\{W_1, \cdots, W_n\}$ に対して θ について連続である。
更に、以下を満たす関数 $Q_0(\theta)$ が存在する。
(iii) $Q_0(\theta_0) > Q_0(\theta)$ for $\forall\theta \neq \theta_0$ in Θ
(iv) $\sup_{\theta\in\Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$

Theorem 11. *A(i),(ii),(iii),(iv)* が成り立つ時、

$$\hat{\theta}_{EE} \xrightarrow{p} \theta_0$$

この定理の仮定 A(i) は制約として強すぎるかもしれない。これを緩める場合は、以下の仮定の下で同じ結果が成立する。

B. パラメータ空間 Θ がコンパクトでないケース

(i)' Θ は R^k のコンパクト集合ではない。また、θ_0 は Θ の内点である。

(ii)' $Q_n(\theta)$ は任意のデータ $\{W_1, \cdots, W_n\}$ に対して θ について concave である。

更に、以下を満たす関数 $Q_0(\theta)$ が存在する。

(iii) $Q_0(\theta_0) > Q_0(\theta)$ for $\forall\theta \neq \theta_0$ in Θ

(iv)' $Q_n(\theta) \xrightarrow{p} Q_0(\theta)$ for $\forall\theta \in \Theta$

Theorem 12. *B(i)',(ii)',(iii),(iv)'* が成り立つ時、

$$\hat{\theta}_{EE} \xrightarrow{p} \theta_0$$

<u>定理 3 の証明</u>

任意に $\eta > 0$ を定めて

$$C = \{\theta : |\theta - \theta_0| < \eta\}$$

$$\bar{C} = \{\theta : |\theta - \theta_0| \geq \eta\}$$

とする。$\bar{C} \cap \Theta$ はコンパクトなので、$\max_{\theta \in \bar{C} \cap \Theta} Q_0(\theta)$ が存在する。$\epsilon = Q_0(\theta_0) - \max_{\theta \in \bar{C} \cap \Theta} Q_0(\theta)$ とおく。

$$P(|Q_n(\theta) - Q_0(\theta)| < \frac{\epsilon}{2} \text{ for } \forall\theta \in \Theta)$$

$$\begin{aligned}
&= P(Q_n(\theta) > Q_0(\theta) - \frac{\epsilon}{2} \text{ and } Q_0(\theta) > Q_n(\theta) - \frac{\epsilon}{2} \text{ for } \forall\theta \in \Theta) \\
&\leq P(Q_n(\theta_0) > Q_0(\theta_0) - \frac{\epsilon}{2} \text{ and } Q_0(\hat{\theta}_{EE}) > Q_n(\hat{\theta}_{EE}) - \frac{\epsilon}{2})
\end{aligned}$$

ここで、$Q_n(\hat{\theta}_{EE}) = \max_{\theta \in \Theta} Q_n(\theta) \geq Q_n(\theta_0)$ なので、

$$\begin{aligned}
&\leq P(Q_n(\theta_0) > Q_0(\theta_0) - \frac{\epsilon}{2} \text{ and } Q_0(\hat{\theta}_{EE}) > Q_n(\theta_0) - \frac{\epsilon}{2}) \\
&\leq P(Q_0(\hat{\theta}_{EE}) > Q_0(\theta_0) - \epsilon) \\
&= P(Q_0(\hat{\theta}_{EE}) > \max_{\theta \in \bar{C} \cap \Theta} Q_0(\theta)) \\
&= P(\hat{\theta}_{EE} \notin \bar{C} \cap \Theta) = P(\hat{\theta}_{EE} \in C)
\end{aligned}$$

両辺の極限を取ると

$$\lim_{n \to \infty} P(|Q_n(\theta) - Q_0(\theta)| < \frac{\epsilon}{2} \text{ for } \forall\theta \in \Theta) \leq \lim_{n \to \infty} P(|\hat{\theta}_{EE} - \theta_0| < \eta)$$

仮定 A(iv) より、左辺は 1 に収束する。したがって、

$$\lim_{n \to \infty} P(|\hat{\theta}_{EE} - \theta_0| < \eta) \to 1$$

（証明終）

(※１) 定理 4 の証明は Newey and McFadden (Handbook of Econometrics, 1994, p.2133-2134) を参照のこと。

（※２）A(iv) が満たされているかを直接調べることは通常難しい。しかし、M 推定量については、以下 Lemma 3 の十分条件が知られている。目的関数は

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} m(W_t;\theta)$$

で、$m(\cdot,\cdot)$ は既知のベクトル値関数である。

Lemma 3. *(Lemma 7.2, Hayashi(2000), P.459)*
W_t はエルゴード性をもつ強定常過程であるとする。
(i) Θ はコンパクトである。
(ii) $m(W_t;\theta)$ はすべての W_t について θ に関して連続である。
(iii) すべての $\theta \in \Theta$ について $||m(W_t;\theta)|| < d(W_t)$、かつ $E\{d(W_t)\} < \infty$ を満たす関数 $d(\cdot)$ が存在する。
そのとき、

$$\sup_{\theta\in\Theta} ||\frac{1}{n}\sum_{t=1}^{n} m(W_t;\theta) - E\{m(W_t;\theta)\}|| \xrightarrow{p} 0$$

であり、かつ $E\{m(W_t;\theta)\}$ は θ について連続である。

（※） (iii) の条件は $E[\sup_{\theta\in\Theta} ||m(W_t;\theta)||] < \infty$ で置き換えられる。

5.3 Extremum estimator の漸近正規性

本節では、Extremum estimator に関する中心極限定理を扱う。前述のとおり、Extremum estimator は M 推定量（従って OLS, ML 推定量）と GMM 推定量を含む広いクラスの推定量である。$Q_n(\theta)$ が θ に関して滑らかで、θ_0 が Θ の内点であれば、Extremum estimator (34) は

$$\frac{\partial Q_n(\hat{\theta}_{EE})}{\partial\theta} = 0, \quad \frac{\partial^2 Q_n(\hat{\theta}_{EE})}{\partial\theta\partial\theta'} < 0 \tag{35}$$

を満たす。Theorem 12 の仮定が満たされていて $Q_n(\theta)$ が θ に関して大域的に凹関数であれば、(35) の解は一意であるが、Theorem 11 の仮定の下では n が十分大きい場合にも (35) を満たす解が複数存在する可能性がある。後者の場合、複数の解のうち、Q_n を最大にするものを選ぶものとする。以下を仮定する。

Theorem 13. *Extremum estimator* の漸近正規性
(i) Theorem 11 または *12* の条件が成立しており、従って $\hat{\theta}_{EE}$ は一致推定量である。
(ii) θ_0 は Θ の内点である。
(iii) $Q_n(\theta)$ は任意の $\{W_1,\cdots,W_n\}$ に対して θ_0 の近傍で θ について 2 回連続微分可能である。
(iv) $\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, B(\theta_0))$, $\quad B(\theta_0) = \lim E\{n\frac{\partial Q_n(\theta_0)}{\partial\theta}\frac{\partial Q_n(\theta_0)}{\partial\theta'}\}$
(v) $\bar{\theta}_n \xrightarrow{p} \theta_0$ を満たすすべての $\bar{\theta}_n$ について、$\frac{\partial^2 Q_n(\bar{\theta}_n)}{\partial\theta\partial\theta'} \xrightarrow{p} \lim E\{\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\} = A(\theta_0)$ が成立し、$A(\theta_0)$ は有界で正則である。

そのとき、

$$\sqrt{n}(\hat{\theta}_{EE} - \theta_0) \xrightarrow{d} N(0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1})$$

が成立する。

（証明）

(ii)、(iii) より、(34) は

$$\frac{\partial Q_n(\hat{\theta}_{EE})}{\partial \theta} = 0$$

を満たす。平均値の定理より、

$$0 = \frac{\partial Q_n(\hat{\theta}_{EE})}{\partial \theta} = \frac{\partial Q_n(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta \partial \theta'}(\hat{\theta}_{EE} - \theta_0)$$

を満たす $\bar{\theta}$ が存在し、$\bar{\theta} = \lambda \hat{\theta}_{EE} + (1-\lambda)\theta_0$ で $\lambda \in [0,1]$ である。これを $\hat{\theta}_{EE} - \theta_0$ について解くと

$$\sqrt{n}(\hat{\theta}_{EE} - \theta_0) = -\{\frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta \partial \theta'}\}^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta}$$

仮定 (iv), (v) より結果を得る。

（証明終）

仮定 (v) を直接確かめるのは通常難しいが、次の Lemma 4 を使って確かめることができる。すなわち n に依存する関数 $g_n(\theta)$ ($\stackrel{p}{\rightarrow} g(\theta)$) について、$\hat{\theta} \stackrel{p}{\rightarrow} \theta_0$ を満たす $\hat{\theta}$ に対して、$g_n(\hat{\theta}) \stackrel{p}{\rightarrow} g(\theta_0)$ が成立する条件を与える。

Lemma 4. （*Theorem 4.1.5, Amemiya (1985), p.113*）

θ_0 の近傍を $N(\theta_0)$ とする。

(i) $\hat{\theta} \stackrel{p}{\rightarrow} \theta_0$

(ii) $\sup_{\theta \in N(\theta_0)} |g_n(\theta) - g(\theta)| \stackrel{p}{\rightarrow} 0$

(iii) $g(\theta)$ は $\theta = \theta_0$ で連続である。

これらの仮定の下で、

$$g_n(\hat{\theta}) \stackrel{p}{\rightarrow} g(\theta_0)$$

である。

（証明）　任意の $\epsilon > 0$ に対して

$$\begin{aligned} P(|g_n(\hat{\theta}) - g(\theta_0)| \geq \epsilon) &\leq P(|g_n(\hat{\theta}) - g(\hat{\theta})| + |g(\hat{\theta}) - g(\theta_0)| \geq \epsilon) \\ &\leq P(|g_n(\hat{\theta}) - g(\hat{\theta})| \geq \frac{\epsilon}{2}) + P(|g(\hat{\theta}) - g(\theta_0)| \geq \frac{\epsilon}{2}) \end{aligned} \tag{36}$$

である。(i), (ii) より、任意の $\epsilon > 0$、$\delta > 0$ に対して、ある n_0 が存在して、

$$P\{\hat{\theta} \notin N(\theta_0)\} < \frac{\delta}{4}$$

$$P\{\sup_{\theta \in N(\theta_0)} |g_n(\theta) - g(\theta)| \geq \frac{\epsilon}{2}\} < \frac{\delta}{4}$$

がすべての $n > n_0$ について成立する。これらを用いると、

$$\begin{aligned} & P(|g_n(\hat{\theta}) - g(\hat{\theta})| \geq \frac{\epsilon}{2}) \\ = \;& P\{|g_n(\hat{\theta}) - g(\hat{\theta})| \geq \frac{\epsilon}{2}, \hat{\theta} \in N(\theta_0)\} + P\{|g_n(\hat{\theta}) - g(\hat{\theta})| \geq \frac{\epsilon}{2}, \hat{\theta} \notin N(\theta_0)\} \\ \leq \;& P\{\sup_{\theta \in N(\theta_0)} |g_n(\theta) - g(\theta)| \geq \frac{\epsilon}{2}\} + P\{\hat{\theta} \notin N(\theta_0)\} \\ \leq \;& \frac{\delta}{2} \end{aligned} \tag{37}$$

となる。仮定 (i), (iii) よりある n_1 が存在して

$$P(|g(\hat{\theta}) - g(\theta_0)| \geq \frac{\epsilon}{2}) < \frac{\delta}{2} \tag{38}$$

がすべての $n > n_1$ について成立する。(37)、(38) から、すべての $n > max(n_0, n_1)$ について

$$P(|g_n(\hat{\theta}) - g(\hat{\theta})| \geq \frac{\epsilon}{2}) + P(|g(\hat{\theta}) - g(\theta_0)| \geq \frac{\epsilon}{2}) < \delta \tag{39}$$

となる。従って、(39)、(36) から

$$P(|g_n(\hat{\theta}) - g(\theta_0)| \geq \epsilon) < \delta$$

（証明終）

5.4 M 推定量の一致性と漸近正規性

本章の目的は Extremum estimator に関する漸近論を紹介することであるが、抽象度が高いので、Extremum estimator の一例である M 推定量に対してその結果を適用してみよう。本題である GMM 推定量に関しては次章で扱う。なお、この節で紹介する M 推定量の結果は、データがエルゴード性をもつ定常過程である場合の LS 推定量に当てはまるため、それ自体にも意味がある。繰り返しになるが、M 推定量はある既知の関数 m について

$$\hat{\theta}_M = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} m(W_t; \theta) \tag{40}$$

によって定義される。

5.4.1 M 推定量の一致性

推定量の一致性の証明には、Theorem 11 （パラメータ空間がコンパクトな場合）または Theorem 12 （パラメータ空間がコンパクトでない場合）の結果を用いる。それらに対応して、以下の 2 つの定理が証明できる。W_t はエルゴード性を持つ定常過程であるとする。

Theorem 14. パラメータ空間がコンパクトな場合（*Proposition 7.3, Hayashi (2000), p.459*）

以下の *(i)-(iv)* が成立すると仮定する。。

(i) パラメータ空間 Θ は R^k のコンパクトな部分集合である。

(ii) $m(W_t; \theta)$ はすべての W_t に対して θ について連続である。

(iii) $Em(W_t; \theta_0) > Em(W_t; \theta)$ *for any* $\theta (\in \Theta) \neq \theta_0$

(iv) $E\{\sup_{\theta \in \Theta} |m(W_t; \theta)|\} < \infty$

そのとき、

$$\hat{\theta}_M \xrightarrow{p} \theta_0$$

Lemma 3 より、(i), (ii), (iv) の仮定のもとで目的関数が $E\{m(W_t; \theta)\}$ に一様確率収束することが保証される。他の条件は Theorem 11 の条件を M 推定量の枠組みに書き直したものである。Theorem 12 に対応して、以下が証明される。

Theorem 15. パラメータ空間がコンパクトでない場合（*Proposition 7.4, Hayashi (2000), p.460*）

以下の *(i)',(ii)',(iii),(iv)'* が成立すると仮定する。

(i)' パラメータ空間 Θ はコンパクトでない R^k の部分集合である。また、θ_0 は Θ の内点である。

(ii)' $m(W_t;\theta)$ はすべての W_t に対して θ について凹関数である。
(iii) $Em(W_t;\theta_0) > Em(W_t;\theta)$ *for any* $\theta(\in\Theta) \neq \theta_0$
(iv)' $E|m(W_t;\theta)| < \infty$ *for any* $\theta \in \Theta$
そのとき、

$$\hat{\theta}_M \xrightarrow{p} \theta_0$$

Theorem 14, 15 の証明は、それぞれ Theorem 11, 12 の条件を M 推定量の目的関数に対して調べればよい。なお、Theorem 15 において、元のモデルそのものは仮定 (ii)' を満たさない場合でも、１対１の連続関数 g に対して $\tau = g(\theta)$ と変換した時に $\tilde{m}(W_t;\tau) = m(W_t;g^{-1}(\tau))$ が τ について凹関数であれば、$\tilde{m}$ を用いた τ の M 推定量 $\hat{\tau}_M$ は Theorem 15 より一致性をもつ。それを用いて $\hat{\theta}_M = g^{-1}(\hat{\tau}_M)$ を構成すれば θ の一致推定量が得られる。（例：iid$N(0,\sigma^2)$ の下での ML 推定量、cf. Hayashi (2000), Example 7.7, p.461）

5.4.2　M 推定量の漸近正規性

この節では、M 推定量の中心極限定理を示す。

Theorem 16. *M* 推定量の漸近正規性（*Proposition 7.8, Hayashi (2000), p.472*）
W_t はエルゴード性を持つ強定常過程であるとする。また、以下の条件が成立しているものとする。
(i) Theorem 14 または *15* の条件が成立しており、従って $\hat{\theta}_M$ は一致推定量である。
(ii) θ_0 は Θ の内点である。
(iii) $m(W_t;\theta)$ は任意の W_t に対して θ について２回連続微分可能である。
(iv) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial m(W_t;\theta_0)}{\partial\theta} \xrightarrow{d} N(0,\Sigma)$、$\Sigma$ は正値定符号行列である。
(v) $\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 m(W_t;\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} E(\frac{\partial^2 m(W_t;\theta_0)}{\partial\theta\partial\theta'}) = H$ で *(*エルゴード性より*)*、H は有界で正則である。
(vi) θ_0 のある近傍を $N(\theta_0)$ として、$E\{\sup_{\theta\in N(\theta_0)}||\frac{\partial^2 m(W_t;\theta)}{\partial\theta\partial\theta'}||\} < \infty$
そのとき、

$$\sqrt{n}(\hat{\theta}_M - \theta_0) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1'})$$

である。

（証明）　(ii)、(iii) より、(40) の解は

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial m(W_t;\hat{\theta}_M)}{\partial\theta} = 0$$

を満たす。仮定 (iii) と平均値の定理より、

$$\frac{\partial m(W_t;\hat{\theta}_M)}{\partial\theta} = \frac{\partial m(W_t;\theta_0)}{\partial\theta} + \frac{\partial^2 m(W_t;\bar{\theta})}{\partial\theta\partial\theta'}(\hat{\theta}_M - \theta_0)$$

を満たす $\bar{\theta}$ が存在し、ある $\lambda \in [0,1]$ に対して $\bar{\theta} = \lambda\hat{\theta}_M + (1-\lambda)\theta_0$ と書ける。Lemma 3 でパラメータ空間を $N(\theta_0)$ の閉包に変更したものを考えると、(ii), (iii),(vi) から、

$$\sup_{\theta\in N(\theta_0)}||\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 m(W_t;\theta)}{\partial\theta\partial\theta'} - E\frac{\partial^2 m(W_t;\theta)}{\partial\theta\partial\theta'}|| \xrightarrow{p} 0 \tag{41}$$

となる。$\hat{\theta}_M$ の一致性より $\bar{\theta} \xrightarrow{p} \theta_0$、(44)、(iii) の下で Lemma 4 を適用すると

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 m(W_t;\bar{\theta})}{\partial\theta\partial\theta'} \xrightarrow{p} E\frac{\partial^2 m(W_t;\theta_0)}{\partial\theta\partial\theta'} = H$$

である。以上の結果と仮定 (iv) から、

$$\begin{aligned}\sqrt{n}(\hat{\theta}_M - \theta_0) &= -\{\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 m(W_t;\bar{\theta})}{\partial\theta\partial\theta'}\}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial m(W_t;\theta_0)}{\partial\theta} \\ &\overset{d}{\to} N(0, H^{-1}\Sigma H^{-1'})\end{aligned}$$

（証明終）

（※１） データが iid の場合には、$\Sigma = E\{\frac{\partial m(W_t;\theta_0)}{\partial\theta}\frac{\partial m(W_t;\theta_0)}{\partial\theta'}\}$ であり、また定常過程なら、Σ は $\frac{\partial m(W_t;\theta_0)}{\partial\theta}$ の long run variance、つまり $\lim_{n\to\infty} Var\{n^{-1/2}\sum_{t=1}^{n}\frac{\partial m(W_t;\theta_0)}{\partial\theta}\}$ である。

（※２） エルゴード性を持つ定常過程に関して仮定 (iv) の中心極限定理が成立するための十分条件としては Gordin の条件（Hayashi (2000), p.402-405）や Scott の条件（White(1999), Asymptotic Theory for Econometricians, p.125）等がある。更に、MD（martingale difference）の条件を課す場合については、Billingsley の条件（Hayashi(2000), p.106) がある。しかし、実際のところはいずれの条件も成立しているか否かをデータに照らして調べることは事実上不可能であり、単に仮定されるのが普通である。

（※３） (v) の収束は定常エルゴード性から成立する結果であり、その仮定の主たる部分は $E\frac{\partial^2 m(W_t;\theta_0)}{\partial\theta\partial\theta'}$ が有界で正則という点である。

6 GMM 推定量の漸近理論とそれに基づく検定

4.1.3 節と同じく以下のモデルを想定する。

$$m(y_t, x_t; \theta_0) = \epsilon_t,\ t = 1, \cdots, n$$

$$E\{m(y_t, x_t; \theta_0)|z_t\} = 0$$

このモデルに対する GMM 推定量は

$$\hat{\theta}_{GMM} = \arg\min_{\theta} Q_n(\theta) \tag{42}$$

$$Q_n(\theta) = \{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\}'\hat{W}\{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\}$$

によって定義される。

6.1 一致性

一致性の証明のために以下の仮定をおく。これは Theorem 11（パラメータ空間がコンパクトである場合）に対応するものである。$\{y_t, x_t, z_t\}$ はエルゴード性をもつ強定常過程であるとする。

A1: $m(y_t, x_t; \theta) = \epsilon_t,\ t = 1, \cdots, n$

A2: $E\{m(y_t, x_t; \theta_0)|z_t\} = 0$

A3: $\theta_0 \in \Theta$、Θ は R^p のコンパクト集合である。

A4: $m(y_t, x_t; \theta)$ は任意の y_t, x_t に対して θ に関して連続である。

更に、以下を満たす関数 $Q_0(\theta)$ が存在する。

A5: （識別条件）$\theta \neq \theta_0$ なら $E\{z_t m(y_t, x_t; \theta)\} \neq 0$ である。

A6: （dominance 条件） $E\{\sup_{\theta\in\Theta} ||z_t m(y_t, x_t; \theta)||\} < \infty$

A7: $\hat{W} \overset{p}{\to} W$、$W$ は対称な正値定符号行列

Theorem 17. *(Proposition 7.7, Hayashi (2000), p.467)* 仮定 *A1-A7* の下で、

$$\hat{\theta}_{GMM} \xrightarrow{p} \theta_0$$

パラメータ空間のコンパクト性の仮定 A3 が成立しない場合は、Theorem 12 を適用するために A3, A4, A6 を以下で置き換える。

A3': $\theta_0 \in \Theta$、Θ は R^p のコンパクト集合ではない。

A4': $Q_n(\theta)$ は任意のデータ $\{W_1, \cdots, W_n\}$ に対して θ について concave である。

A6': $||E\{z_t m(y_t, x_t; \theta)\}|| < \infty$ for $\forall \theta \in \Theta$

Theorem 18. 仮定 *A1, A2, A3',A4',A5, A6', A7* の下で、

$$\hat{\theta}_{GMM} \xrightarrow{p} \theta_0$$

Theorem 17,18 の証明は、それぞれ Theorem 11, 12 の条件を GMM の目的関数に合わせて調べればよい。

6.2 漸近正規性

中心極限定理のために、Theorem 17 または 18 の仮定に加えて以下を仮定する。

A8: θ_0 は Θ の内点である。

A9: $m(y_t, x_t; \theta)$ は任意の y_t, x_t に対して θ について連続微分可能である。

A10: $\frac{1}{\sqrt{n}} \sum_{t=1}^n z_t m(y_t, x_t; \theta_0) \xrightarrow{d} N(0, S)$、$S$ は正値定符号行列である。

A11: θ_0 の近傍 $N(\theta_0)$ に対して

$$E\{\sup_{\theta \in N(\theta_0)} ||\frac{\partial z_t m(y_t, x_t; \theta)}{\partial \theta'}||\} < \infty$$

A12: $G = E\{\frac{\partial z_t m(y_t, x_t; \theta_0)}{\partial \theta'}\}$ として、$rank(G) = k$

A13: $\hat{S} \xrightarrow{p} S$

Theorem 19. *GMM* 推定量の漸近正規性

Theorem 17 または *18* の条件に加えて、*A8-A12* が成立するとき、

(a) $\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(0, \Omega)$、ただし $\Omega = (G'WG)^{-1}G'WSWG(G'WG)^{-1}$

更に *A13* を加えると

(b) $\hat{\Omega} = (\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\hat{S}\hat{W}\hat{G}(\hat{G}'\hat{W}\hat{G})^{-1} \xrightarrow{p} \Omega$

ただし $\hat{G} = \frac{1}{n}\sum_{t=1}^n \frac{\partial z_t m(y_t, x_t; \hat{\theta}_{GMM})}{\partial \theta'}$。

(証明)

A8、A9 より、(42) の解は

$$2\{\frac{1}{n}\sum_{t=1}^n \frac{z_t \partial m(y_t, x_t; \hat{\theta}_{GMM})}{\partial \theta'}\}'\hat{W}\{\frac{1}{n}\sum_{t=1}^n z_t m(y_t, x_t; \hat{\theta}_{GMM})\} = 0$$

を満たす。A9 と平均値の定理より、

$$\frac{1}{n}\sum_{t=1}^n z_t m(y_t, x_t; \hat{\theta}_{GMM}) = \frac{1}{n}\sum_{t=1}^n z_t m(y_t, x_t; \theta_0) + \frac{1}{n}\sum_{t=1}^n z_t \frac{\partial m(y_t, x_t; \bar{\theta})}{\partial \theta'}(\hat{\theta}_{GMM} - \theta_0)$$

を満たす $\bar{\theta}$ が存在し、ある $\lambda \in [0,1]$ に対して $\bar{\theta} = \lambda\hat{\theta}_{GMM} + (1-\lambda)\theta_0$ と書ける。従って

$$\begin{aligned} &\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \\ = \quad & -\left[\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t, x_t; \hat{\theta}_{GMM})}{\partial\theta'}\}'\hat{W}\{\frac{1}{n}\sum_{t=1}^{n} z_t\frac{\partial m(y_t, x_t; \bar{\theta})}{\partial\theta'}\}\right]^{-1} \\ & \times\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t, x_t; \hat{\theta}_{GMM})}{\partial\theta'}\}'\hat{W}\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta_0)\} \end{aligned} \tag{43}$$

Lemma 3 でパラメータ空間を $N(\theta_0)$ の閉包に変更したものを考えると、A8, A9, A11 から、

$$\sup_{\theta\in N(\theta_0)} ||\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t, x_t; \theta)}{\partial\theta'} - E\frac{z_t\partial m(y_t, x_t; \theta)}{\partial\theta'}|| \xrightarrow{p} 0 \tag{44}$$

となる。$\hat{\theta}_{GMM}$ の一致性より $\bar{\theta} \xrightarrow{p} \theta_0$、(44)、A9 が成立しているため Lemma 4 が適用できて、

$$\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t, x_t; \hat{\theta}_{GMM})}{\partial\theta'} \xrightarrow{p} E\frac{z_t\partial m(y_t, x_t; \theta_0)}{\partial\theta'} = G \tag{45}$$

$$\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t, x_t; \bar{\theta})}{\partial\theta'} \xrightarrow{p} E\frac{z_t\partial m(y_t, x_t; \theta_0)}{\partial\theta'} = G \tag{46}$$

である。(25), (27), (28), $\hat{W} \xrightarrow{p} W$ (Theorem 17 または 18 の A7 より), 仮定 A10 から、

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \quad \xrightarrow{d} \quad N(0, (G'WG)^{-1}G'WSWG(G'WG)^{-1})$$

(b) は $\hat{W} \xrightarrow{p} W$, A13, (27) より明らか。（証明終）

（※１） $\hat{S}$ としては次のような推定量が考えられる。$\hat{g}_t = z_t m(y_t, x_t; \hat{\theta}_{GMM})$ とする。もしも標本が iid なら、単純に

$$\hat{S} = \frac{1}{n}\sum_{t=1}^{n}\hat{g}_t\hat{g}_t'$$

でよい。自己相関があるなら、例えば Newey and West (1987) の HAC(Heteroscedasticity and Autocorrelation Consistent) 推定量を用いることができる（Hayashi (2000), p.409）。

$$\hat{S} = \hat{\Gamma}_0 + \sum_{s=1}^{q_n}(1 - \frac{s}{q_n+1})(\hat{\Gamma}_s + \hat{\Gamma}_s')$$

ただし $\hat{\Gamma}_s = \frac{1}{n}\sum_{t=s+1}^{n}\hat{g}_t\hat{g}_{t-s}'$ で、q_n は n よりもゆっくり発散していくように分析者が決める定数である。

（※２） $W = S^{-1}$ のとき、漸近分散 Ω は最小になる。これを efficient GMM という。

（※３） 複数方程式に対する GMM 推定の場合は、上の $z_t m(y_t, x_t; \theta)$ を

$$g(y_t, x_t, z_t; \theta) = \begin{pmatrix} z_t m_1(y_t, x_t; \theta_1) \\ z_t m_2(y_t, x_t; \theta_2) \\ \vdots \\ z_t m_p(y_t, x_t; \theta_p) \end{pmatrix}$$

で置き換えればよい。

6.3 Efficient GMM 推定

前節に述べたように、$W = S^{-1}$ のとき、GMM 推定量の漸近分散 Ω は最小になる。つまり、任意の対称な正値定符号行列 W に対して、

$$(G'WG)^{-1}G'WSWG(G'WG)^{-1} \geq (G'S^{-1}G)^{-1}$$

が成立する。従って、ウェイト行列を $W = S^{-1}$ とおいて推定を行うのが良いが、実際には S は未知なので、実行不可能である。そこで、以下の二つの手法が提案されている。

(1) 2 step GMM

１段階目に $W = I$ として GMM 推定を行い、$\hat{\theta}^{(1)}_{GMM}$ を得る。それを用いて Newey-West HAC 推定等によって $\hat{S}$ を得る。２段階目に $\hat{S}^{-1}$ をウェイト行列に使って $\hat{\theta}^{(2)}_{GMM}$ を得る。これは、漸近分散 $(G'S^{-1}G)^{-1}$ を達成する efficient GMM である。

(2) Continuous updating GMM

θ が未知であるために S の推定ができないわけであるが、θ を未知のまま S の「推定量」を構成することは可能である。たとえば、iid の標本であれば、

$$\hat{S}(\theta) = \frac{1}{n}\sum_{t=1}^{n} z_t z_t' m(y_t, x_t; \theta)^2$$

とすればよい。Newey-West 推定量は形が表現が複雑なので省略するが、同様である。Contiunous updating estimator は、これを用いて、

$$\hat{\theta}_{CUE} = \arg\min_{\theta}\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\}'\hat{S}(\theta)^{-1}\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\}$$

によって定義されるもので、漸近分散 $(G'S^{-1}G)^{-1}$ を達成する efficient GMM である。

この二つの推定量について、以下の結果を得る。

Corollary 1. *Efficient GMM* 推定量の漸近正規性

Theorem 19 の条件を仮定する。ただし、$\hat{W}$ には $\hat{S}^{-1}$ を用いて、上のいずれかの方法による効率的 *GMM* 推定を考えるものとする。そのとき、

(a) $\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(0, \Omega)$、ただし $\Omega = (G'S^{-1}G)^{-1}$

(b) $\hat{\Omega} = (\hat{G}'\hat{S}^{-1}\hat{G})^{-1} \xrightarrow{p} \Omega$

一般に、2 step GMM の方が計算が簡単であるが、バイアスが大きいことが知られている (Hansen, Heaton and Yaron (1996, Journal of Business and Economic Statistics)、Newey and Smith (2004, Econometrica))。

6.4 GMM に関わる検定

6.4.1 節では、パラメータに関する仮説が正しいか否かの検定法を説明する。ML 法の枠組みでは、ワルド検定、ラグランジュ乗数検定、尤度比検定という３つの同等な大標本検定法があるが、GMM の枠組みでも原理的にそれらと同様にして検定法を構成できる。6.4.2 節で、GMM 推定に用いるモーメント条件が正しいかどうかを確かめる検定法を扱う。それは J テストと呼ばれ、過剰識別性 (over-identification) の検定ともいう。通常わざわざ非効率な GMM 推定量を使うことはないので、前節で示した efficient GMM 推定量を用いるものとする。

6.4.1　パラメータに関する検定（Hayashi (2000), Section 7.4, p.487-497）

$a(\theta)$ を $r(<k)$ 次元の連続微分可能なベクトル値関数とし、以下の検定を考える。

$$\begin{aligned} H_0: \ a(\theta_0) &= 0 \\ H_1: \ a(\theta_0) &\neq 0 \end{aligned}$$

また、$A(\theta) = \partial a(\theta)/\partial\theta'$、$A_0 = A(\theta_0)$ とする。

　ワルド検定 (W) では、帰無仮説の制約なしの推定値が制約を満たしているかどうかを調べる。ラグランジュ乗数 (LM) 検定では帰無仮説を制約として推定を行ったときに、その制約が効いているかどうかを調べる。尤度比検定 (LR) では制約つき推定値と制約なし推定値で目的関数が達成する最大値に違いがあるかどうかを調べる。LM と LR では制約付きの GMM 推定量 $\tilde{\theta}_{GMM}$ が必要になる。それは以下の解として定義される。

$$\min_{\theta\in\Theta}\{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\}'\tilde{S}^{-1}\{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\} \text{ s.t. } 2a(\theta) = 0 \tag{47}$$

すなわち $\Theta_1 = \Theta \cap \{\theta : a(\theta) = 0\}$ として

$$\tilde{\theta} = \arg\min_{\theta\in\Theta_1}\{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\}'\tilde{S}^{-1}\{\frac{1}{n}\sum_{t=1}^{n} z_t m(y_t, x_t; \theta)\}$$

なお、制約なしの推定量は前節までと同じく $\hat{\theta}_{GMM}$ とする。以下、表現を簡潔にするために $\hat{\theta}_{GMM}, \tilde{\theta}_{GMM}$ を $\hat{\theta}, \tilde{\theta}$ と書くことにする。$\hat{\theta}$ は常に一致推定量であるが、$\tilde{\theta}$ は帰無仮説（制約 $a(\theta_0) = 0$）が正しいときのみ一致性をもつ。つまり、

$$\theta_1 = \arg\min_{\theta\in\Theta_1} E\{z_t m(y_t, x_t; \theta)\}'S^{-1}E\{z_t m(y_t, x_t; \theta)\}$$

として、帰無仮説が正しいときは $\theta_1 = \theta_0$、対立仮説が正しい時は $\theta_1 \neq \theta_0$ であるから

$$\hat{\theta} \xrightarrow{p} \theta_0 \quad \text{under } H_0 \,\&\, H_1$$

$$\begin{aligned} \tilde{\theta} \ &\xrightarrow{p} \ \theta_0 \ \text{ under } H_0 \\ &\xrightarrow{p} \ \theta_1(\neq \theta_0) \ \text{ under } H_1 \end{aligned}$$

となる。

　(1) ワルド検定

　もし帰無仮説が正しければ、H_0 の制約なし推定量 $\hat{\theta}$ は制約をほぼ満たすはずである。そこで、ワルド検定では $a(\hat{\theta}) \approx 0$ かどうかを調べる。検定統計量は以下で定義される。$\hat{S}$ を S の一致推定量として、

$$T_W = na(\hat{\theta})'[A(\hat{\theta})(\hat{G}'\hat{S}^{-1}\hat{G})^{-1}A(\hat{\theta})']^{-1}a(\hat{\theta})$$

この統計量の構成は、k 次元正規確率変数ベクトル u でその分散行列 $V(u)$ の逆行列をはさむ 2 次形式 $u'\{V(u)\}^{-1}u$ を作ると、それは自由度 k のカイ二乗分布に従うという性質を用いたものである。

Theorem 20. ワルド検定

　(i) Theorem 19 の仮定が成立する。

　(ii) $a(\theta)$ は $r(<k)$ 次元の連続微分可能なベクトル値関数とする。

　(iii) A_0 は *full row rank* である。

(i)-(iii) の仮定が成立するとき、

$$T_W \quad \xrightarrow{d} \quad \chi_r^2 \;\; under\; H_0$$

$$\frac{T_W}{n} \xrightarrow{p} a(\theta_0)'\{A_0(G'S^{-1}G)^{-1}A_0'\}^{-1}a(\theta_0) = const.(>0) \;\; under\; H_1$$

である。

（証明）

$a(\theta)$ は連続微分可能なので、平均値の定理から

$$\sqrt{n}\{a(\hat{\theta}) - a(\theta_0)\} = A(\bar{\theta})\sqrt{n}(\hat{\theta} - \theta_0)$$

かつ

$$A(\bar{\theta}) \xrightarrow{p} A(\theta_0)$$

である。従って、Corollary 1 の結果と合わせると

$$\sqrt{n}\{a(\hat{\theta}) - a(\theta_0)\} \xrightarrow{d} N(0, A_0(G'S^{-1}G)^{-1}A_0') \tag{48}$$

(27) より $\hat{G} \xrightarrow{p} G$、Theorem 19 の A13 より $\hat{S} \xrightarrow{p} S$、(ii) より $A(\hat{\theta}) \xrightarrow{p} A_0$ なので、

$$A(\hat{\theta})(\hat{G}'\hat{S}^{-1}\hat{G})^{-1}A(\hat{\theta})' \xrightarrow{p} A(\theta_0)(G'S^{-1}G)^{-1}A(\theta_0)' \tag{49}$$

(48)、(49) より

$$n\{a(\hat{\theta}) - a(\theta_0)\}'\{A(\hat{\theta})(\hat{G}'\hat{S}^{-1}\hat{G})^{-1}A(\hat{\theta})'\}^{-1}\{a(\hat{\theta}) - a(\theta_0)\} \xrightarrow{d} \chi_r^2 \tag{50}$$

である。(50),(49) は帰無仮説、対立仮説のどちらが正しい場合にも成立する。帰無仮説 $a(\theta_0) = 0$ が正しい場合、(50) の左辺で $a(\theta_0) = 0$ とおくと T_W に一致することから

$$T_W \xrightarrow{d} \chi_r^2 \;\; \text{under } H_0$$

が示される。

次に H_1 の下での T_W の性質を調べる。$a'Va$, $b'Vb$ が計算できるような任意の a, b, V に対して $a'Va = (a-b)'V(a-b) + (a-b)'Vb + b'V(a-b) + b'Vb$ が成立するので、

$$\begin{aligned}
\frac{T_W}{n} &= \{a(\hat{\theta}) - a(\theta_0)\}'\{A(\hat{\theta})(\hat{G}'\hat{S}^{-1}\hat{G})^{-1}A(\hat{\theta})'\}^{-1}\{a(\hat{\theta}) - a(\theta_0)\} \\
&+ \{a(\hat{\theta}) - a(\theta_0)\}'\{A(\hat{\theta})(\hat{G}'\hat{S}^{-1}\hat{G})^{-1}A(\hat{\theta})'\}^{-1}a(\theta_0) \\
&+ a(\theta_0)'\{A(\hat{\theta})(\hat{G}'\hat{S}^{-1}\hat{G})^{-1}A(\hat{\theta})'\}^{-1}\{a(\hat{\theta}) - a(\theta_0)\} \\
&+ a(\theta_0)'\{A(\hat{\theta})(\hat{G}'\hat{S}^{-1}\hat{G})^{-1}A(\hat{\theta})'\}^{-1}a(\theta_0)
\end{aligned}$$

と書ける。対立仮説が正しい場合、右辺第１、２、３項は $a(\theta)$ の連続性、$\hat{\theta} \xrightarrow{p} \theta_0$、(49) から 0 に確率収束する。また、(49) から第４項は $a(\theta_0)'\{A_0(G'S^{-1}G)^{-1}A_0'\}^{-1}a(\theta_0)$ に収束する。従って、対立仮説が正しい時には

$$\frac{T_W}{n} \xrightarrow{p} a(\theta_0)'\{A_0(G'S^{-1}G)^{-1}A_0'\}^{-1}a(\theta_0) = const.(>0)$$

（証明終）

（※１）H_1 の下では T_W/n がある正の定数に確率収束するので、「T_W は発散する」と解釈できる。実際、H_1 が正しい時、n が大きくなると H_0 が棄却される確率が１に近づく。それを、$T_W \xrightarrow{p} \infty$ と書くこともあるが、それは任意の（大きい）$C > 0$、（小さい）$\epsilon > 0$ に対して、ある n_0 が存在して、すべての $n > n_0$ について

$$P(T_W < C) < \epsilon$$

が成立する、ということを意味する。

（※２） (iii) は重複した制約が含まれないことを保証している。また、(iii) が成立しないなら逆行列が計算できないという問題が生じる。

(2) ラグランジュ乗数（LM）検定

この検定は、制約付き最大化問題において、制約が binding でなければラグランジュ乗数が 0 になり、binding であればラグランジュ乗数が 0 でないという性質を用いた検定である。制約の下での S の推定量を $\tilde{S}$、ラグランジュ乗数を ν として、(47) の解は以下の一階の条件を満たす。

$$\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t,x_t;\tilde{\theta})}{\partial\theta'}\}'\tilde{S}^{-1}\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t m(y_t,x_t;\tilde{\theta})\}+\sqrt{n}A(\tilde{\theta})'\tilde{\nu} = 0 \quad (51)$$

$$\sqrt{n}a(\tilde{\theta}) = 0 \quad (52)$$

ラグランジュ乗数検定統計量は

$$T_{LM} = n\tilde{\nu}'\{A(\tilde{\theta})(\tilde{G}'\tilde{S}^{-1}\tilde{G})^{-1}A(\tilde{\theta})'\}^{-1}\tilde{\nu}$$

である。ただし $\tilde{G} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial z_t m(y_t,x_t;\tilde{\theta})}{\partial\theta'}$ である。

Theorem 21. ラグランジュ乗数検定

以下の仮定を導入する。

(i) Theorem 19 の仮定が *A13* を除いて成立する。

(ii) $a(\theta)$ は $r(<k)$ 次元の連続微分可能なベクトル値関数とする。

(iii) A_0 は *full row rank* である。

(iv) 帰無仮説のもとで $\tilde{S} \xrightarrow{p} S$

(v) 対立仮説のもとで $\tilde{S} \xrightarrow{p} S_1$ で、S_1 は正値定符号行列である。

(vi) $A_1 = A(\theta_1)$ は *full row rank*、$G_1 = E\{\frac{z_t\partial m(y_t,x_t;\theta_1)}{\partial\theta'}\}$ は *full column rank* である。

仮定 *(i)-(iv)* の下で、

$$T_{LM} \xrightarrow{d} \chi^2_r \ \ under\ H_0$$

仮定 *(i)*、*(ii)*、*(v)*、*(vi)* の下で、

$$\frac{T_{LM}}{n} \xrightarrow{p} \gamma'\{A_1(G_1'S_1^{-1}G_1)^{-1}A_1'\}^{-1}\gamma = const.(>0)\ \ under\ H_1$$

(証明)

平均値の定理より

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t m(y_t,x_t;\tilde{\theta}) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t m(y_t,x_t;\theta_1)+\frac{1}{n}\sum_{t=1}^{n}z_t\frac{\partial m(y_t,x_t;\bar{\theta}_1)}{\partial\theta'}\sqrt{n}(\tilde{\theta}-\theta_1) \quad (53)$$

$$\sqrt{n}a(\tilde{\theta}) = A(\bar{\theta}_2)\sqrt{n}(\tilde{\theta}-\theta_1) \quad (54)$$

を満たす $\bar{\theta}_1=\lambda_1\tilde{\theta}+(1-\lambda_1)\theta_1$、$\bar{\theta}_2=\lambda_2\tilde{\theta}+(1-\lambda_2)\theta_1$ が存在する。(51)、(52) に (53)、(54) を代入すると

$$\tilde{\Psi}=\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t,x_t;\tilde{\theta})}{\partial\theta'}\}'\tilde{S}^{-1}\{\frac{1}{n}\sum_{t=1}^{n}z_t\frac{\partial m(y_t,x_t;\bar{\theta}_1)}{\partial\theta'}\}$$

として

$$\begin{aligned}\tilde{\Psi}\sqrt{n}(\tilde{\theta}-\theta_1)+\sqrt{n}A(\tilde{\theta})'\tilde{\nu} &= -\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t,x_t;\tilde{\theta})}{\partial\theta'}\}'\tilde{S}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t m(y_t,x_t;\theta_1)\\ A(\bar{\theta}_2)\sqrt{n}(\tilde{\theta}-\theta_1) &= 0\end{aligned}$$

を得る。これを行列表記すると

$$\begin{pmatrix}\tilde{\Psi} & A(\tilde{\theta})'\\ A(\bar{\theta}_2) & 0\end{pmatrix}\begin{pmatrix}\sqrt{n}(\tilde{\theta}-\theta_1)\\ \sqrt{n}\tilde{\nu}\end{pmatrix}=\begin{pmatrix}-\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t,x_t;\tilde{\theta})}{\partial\theta'}\}'\tilde{S}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t m(y_t,x_t;\theta_1)\\ 0\end{pmatrix}$$

分割行列の逆行列の公式 (Hayashi (2000), p.673) を使うと、

$$\sqrt{n}\tilde{\nu}=-\{A(\bar{\theta}_2)\tilde{\Psi}^{-1}A(\tilde{\theta})'\}^{-1}A(\bar{\theta}_2)\tilde{\Psi}^{-1}\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t,x_t;\tilde{\theta})}{\partial\theta'}\}'\tilde{S}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t m(y_t,x_t;\theta_1) \tag{55}$$

帰無仮説が正しい時、$\theta_1=\theta_0$ で、$\tilde{\theta}$,$\bar{\theta}_1,\bar{\theta}_2$ はすべて θ_0 に確率収束するので、仮定の下で Lemma 4 が適用できて

$$\tilde{G}\xrightarrow{p}G,\ \ \tilde{\Psi}\xrightarrow{p}G'S^{-1}G, \tag{56}$$

また、仮定 (ii) より

$$A(\bar{\theta}_2)\xrightarrow{p}A(\theta_0) \tag{57}$$

である。一方、(56), 仮定 (iv), Theorem 19 の仮定 A10 から

$$\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t,x_t;\tilde{\theta})}{\partial\theta'}\}'\tilde{S}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t m(y_t,x_t;\theta_0)\xrightarrow{d}N(0,G'S^{-1}G) \tag{58}$$

である。(55), (56), (57), (58) より

$$\sqrt{n}\tilde{\nu}\xrightarrow{d}N(0,(G'S^{-1}G)^{-1})$$

仮定 (ii) と $\tilde{\theta}\xrightarrow{p}\theta_0$ より $A(\tilde{\theta})\xrightarrow{p}A(\theta_0)$、(56)、仮定 (iv) より、

$$T_{LM}\xrightarrow{d}\chi_r^2\ \text{ under } H_0$$

が示される。

対立仮説のもとでは、(55) は

$$\begin{aligned}&\sqrt{n}\tilde{\nu}\\ =\ &-\sqrt{n}\{A(\bar{\theta}_2)\tilde{\Psi}^{-1}A(\tilde{\theta})'\}^{-1}A(\bar{\theta}_2)\tilde{\Psi}^{-1}\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t\partial m(y_t,x_t;\tilde{\theta})}{\partial\theta'}\}'\tilde{S}^{-1}\frac{1}{n}\sum_{t=1}^{n}z_t m(y_t,x_t;\theta_1)\end{aligned}$$

右辺を $\sqrt{n}$ で割ったものは

$$\gamma=-\{A_1(G_1'S_1^{-1}G)_1^{-1}A_1'\}^{-1}A_1(G_1'S_1^{-1}G_1)^{-1}G_1'S_1^{-1}E\{z_t m(y_t,x_t;\theta_1)\}$$

に収束する。Theorem 17 または 18 の A5 の仮定より $E\{z_t m(y_t, x_t; \theta_1)\} \neq 0$ なので、(v), (vi) より $\gamma \neq 0$。以上から、対立仮説のもとでのワルド検定の収束の証明と同様にして対立仮説のもとでは

$$\frac{T_{LM}}{n} \xrightarrow{p} \gamma'\{A_1(G_1' S_1^{-1} G_1)^{-1} A_1'\}^{-1}\gamma = const.(> 0)$$

（証明終）

（※）　仮定 (vi) の「$A_1 = A(\theta_1)$ は full row rank」は、対立仮説のもとでの収束の形で定理を記述するために必要であるが、実際には成立していなくても検定上は困らない。むしろ、発散が早くなると考えられ、望ましいであろう。

(3) 尤度比（LR）検定

ここで考える枠組みは最尤法ではないので、厳密には尤度比検定という言葉は適切ではないかもしれないが、最尤法の枠組みにおける尤度比検定と全く同等の考え方から統計量が導かれるため、そのように呼ぶことにする。検定統計量は

$$T_{LR} = -2n\{Q_n(\hat{\theta}) - Q_n(\tilde{\theta})\}$$

であり、適当な条件のもとで、

$$T_{LR}\begin{cases} \xrightarrow{d} & \chi_r^2 \text{ under } H_0 \\ \xrightarrow{p} & \infty \text{ under } H_1 \end{cases}$$

が示される。

6.4.2　J 検定（過剰識別の検定）

通常、GMM 推定に用いられるモーメント条件の個数（p）は、パラメータの次元（k）よりも多い。もし $p = k$ なら、just-identified であり、目的関数の最小値は 0 になる。しかし、$p > k$ のときは、用いるモーメント条件が正しければ目的関数の最小値 0 に近い値になるはずである。これを用いて検定を行うのが J テストである。なお、もしもモーメント条件の中に間違った制約が含まれていると、推定結果は一致性を持たない。検定の帰無仮説と対立仮説は

$$\begin{aligned} H_0: \ E\{z_t m(y_t, x_t; \theta_0)\} &= 0 \\ H_1: \ E\{z_t m(y_t, x_t; \theta_0)\} &= \delta(\neq 0) \end{aligned}$$

である。δ は p 次元ベクトルであるが、対立仮説はその p 個の要素のうち、ひとつでもゼロでないものがあるという仮説である。検定統計量を

$$J = nQ_n(\hat{\theta})$$

すると、以下の結果が成り立つ。

Theorem 22. J 検定

Corollary 1 と同じ条件を仮定する。ただし、*Theorem 19* の *A10* を

A10': $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[z_t m(y_t, x_t; \theta_0) - E\{z_t m(y_t, x_t; \theta_0)\}] \xrightarrow{d} N(0, S)$

で置き換えるものとする。M_1 を証明中で定義する、ある定数行列として、

$$J\begin{cases} \xrightarrow{d} & \chi_{p-k}^2 \ \ under\ H_0 \\ \xrightarrow{p} & \infty \quad under\ H_1\ if\ \delta'(S^{-1} - S^{-\frac{1}{2}} M_1 S^{-\frac{1}{2}})\delta \neq 0 \end{cases}$$

（証明）

平均値の定理より

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\hat{\theta}) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta_0) + \frac{1}{n}\sum_{t=1}^{n} z_t \frac{\partial m(y_t,x_t;\bar{\theta})}{\partial \theta'}\sqrt{n}(\hat{\theta}-\theta_0)$$

で、$\bar{\theta} \xrightarrow{p} \theta_0$ である。(43) を代入すると、ウェイト行列に $\hat{S}^{-1}$ を用いることに注意して、

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\hat{\theta}) &= S^{\frac{1}{2}}\{S^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta_0)\} \\
&- \frac{1}{n}\sum_{t=1}^{n} z_t \frac{\partial m(y_t,x_t;\bar{\theta})}{\partial \theta'} \\
&\times \left[\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t \partial m(y_t,x_t;\hat{\theta})}{\partial \theta'}\}'\hat{S}^{-1}\{\frac{1}{n}\sum_{t=1}^{n} z_t \frac{\partial m(y_t,x_t;\bar{\theta})}{\partial \theta'}\}\right]^{-1} \\
&\times\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t \partial m(y_t,x_t;\hat{\theta})}{\partial \theta'}\}'\hat{S}^{-1}S^{\frac{1}{2}}\{S^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta_0)\}
\end{aligned}$$

ここで、

$$\begin{aligned}
\hat{M} &= S^{-1/2}\{\frac{1}{n}\sum_{t=1}^{n} z_t \frac{\partial m(y_t,x_t;\bar{\theta})}{\partial \theta'}\}\left[\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t \partial m(y_t,x_t;\hat{\theta})}{\partial \theta'}\}'\hat{S}^{-1}\{\frac{1}{n}\sum_{t=1}^{n} z_t \frac{\partial m(y_t,x_t;\bar{\theta})}{\partial \theta'}\}\right]^{-1} \\
&\quad \times\{\frac{1}{n}\sum_{t=1}^{n}\frac{z_t \partial m(y_t,x_t;\hat{\theta})}{\partial \theta'}\}'\hat{S}^{-1}S^{\frac{1}{2}}
\end{aligned}$$

とおくと、

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\hat{\theta}) = S^{\frac{1}{2}}(I-\hat{M})\{S^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta_0)\}$$

なので

$$\begin{aligned}
J &= \{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\hat{\theta})\}'\hat{S}^{-1}\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\hat{\theta})\} \\
&= \{S^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta_0)\}'(I-\hat{M}')S^{\frac{1}{2}}\hat{S}^{-1}S^{\frac{1}{2}}(I-\hat{M})\{S^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta_0)\}
\end{aligned} \tag{59}$$

となる。

帰無仮説が正しい時、仮定 A10' より

$$S^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t m(y_t,x_t;\theta_0) \xrightarrow{d} N(0,I), \tag{60}$$

(45)、(46)、Theorem 19 の仮定 A13 より

$$\hat{M} \xrightarrow{p} M = S^{-\frac{1}{2}}G(G'S^{-1}G)^{-1}G'S^{-\frac{1}{2}} \tag{61}$$

である。また、$I - M$ はべき等行列であるから、$P = S^{-\frac{1}{2}}G$ とおくと

$$\begin{aligned} rank(I-M) &= rank(I - P(P'P)^{-1}P') \\ &= tr(I - P(P'P)^{-1}P') \\ &= p - k \end{aligned} \tag{62}$$

である。(59)、(60)、(61)、(62) より

$$J \xrightarrow{d} \chi^2_{p-k}$$

であることがわかる。

対立仮説が正しいとき、$E\{z_t m(y_t, x_t; \theta_0)\} = \delta \neq 0$ であり、

$$S^{-\frac{1}{2}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_t m(y_t, x_t; \theta_0) = \sqrt{n} S^{-\frac{1}{2}} \delta + S^{-\frac{1}{2}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{z_t m(y_t, x_t; \theta_0) - \delta\}$$

と書ける。A10' より、右辺第２項は標準正規分布に収束するが、第１項は ∞ か $-\infty$ に発散する。また、

$$\hat{\theta} \xrightarrow{p} \theta_1 = \arg\min_{\theta} E\{z_t m(y_t, x_t; \theta)\}' S^{-1} E\{z_t m(y_t, x_t; \theta)\}$$

である（一般に $\theta_1 \neq \theta_0$ なので、対立仮説が正しい時には $\hat{\theta}$ は θ_0 の一致推定量ではない）ため、(若干の追加的仮定の下で)$\hat{M}$ は M でない別の行列 M_1 に確率収束する。従って、J のうち、オーダーの大きい部分を取り出すと

$$J \approx n\delta' S^{-\frac{1}{2}} (I - M_1) S^{-\frac{1}{2}} \delta$$

である。$I - M_1$ はべき等行列なので正値定符号ではないが（半正値定符号である）、$\delta'(S^{-1} - S^{-\frac{1}{2}} M_1 S^{-\frac{1}{2}})\delta \neq 0$ ならば J は ∞ に発散する。（証明終）

（※１）証明からわかるように、もしも偶然に $\delta'(S^{-1} - S^{-\frac{1}{2}} M_1 S^{-\frac{1}{2}})\delta = 0$ であれば、帰無仮説が間違っていても、$J \xrightarrow{d} \chi^2_{p-k}$ となってしまい、Ｊ検定は検出力がないことになってしまう。ただし、一般にはそうなっている可能性は極めて低いため、実用上は問題ない。

（※２）パラメータの個数（k）と操作変数の個数（p）が同じ場合は、目的関数がゼロになるためＪ検定は機能しない。

（※３）$k < p$ なら、操作変数に一つでもモーメント条件を満たさないものがある場合にはＪ検定で帰無仮説は棄却される。しかし、どのモーメント条件が間違っているか、あるいはいくつ間違っているかはわからない。従って、実際にＪ検定で棄却された時は、いくつかのモーメント条件を取り除いてＪ検定を行い、検定をパスしたセットのモーメント条件を用いた推定を最終結果とする。

（※４）（※３）に述べたように、実用上は、Ｊ検定で棄却されなければ、モーメント条件は正しいものとしてその推定結果を採択することになる。しかし、そのためには p 個のうち正しいモーメント条件が少なくとも $k+1$ 個なければならない。（$\because$ もし正しいモーメント条件が k 個以下なら、どの $k+1$ 個のモーメント条件を用いてもＪ検定で棄却される。そこからモーメント条件を１つ抜くとモーメント条件の数とパラメータ数が一致してＪ検定が機能しなくなり、どれが間違っているのか判別できない。逆に正しいモーメント条件が $k+1$ 個あれば、それ以外の条件を除いた場合にＪ検定をパスするはずである。）

7 Limited dependent variable models

従属変数が、家や車などの耐久消費財を保有しているか否か、といった選択データの分析には質的選択モデル（qualitative response model）が用いられる。その最も基本的なモデルが２値選択モデル（binary choice molde）である。３種類以上の選択肢があるときには、多項選択モデル（multiple choice model）が用いられるが、それには選択肢に順番がある場合（ordered model）とない場合がある。

　また、被説明変数に打切りがある場合、censored regression, truncated regression model が用いられる。

7.1 ２値選択モデルとその推定

y_i を 0 か 1 の値を取る従属変数、x_i を k 次元の説明変数、F をある分布関数として、

$$P(y_i = 1|x_i) = F(x_i'\beta) \tag{63}$$

を２値選択モデルという。より一般的には、$F(x_i'\beta)$ でなく、g をある関数としてより一般的に $F(g(x_i, \beta))$ とすることも考えられる。(63) は、以下のような状況から導出される。

　（１）Latent variable model

　例えば、x を個人の属性として、家を保有している場合は $y = 1$、保有していない場合は $y = 0$ とする。個人 i は $y_i^* = x_i'\beta + \epsilon_i$ の値が正か負かによって意思決定を行うが、それ自体は観測されない変数（latent variable）である。y^* が正の場合は $y = 1$、負の場合は $y = 0$ となるものとする。また、x_i を条件とする $-\epsilon_i$ の分布関数を F とする。

$$\begin{aligned} y_i^* &= x_i'\beta + \epsilon_i \\ y_i &= I(y_i^* \geq 0) \end{aligned}$$

そのとき、

$$\begin{aligned} P(y_i = 1|x_i) &= P(y_i^* \geq 0|x_i) \\ &= P(x_i'\beta + \epsilon_i \geq 0|x_i) \\ &= P(-\epsilon_i \leq x_i'\beta|x_i) \\ &= F(x_i'\beta) \end{aligned}$$

となる。

　（２）効用モデル

　例えば、個人 i が通勤に車を使う $(y_i = 1)$ か、公共交通機関を使う $(y_i = 0)$ かを決める問題を考える。前者の場合の効用は $u_{1i} = z_{1i}'\beta + \epsilon_{1i}$、後者の場合の効用は $u_{0i} = z_{0i}'\beta + \epsilon_{0i}$ とする。z は説明変数で、たとえば所要時間、費用等である。また、$\epsilon_{0i} - \epsilon_{1i}$ の分布関数を F とし、$x_i = z_{1i} - z_{0i}$ とする。そのとき、個人 i の意思決定は次のように表される。

$$\begin{aligned} P(y_i = 1|x_i) &= P(u_{1i} \geq u_{0i}|x_i) \\ &= P\{(z_{1i}'\beta + \epsilon_{1i}) - (z_{0i}'\beta + \epsilon_{0i}) \geq 0|x_i\} \\ &= P(\epsilon_{0i} - \epsilon_{1i} \leq (z_{1i} - z_{0i})'\beta|x_i) \\ &= F(x_i'\beta) \end{aligned}$$

　（３）推定

　(y_i, x_i), $i = 1, \cdots, n$ を無作為標本とする。y を説明変数、x を被説明変数として OLS 回帰しても β は推定できない。なぜなら、単純化のために x はスカラーとして

$$E(\beta_{OLS}|x) = \frac{\sum(x_i - \bar{x})E(y_i|x_i)}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})F(x_i'\beta)}{\sum(x_i - \bar{x})^2} \neq \beta$$

だからである。F を特定することによって、β を ML 推定することができる。ML 推定で実証上よく用いられる特定化は

- 線形確率モデル： $F(x) = x$
- プロビットモデル： $F(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})dt$（正規分布）
- ロジットモデル： $F(x) = \Lambda(x) = \frac{\exp(x)}{1+\exp(x)}$（ロジスティック分布）

である。ただし、線形確率モデルでは x が負値や 1 より大きい値はとれないので、注意して適用する必要がある。各個人の確率関数は

$$
\begin{aligned}
P(y_i &= 1|x_i) = F(x_i'\beta) \\
P(y_i &= 0|x_i) = 1 - F(x_i'\beta)
\end{aligned}
$$

なので、

$$
\begin{aligned}
P(y_i = y|x_i) &= \begin{cases} F(x_i'\beta) & y = 1 \\ 1 - F(x_i'\beta) & y = 0 \end{cases} \\
&= F(x_i'\beta)^y \{1 - F(x_i'\beta)\}^{1-y} \quad (y = 1, 0)
\end{aligned}
$$

と書き換えることができる。したがって、尤度関数は、

$$
L(\beta) = \prod_{i=1}^{n} F(x_i'\beta)^{y_i} \{1 - F(x_i'\beta)\}^{1-y_i}
$$

であり、対数尤度関数は

$$
\log L(\beta) = \sum_{i=1}^{n} y_i \log F(x_i'\beta) + (1 - y_i) \log\{1 - F(x_i'\beta)\} \tag{64}
$$

となる。また、パラメータ空間を Θ として、最尤推定量は

$$
\hat{\beta} = \arg\max_{\beta\in\Theta} \log L(\beta) \tag{65}
$$

で定義される。微分可能性等の条件のもとでは、$f(x) = dF(x)/dx$ とすると、MLE の一階の条件は

$$
\frac{\partial \log L(\hat{\beta})}{\partial \beta} = \sum_{i=1}^{n} \frac{y_i - F(x_i'\hat{\beta})}{F(x_i'\hat{\beta})\{1 - F(x_i'\hat{\beta})\}} f(x_i'\hat{\beta}) x_i = 0 \tag{66}
$$

となる。また、２階の条件は

$$
\begin{aligned}
\frac{\partial^2 \log L(\hat{\beta})}{\partial \beta \partial \beta'} &= -\sum_{i=1}^{n} \left[\frac{y_i - F(x_i'\hat{\beta})}{F(x_i'\hat{\beta})\{1 - F(x_i'\hat{\beta})\}} \right]^2 f(x_i'\hat{\beta})^2 x_i x_i' \\
&\quad + \sum_{i=1}^{n} \frac{y_i - F(x_i'\hat{\beta})}{F(x_i'\hat{\beta})\{1 - F(x_i'\hat{\beta})\}} f'(x_i'\hat{\beta}) x_i x_i' \\
&< 0
\end{aligned} \tag{67}
$$

である。この推定量は、M 推定量で $W_i = (y_i, x_i)$、

$$
m(W_i; \beta) = y_i \log F(x_i'\beta) + (1 - y_i) \log\{1 - F(x_i'\beta)\}
$$

とした特殊ケースなので、Thoerem 6, 7, 8 を適用できる。Proposition 7, 8, 9 において条件付き最尤推定量に限って漸近理論を述べる。(y_t, x_t) はエルゴード性をもつ定常過程とし、x_t を条件とする y_t の条件付き密度（または確率）$f(y_t|x_t;\theta)$ であるとする。そのとき、θ の疑似最尤推定量（QML: quasi-ML）は、

$$\hat{\theta} = \arg\max_{\theta} LL(\theta) = \sum_{t=1}^{T} \log f(y_t|x_t;\theta)$$

で定義される。「疑似」と言われるのは、この「対数尤度関数」$LL(\theta) = \sum_{t=1}^{T} \log f(y_t|x_t;\theta)$ は (y_t, x_t), $t = 1, \cdots, T$ の本当の尤度関数でなく、あたかも (y_t, x_t), $t = 1, \cdots, T$ を iid の観測値と考えて作った尤度関数だからである。もしそれが本当に iid であれば、これは厳密な意味で最尤推定になる。また、一般には同時密度関数の関数形の特定に誤りがある場合も QML と言われるが、以下では条件付き密度の特定化 $f(y|x;\theta)$ は正しいものとする。

Proposition 7. 条件付き（疑似）最尤推定量の一致性（*compact parameter space case*）

以下の仮定をおく。

(i) パラメータ空間 Θ は k 次元ユークリッド空間の *compact* な集合である。

(ii) $f(y_t|x_t;\theta)$ はすべての (y_t, x_t) について連続である。

(iii) (identification) $P[f(y_t|x_t;\theta) \neq f(y_t|x_t;\theta_0)] > 0$ *for all* $\theta \neq \theta_0$ *in* Θ

(iv) (dominance) $E[\sup_{\theta\in\Theta} |\log f(y_t|x_t;\theta)|] < \infty$

そのとき、

$$\hat{\theta} \xrightarrow{p} \theta_0$$

(iii) は若干回りくどい表現に思われるが、言いかえると、$\theta = \theta_0$ のときのみ $P[f(y_t|x_t;\theta) \neq f(y_t|x_t;\theta_0)] = 0$ ということである。次に、パラメータ空間のコンパクト性を仮定しない結果を示す。

Proposition 8. 条件付き（疑似）最尤推定量の一致性（*without compactness*）

以下の仮定をおく。

(i)' パラメータ空間 Θ は k 次元ユークリッド空間の *convex* な集合で、真の値 θ_0 は Θ の内点である。

(ii)' $\log f(y_t|x_t;\theta)$ はすべての (y_t, x_t) に対して θ について *concave* である。

(iii) (identification) $P[f(y_t|x_t;\theta) \neq f(y_t|x_t;\theta_0)] > 0$ *for all* $\theta \neq \theta_0$ *in* Θ

(iv)' (dominance) $E[|\log f(y_t|x_t;\theta)|] < \infty$ *for all* $\theta \in \Theta$

そのとき、

$$\hat{\theta} \xrightarrow{p} \theta_0$$

次に、中心極限定理を述べる。Proposition 7 または 8 の条件が成立し、従って $\hat{\theta} \xrightarrow{p} \theta_0$ であるとする。

Proposition 9. 条件付き（疑似）最尤推定量の漸近正規性

(i) θ_0 は Θ の内点である。（*Proposition 8* では仮定済み）

(ii) $f(y_t|x_t;\theta)$ はすべての (y_t, x_t) に対して θ について 2 回連続微分可能である。

(iii) $s(y_t, x_t;\theta) = \frac{\partial \log f(y_t|x_t;\theta)}{\partial\theta}$、$H(y_t, x_t;\theta) = \frac{\partial \log f(y_t|x_t;\theta)}{\partial\theta\partial\theta'}$ として、

$$E\{s(y_t, x_t;\theta_0)\} = 0, \quad -E\{H(y_t, x_t;\theta_0)\} = E\{s(y_t, x_t;\theta_0)s(y_t, x_t;\theta_0)'\}$$

が成り立つ。

(iv) (local dominance on the Hessian) θ_0 の近傍 $\mathcal{N}$ に対して、

$$E[\sup_{\theta\in\mathcal{N}} ||H(y_t, x_t;\theta)||] < \infty$$

であり、従って任意の $\bar{\theta} \xrightarrow{p} \theta_0$ に対して

$$\frac{1}{n}\sum_{i=1}^{n} H(y_t, x_t; \bar{\theta}) \xrightarrow{p} E\{H(y_t, x_t; \theta_0)\}$$

(v) $I(\theta) = -E\{H(y_t, x_t; \theta_0)\}$ は非特異である。
そのとき、

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I(\theta)^{-1})$$

である。

(iii) は $w = (y, x)$ として $a(w; \theta) = f(w, \theta)$, $\frac{\partial \log f(w,\theta)}{\partial \theta} f(w; \theta)$ について、以下の微分と積分の入れ替えが成立すれば成立する。

$$\frac{\partial}{\partial \theta}\int a(w, \theta)dw = \int \frac{\partial}{\partial \theta} a(w, \theta)dw$$

通常の計量経済モデルではこれは成立するが、そのための十分条件は Newey and McFadden (1994, Handbook of Econometrics, Ch.36, Lemma 3.6, p.2152) を参照。

以下に、Proposition 8 に基づいてロジットモデル、プロビットモデルの場合について (65) の解の一致性を示す。

Theorem 23. ロジットモデルとプロビットモデルの一致性
$\Theta = R^k$ で $E(x_i x_i')$ が非特異行列であるとき、

$$\hat{\beta} \xrightarrow{p} \beta_0$$

である。

（証明）　Proposition 8 の各条件が成立していることを示せばよい。(i)' は明らかである。

(ii)' について、ロジットモデルの場合のみ示す（プロビットの場合は若干厄介であるが、Amemiya (1985, p.273-274; p273 最終式に誤植の可能性あり) を参照）。簡単な計算から

$$f(u) = \Lambda'(u) = \Lambda(u)\{1 - \Lambda(u)\}$$

$$f'(u) = \Lambda''(u) = \{1 - 2\Lambda(u)\}\Lambda(u)\{1 - \Lambda(u)\}$$

であることがわかる。したがって、(67) の等式部分に $F = \Lambda$ を代入して 2 次微分を計算すると、$y_i(1 - y_i) = 0$ を用いて

$$\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} \Lambda(x_i'\beta)\{1 - \Lambda(x_i'\beta)\}x_i x_i' \tag{68}$$

であることがわかる。$0 < \Lambda(x_i'\beta) < 1$、$x_i x_i' > 0$ なので、(ii)' が成立する。

(iii)　仮定より $E(x_i x_i')$ は正値定符号であるから、$E(x_i'\beta - x_i'\beta_0)^2 = (\beta - \beta_0)' E(x_i x_i')(\beta - \beta_0)$ より、$\beta \neq \beta_0$ なら $E(x_i'\beta - x_i'\beta_0)^2 \neq 0$ である。従って、$\beta \neq \beta_0$ のとき $P(x_i'\beta \neq x_i'\beta_0) > 0$ でなければならない。すると、正規分布、ロジット分布の分布関数は狭義の増加関数なので、$P[F(x_i'\beta) \neq F(x_i'\beta_0)] > 0$ が成立する。さて、

$$f(y_i|x_i; \beta) \quad = \quad F(x_i'\beta)^{y_i}\{1 - F(x_i'\beta)\}^{1-y_i}$$

なので、

$$\begin{aligned}
& P[f(y_t|x_t;\beta) = f(y_t|x_t;\beta_0)] \\
= \; & P[F(x_i'\beta)^{y_i}\{1-F(x_i'\beta)\}^{1-y_i} = F(x_i'\beta_0)^{y_i}\{1-F(x_i'\beta_0)\}^{1-y_i}] \\
= \; & P[F(x_i'\beta)^{y_i}\{1-F(x_i'\beta)\}^{1-y_i} = F(x_i'\beta_0)^{y_i}\{1-F(x_i'\beta_0)\}^{1-y_i}|y_i=1]P(y_i=1) \\
+ \; & P[F(x_i'\beta)^{y_i}\{1-F(x_i'\beta)\}^{1-y_i} = F(x_i'\beta_0)^{y_i}\{1-F(x_i'\beta_0)\}^{1-y_i}|y_i=0]P(y_i=0) \\
= \; & P[F(x_i'\beta) = F(x_i'\beta_0)|y_i=1]P(y_i=1) + P[F(x_i'\beta) = F(x_i'\beta_0)|y_i=0]P(y_i=0) \\
= \; & P[F(x_i'\beta) = F(x_i'\beta_0)]
\end{aligned}$$

となる。両辺をそれぞれ 1 から引くと

$$P[f(y_t|x_t;\theta) \neq f(y_t|x_t;\theta_0)] = P[F(x_i'\beta) \neq F(x_i'\beta_0)] > 0$$

を得る。なお、ここで $f(y_i|x_i;\beta)$ は密度ではなく確率である。

(iv)' $\log f(y_i|x_i;\beta) = y_i \log F(x_i'\beta) + (1-y_i)\log\{1-F(x_i'\beta)\}$ なので、$y_i = 0, 1$、$F(-u) = 1 - F(u)$ に注意して、

$$E[|\log f(y_i|x_i;\beta)|] \leq E[|\log F(x_i'\beta)|] + E[|\log F(-x_i'\beta)|]$$

である。まず、プロビットモデル（$F = \Phi$）のとき、$|\log\Phi(x_i'\beta)| = |-\frac{1}{2}\log(2\pi) - \frac{1}{2}(x_i'\beta)^2| \leq c + (x_i'\beta)^2$ であり、仮定（非特異）から $E(x_i x_i')$ は有界なので、

$$E[|\log\Phi(x_i'\beta)|] + E[|\log\Phi(-x_i'\beta)|] \leq 2c + 2E(x_i'\beta)^2 = 2c + 2\beta' E(x_i x_i')\beta < \infty$$

となる。次にロジットモデル（$F = \Lambda$）の場合を考える。$u \geq 0$ なら、$\Lambda(0) \leq \Lambda(u) < 1$ なので、

$$|\log\Lambda(u)| \leq |\log\Lambda(0)| \leq |u| + |\log\Lambda(0)|$$

である。$\exp(\cdot)$, $\log(\cdot)$ は単調増加関数なので、$u < 0$ なら $0 > \log\{\frac{1}{1+\exp(u)}\} > \log\{\frac{1}{1+\exp(0)}\} = \log\Lambda(0)$ である。よって

$$|\log\Lambda(u)| = |u + \log\{\frac{1}{1+\exp(u)}\}| < |u + \log\Lambda(0)| < |u| + |\log\Lambda(0)|$$

なので、任意の u について $|\log\Lambda(u)| \leq |u| + |\log\Lambda(0)|$ が成り立つ。よって、$E|\beta' x_i| < \infty$ なら dominance condition が成り立つが、それは $E(x_i x_i')$ の有界性から成立する。（証明終）

Theorem 24. 漸近正規性

$\Theta = R^k$ で $E(x_i x_i')$ が非特異行列であるとき、

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, I(\beta_0)^{-1})$$

である。

（証明） Proposition 9 の (i)-(v) を示せばよい。(i) は $\Theta = R^k$ として、成立する。(ii) は明らかに成立する。

(iii) スコアとヘシアンは $F = \Lambda, \Phi$ に対して

$$\begin{aligned}
s(y_i, x_i;\beta_0) &= \frac{\partial \log f(y_i|x_i;\beta_0)}{\partial\beta} \\
&= \frac{y_i - F(x_i'\beta_0)}{F(x_i'\beta_0)\{1-F(x_i'\beta_0)\}} f(x_i'\beta_0) x_i
\end{aligned}$$

$$
\begin{aligned}
H(y_i, x_i; \beta_0) &= \frac{\partial \log f(y_i|x_i;\beta_0)}{\partial\beta\partial\beta'} \\
&= -\left[\frac{y_i - F(x_i'\beta_0)}{F(x_i'\beta_0)\{1-F(x_i'\beta_0)\}}\right]^2 f(x_i'\beta_0)^2 x_i x_i' \\
&\quad + \frac{y_i - F(x_i'\beta_0)}{F(x_i'\beta_0)\{1-F(x_i'\beta_0)\}} f'(x_i'\beta_0) x_i x_i'
\end{aligned}
$$

である。$E(y_i|x_i) = E(y_i^2|x_i) = F(x_i'\beta_0)$ なので、$E\{s(y_i, x_i; \beta_0)\} = 0$ である。

$$
\begin{aligned}
E\{s(y_i, x_i; \beta_0)s(y_i, x_i; \beta_0)'|x_i\} &= \frac{E[\{y_i - F(x_i'\beta_0)\}^2|x_i]}{F(x_i'\beta_0)^2\{1-F(x_i'\beta_0)\}^2} f(x_i'\beta_0)^2 x_i x_i' \\
&= \frac{E(y_i^2|x_i) - F(x_i'\beta_0)^2}{F(x_i'\beta_0)^2\{1-F(x_i'\beta_0)\}^2} f(x_i'\beta_0)^2 x_i x_i' \\
&= \frac{1}{F(x_i'\beta_0)\{1-F(x_i'\beta_0)\}} f(x_i'\beta_0)^2 x_i x_i'
\end{aligned}
$$

であり、$E\{y_i - F(x_i'\beta_0)|x_i\} = 0$ なので

$$
E\{H(y_i, x_i; \beta_0)|x_i\} = -\frac{1}{F(x_i'\beta_0)\{1-F(x_i'\beta_0)\}} f(x_i'\beta_0)^2 x_i x_i'
$$

である。したがって、(iii) が成立する。

(iv) プロビットモデルに関しては面倒である (Newey & McFadden (1994, Handbook of Econometrics, Ch.36, p.2147, Example 1.2 参照) が、ロジットモデルについては比較的簡単である。$0 \le \Lambda(x_i'\beta) \le 1$ と (68) より、すべての β にたいして

$$
||H(y_i, x_i; \beta)|| \le ||x_i x_i'||
$$

となるが、$E(x_i x_i')$ が非特異であることより (iv) が成り立つ。

(v) $1 - \Lambda(x_i'\beta_0) = \Lambda(-x_i'\beta_0)$ なので、任意の $c > 0$ に対して、

$$
\begin{aligned}
I(\beta_0) &= -E\{H(y_i, x_i; \beta_0)\} = E[\Lambda(x_i'\beta_0)\{1-\Lambda(x_i'\beta_0)\}x_i x_i'] \\
&\ge E[I(|x_i'\beta_0| \le c)\Lambda(x_i'\beta_0)\Lambda(-x_i'\beta_0)x_i x_i'] \\
&\ge \Lambda(-c)^2 E\{I(|x_i'\beta_0| \le c)x_i x_i'\}
\end{aligned}
$$

最後の行列は c を十分大きくとれば非特異になる。したがって、$I(\beta_0)$ は非特異である。(証明終)

7.2 打切り回帰モデル（truncated regression, censored regression model, Tobit model）

基本的に回帰モデルであるが、被説明変数がある閾値以下の個体に関しては観測値が得られない場合（truncated sample）や、ある閾値以下のときは被説明変数が（たとえば非負制約によって）すべて 0 になってしまう場合を考える。例えば、労働供給時間、購入する耐久消費財の価格、といったものが被説明変数である場合に生ずる。その場合、得られたサンプルで OLS 回帰を行うとバイアスが生ずる。なお、説明変数に閾値があって打ち切られる場合はバイアスの問題は生じないため、問題なく OLS 推定ができる。

7.2.1 Truncated regression model

y_i^* は観測されない変数として、

$$y_i^* = x_i'\beta + \epsilon_i$$

$$y_i = \begin{cases} y_i^* & if\ y_i^* \geq 0 \\ 0 & if\ y_i^* < 0 \end{cases}$$

という状況を考える。例えば、女性の労働力市場への参加・不参加の問題を考える際に、個人 i の労働時間を y_i、属性を x_i とすると、上のような構造を考えることができる。また、購入した耐久消費財の価格を説明するモデルにも上のような定式化が考えられる。$y_i = 0$ の個体については観測値がないとき、そのようなサンプルを truncated sample という。このモデルも OLS 推定を行うことができない。ϵ の密度関数、分布関数を f, F とする。上のモデルは、次のように書き換えられる。

$$y_i = \begin{cases} x_i'\beta + \epsilon_i & if\ x_i'\beta + \epsilon_i \geq 0 \\ 0 & if\ x_i'\beta + \epsilon_i < 0 \end{cases}$$

このようなモデルを分析するには、切断のある分布を考える必要がある。u の分布が c より下で切断されているとき、切断された分布は

$$f(u|u > c) = \frac{f(u)}{P(u > c)} = \frac{f(u)}{1 - F(c)} \tag{69}$$

である。$u \sim N(\mu, \sigma^2)$ とすると、切断された分布の期待値と分散は

$$\begin{aligned} E(u|u > c) &= \mu + \sigma\lambda(v) \\ Var(u|u > c) &= \sigma^2[1 - \lambda(v)\{\lambda(v) - v\}] \end{aligned}$$

となる。ただし、$v = (c - \mu)/\sigma$、$\lambda(v) = \frac{\phi(v)}{1-\Phi(v)}$ で、ϕ, Φ はそれぞれ標準正規分布の密度関数と分布関数である。$\lambda(v)$ は逆ミルズ比 (inverse Mill's ratio)、ハザード関数 (hazard function) と呼ばれる。以下、$y^*|x \sim N(x'\beta, \sigma^2)$ を仮定して OLS と ML 推定法を紹介する。

(1)OLS 推定

$u = y^*$、$c = 0$ と考えて上の結果を用いると、切断回帰モデルについて

$$\begin{aligned} E(y_i|x_i) &= x_i'\beta + \sigma\lambda(\frac{-x_i'\beta}{\sigma}) \\ Var(y_i|x_i) &= \sigma^2\left\{1 - \lambda(\frac{-x_i'\beta}{\sigma})[\lambda(\frac{-x_i'\beta}{\sigma}) - \frac{-x_i'\beta}{\sigma}]\right\} \end{aligned} \tag{70}$$

を得る。従って、OLS 推定では β を推定できないことがわかる。なぜなら、OLS 推定を行うと、

$$\begin{aligned} E(\hat{\beta}|X) &= (\sum x_i x_i')^{-1} \sum x_i E(y_i|x_i) \\ &= (\sum x_i x_i')^{-1} \sum x_i \{x_i'\beta + \sigma\lambda(\frac{-x_i'\beta}{\sigma})\} \\ &= \beta + \sigma(\sum x_i x_i')^{-1} \sum x_i \lambda(\frac{-x_i'\beta}{\sigma}) \end{aligned}$$

となり、バイアスを生ずる。

(2) ML 推定

個体 i の打切り前の条件付き密度関数は

$$f(y_i|x_i;\beta,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y_i-x_i'\beta)^2}{2\sigma^2}\right\}=\frac{1}{\sigma}\phi(\frac{y_i-x_i'\beta}{\sigma})$$

であり、

$$\begin{aligned}P(x_i'\beta+\epsilon_i\geq 0|x_i) &= 1-P(x_i'\beta+\epsilon_i<0|x_i)\\ &= 1-\Phi(\frac{-x_i'\beta}{\sigma})\\ &= \Phi(\frac{x_i'\beta}{\sigma})\end{aligned}$$

なので、(69) より打切り条件付き密度関数は

$$f(y_i|x_i,x_i'\beta+\epsilon_i\geq 0;\beta,\sigma^2)=\frac{\frac{1}{\sigma}\phi(\frac{y_i-x_i'\beta}{\sigma})}{\Phi(\frac{x_i'\beta}{\sigma})} \tag{71}$$

である。従って、観測値 (y_i,x_i), $i=1,\cdots,n$ を得たときの対数尤度関数は

$$LL(\beta,\sigma^2)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\sigma^2-\frac{1}{2}\sum_{i=1}^{n}(\frac{y_i-x_i'\beta}{\sigma})^2-\sum_{i=1}^{n}\log\Phi(\frac{x_i'\beta}{\sigma})$$

である。これを最大化することにより、MLE が得られる。一般には、パラメータを変換して考えるのが簡単である。これは CAN 推定量であるが、漸近的性質とその導出は、Hayashi (2000, p515-516) を参照。

7.2.2 Type 1 Tobit model

切断回帰モデルでは、$y_i^*<0$ の個体については、(y_i,x_i) が観測されない状況を考えた。しかし、そのデータが取得される場合もあり、それを Tobit model (censored regression model) という。このモデルに対しても、OLS 推定量はバイアスを生ずる。

（１） Heckman の 2 段階推定

Heckman(1976) は、回帰に基づく 2 段階推定量を提案した。１段階目は $y>0$ の観測値を $y=1$ で置き換えたデータを用意してプロビット ML で β/σ を推定する。それを $\lambda(\frac{-x_i'\beta}{\sigma})$ に代入して $\hat{\lambda}_i=\lambda(-x_i'(\frac{\hat{\beta}}{\sigma}))$ を作り、2 段階目に y_i が正の観測値のみを用いて (70) に基づいて y_i を $x_i,\hat{\lambda}_i$ に OLS 回帰することによって β の推定量を得る、という方法である。これは CAN 推定量である。詳細は Amemiya(1985), Sec. 10.4.3, p.368-を参照。

（２） MLE

y_i の分布関数は

$$\begin{aligned}P(y_i\leq y) &= P(y_i\leq y|y_i^*\geq 0)P(y_i^*\geq 0)+P(y_i\leq y|y_i^*<0)P(y_i^*<0)\\ &= \begin{cases}0 & for\ y<0\\ P(y_i^*\leq 0) & for\ y=0\\ P(y_i^*<0)+P(y_i\leq y|y_i^*\geq 0)P(y_i^*\geq 0) & for\ y>0\end{cases}\end{aligned}$$

である。更に (71) と $P(y_i^*\geq 0)=P(x_i'\beta+\epsilon_i\geq 0)=1-\Phi(-\frac{x_i'\beta}{\sigma})=\Phi(\frac{x_i'\beta}{\sigma})$ を用いて、

$$P(y_i\leq y)=\begin{cases}0 & for\ y<0\\ \Phi(-\frac{x_i'\beta}{\sigma}) & for\ y=0\\ \Phi(-\frac{x_i'\beta}{\sigma})+\frac{1}{\sigma}\int_0^y\phi(\frac{u-x_i'\beta}{\sigma})du & for\ y>0\end{cases}$$

となる。ここから、サンプル全体に対する条件付き尤度関数は $n_1 = \sum_{i=1}^{n} I(y_i > 0)$ として

$$
\begin{aligned}
LL(\beta, \sigma^2) &= -\frac{n_1}{2}\log(2\pi) - \frac{n_1}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n} 1(y_i > 0)(\frac{y_i - x_i'\beta}{\sigma})^2 - \sum_{i=1}^{n} 1(y_i = 0)\log\Phi(\frac{x_i'\beta}{\sigma}) \\
&= -\frac{n_1}{2}\log(2\pi) - \frac{n_1}{2}\log\sigma^2 - \frac{1}{2}\sum_{i:y_i>0}(\frac{y_i - x_i'\beta}{\sigma})^2 - \sum_{i:y_i=0}\log\Phi(\frac{x_i'\beta}{\sigma})
\end{aligned}
$$

これを最大化することによって MLE を得ることができ、それは CAN である。

7.2.3 その他の Tobit model

切断がどのような変数に対して生じるか、何が観測されるかによって色々な Tobit model を考えることができる。

$$
\begin{aligned}
y_{1i}^* &= x_{1i}'\beta_1 + \epsilon_{1i} \\
y_{2i}^* &= x_{2i}'\beta_2 + \epsilon_{2i}
\end{aligned}
$$

で、$*$ のついた変数は観測されない変数とする。
(1) Type 2 Tobit model

$$
y_i = \begin{cases} y_{2i}^* & if\ y_{1i}^* \geq 0 \\ 0 & if\ y_{1i}^* < 0 \end{cases}
$$

(2) Type 3 Tobit model

$$
\begin{aligned}
y_{1i} &= \begin{cases} y_{1i}^* & if\ y_{1i}^* \geq 0 \\ 0 & if\ y_{1i}^* < 0 \end{cases} \\
y_{2i} &= \begin{cases} y_{2i}^* & if\ y_{1i}^* \geq 0 \\ 0 & if\ y_{1i}^* < 0 \end{cases}
\end{aligned}
$$

これらのモデルは原則的に ML 法で推定されるが、他にも実際のデータに合わせて様々なモデル、推定法が提案されている。

8 パネル分析

8.1 パネルデータ

時系列方向とクロスセクション方向の両方に広がりをもつデータをパネルデータという。よく知られたパネルデータとしては、NLS(National Longitudinal Survey of Labor Market Experience), PSID(Michigan Panel Study of Income Dynamics), 日本では慶應大学が収集している日本家計パネル調査などがある。同一個体を時系列的に追跡しながらデータを集める必要があるため、データ収集にはコストがかかる。収集が少し簡易なものに、疑似パネルデータと呼ばれるものがあるが、これはクロスセクションデータを繰り返しとったものである。これは、観測される個体が毎年変わる点がパネルデータと根本的に異なる。また、パネルデータの中には期間の途中でサンプルから脱落したり、途中からサンプルに入ってくる個体もある。そのようなパネルデータを、区別して unbalanced panel data という。また、対応して普通のパネルデータを balanced panel と呼ぶ。実際は、ほとんどすべてのパネルデータは unbalanced である。

8.2 パネルモデル

i、t をそれぞれ個体、時点を表わす添え字として、パネルデータ分析する際に用いるモデルとしては、普通の回帰モデルと同様に以下のような線形モデルが考えられる。

$$y_{it} = \alpha_{0i} + x_{it}'\beta_0 + \epsilon_{it},\ i = 1, \cdots, n,\ t = 1, \cdots, T \tag{72}$$

ここで、y_{it} はスカラーの被説明変数、α_{0i} は個体 i に特有の効果、x_{it} は定数項を含まない k 次元の説明変数、ϵ_{it} は誤差項である。α_{0i} は次のように解釈される。時間と共に変化せず、また観測されない変量 z_i、関数 g があって本来的には

$$y_{it} = g(z_i) + x_{it}'\beta_0 + \epsilon_{it}$$

によって y_{it} が決まっているが、まとめて $\alpha_{0i} = g(z_i)$ とおいて推定モデルとしている。また、$E(\epsilon_{it}|x_{it}) = 0$ であるとする。このモデルでは通常興味対象となるパラメータは β_0 である。

以下の行列を定義する。

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ \vdots \\ y_{iT} \end{bmatrix},\ \mathrm{i} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{T\times 1},\ X_i = \begin{bmatrix} x_{i1}' \\ x_{i2}' \\ x_{i3}' \\ \vdots \\ x_{iT}' \end{bmatrix}_{T\times k},\ \epsilon_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \vdots \\ \epsilon_{iT} \end{bmatrix}$$

として、

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}_{nT\times 1},\ D = \begin{bmatrix} \mathrm{i} & 0 & 0 & \cdots & 0 \\ 0 & \mathrm{i} & 0 & \cdots & 0 \\ 0 & 0 & \mathrm{i} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathrm{i} \end{bmatrix}_{nT\times n},\ X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{bmatrix}_{nT\times k},\ \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

とする。$Z = [D\ X]$,$\alpha = (\alpha_1, \cdots, \alpha_n)'$ とすると、データに関して

$$Y = D\alpha + X\beta + \epsilon$$

と書ける。これを LSDV(Least squares dummy variables) モデルと呼ぶ。

8.2.1 固定効果モデル

個体 i に特有の効果 α_{0i} が定数（非確率的）である場合、これは単に定数項が各個体で異なる回帰モデルである。これを固定効果モデル（Fixed effect model）という（注：この定義は後で再考）。そのとき、OLS 推定が可能である。OLS 推定量は

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (Z'Z)^{-1}Z'Y = \begin{bmatrix} D'D & D'X \\ X'D & X'X \end{bmatrix}^{-1} \begin{bmatrix} D'Y \\ Z'Y \end{bmatrix}$$

で与えられる。分割行列の逆行列の公式を用いると、$M_D = I - D(D'D)^{-1}D'$ として

$$\begin{aligned} \hat{\beta}_{FE} &= (X'M_DX)^{-1}X'M_DY \\ &= \{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\}^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)y_{it} \\ \hat{\alpha}_{FE} &= (D'D)^{-1}D'(y - X\hat{\beta}_{FE}) \end{aligned}$$

となる。ただし $\bar{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$ である。これを Fixed effect (FE) estimator、または within estimator（個体平均の周りでの変動を使った推定）という。この推定量は、(72) を各個体ごとに t 方向に平均をとったもの

$$\bar{y}_i = \alpha_{0i} + \bar{x}_i'\beta_0 + \bar{\epsilon}_i \tag{73}$$

を (72) から引いて α_{0i} を消去したモデル

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta_0 + \epsilon_{it} - \bar{\epsilon}_i \tag{74}$$

に最小二乗法を適用したものになっている。

次の定理で FE 推定量の性質を述べる。パネル推定の漸近理論を構築する場合、n を大きくする漸近理論と T を大きくする漸近理論、両方を大きくする漸近理論が考えられる。多くのパネルデータは T 方向よりも n 方向に長いため、以下では T は固定、$n \to \infty$ の漸近論を考える。

Theorem 25. *FE estimator* の漸近的性質

(a) モデル *(72)* が正しい特定化で、α_{0i} は定数（非確率的）である。また、(y_i, X_i), $i = 1, \cdots, n$ は *i.i.d.* であるとする。

(b) $rank(E\{\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\}) = k$

(c) $E(\epsilon_i|X_i) = 0$

(d) $E(\epsilon_i\epsilon_i'|X_i) = \sigma_\epsilon^2 I_T$

以上の仮定の下で、$n \to \infty$ のとき、

(i) $\hat{\beta}_{FE} \xrightarrow{p} \beta_0$

(ii) $\sqrt{n}(\hat{\beta}_{FE} - \beta_0) \xrightarrow{d} N(0, \sigma_\epsilon^2 T^{-1}V^{-1})$, $V = E\{T^{-1}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\}$

(iii) $\hat{V} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \xrightarrow{p} V$, $s^2 = \frac{1}{nT-n}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{\alpha}_{iFE} - x_{it}'\hat{\beta}_{FE})^2 \xrightarrow{p} \sigma_\epsilon^2$

である。

8.2.2 変量効果モデル

個体 i に特有の効果 α_{0i} が確率変数である場合、以下添え字の 0 を落として書くことにして、$E(\alpha_i) = \alpha_0$, $u_i = \alpha_i - \alpha_0$ として、(72) を次のように書き換えることができる。

$$\begin{aligned} y_{it} &= \alpha_0 + x_{it}'\beta_0 + u_i + \epsilon_{it} \\ &= w_{it}'\gamma_0 + u_i + \epsilon_{it},\ i = 1, \cdots, n,\ t = 1, \cdots, T \end{aligned} \tag{75}$$

ただし、$w_{it} = (1, x'_{it})'$, $\gamma_0 = (\alpha_0, \beta'_0)'$ である。これを変量効果モデル (Random effect model) という（この定義は後に再考)。

もし u_i と x_{it} が相関をもてば LS 法は適用できないが、ひとつの解決法は IV 推定の適用である。また、α_i は確率変数であるけれども FE 推定によって内生性の問題を回避して β_0 の一致推定量を得ることができる（後述)。

もし u_i が x_{it} と無相関なら、y_{it} を被説明変数、x_{it} を説明変数とする LS によって β_0 の推定が可能である。OLS でも一致推定は可能であるが、モデルの構造上誤差項 $\eta_{it} = u_i + \epsilon_{it}$ は以下のような自己相関をもつため、GLS によって効率的な推定が可能である。u_i と ϵ_{it} は X を条件づけたとき無相関であるとして、

$$\begin{aligned} E(\eta_{it}^2|X_i) &= \sigma_u^2 + \sigma_\epsilon^2 \\ E(\eta_{it}\eta_{is}|X_i) &= \sigma_u^2 \quad (t \neq s) \\ E(\eta_{it}\eta_{js}|X_i, X_j) &= 0 \quad (i \neq j) \end{aligned}$$

よって、$\eta_i = (\eta_{i1}, \cdots, \eta_{iT})'$ として、

$$\Sigma = E(\eta_i\eta_i'|X_i) = \begin{bmatrix} \sigma_u^2+\sigma_\epsilon^2 & \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2+\sigma_\epsilon^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2+\sigma_\epsilon^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 & \sigma_u^2+\sigma_\epsilon^2 \end{bmatrix}_{T\times T} \tag{76}$$

さらに、$\eta = (\eta_1, \cdots, \eta_n)'$ として、

$$\Omega = E(\eta\eta'|X) = \begin{bmatrix} \Sigma & 0 & 0 & \cdots & 0 \\ 0 & \Sigma & 0 & \cdots & 0 \\ 0 & 0 & \Sigma & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \Sigma \end{bmatrix}_{nT\times nT}$$

となる。OLS 推定量は、$W_i = (w_{i1}, w_{i2}, \cdots, w_{iT})'$ として

$$\hat{\gamma}_{OLS} = (\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}w'_{it})^{-1}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}y_{it}$$

であるから、$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}w'_{it} \xrightarrow{p} Q_1$, $\frac{1}{nT}\sum_{i=1}^{n} W'_i\Sigma W_i \xrightarrow{p} Q_2$ とすると、適当な条件のもとで

$$\hat{\gamma}_{OLS} \xrightarrow{p} \gamma_0$$

$$\sqrt{n}(\hat{\gamma}_{OLS} - \gamma_0) \xrightarrow{d} N(0, Q_1^{-1}Q_2Q_1^{-1})$$

を得る。これを pooled OLS 推定という。FE estimator に対応して、Random effect estimator と呼ばれるのは、上のモデルに対する GLS 推定量である。それは、以下で定義される。

$$W = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ \vdots \\ W_n \end{bmatrix}_{nT\times(k+1)}, \quad \Omega^{-1} = \begin{bmatrix} \Sigma^{-1} & 0 & 0 & \cdots & 0 \\ 0 & \Sigma^{-1} & 0 & \cdots & 0 \\ 0 & 0 & \Sigma^{-1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \Sigma^{-1} \end{bmatrix}$$

として、

$$\hat{\gamma}_{RE} = (W'\Omega^{-1}W)^{-1}W'\Omega^{-1}Y$$

Theorem 26. *RE estimator* の漸近的性質

(a) モデル *(75)* が正しい特定化で、α_i は観測されない確率変数である。また、(y_i, X_i, α_i), $i = 1, \cdots, n$ は *i.i.d.* であるとする。

(b) $rank(E\{W_i'\Sigma^{-1}W_i\}) = k+1$

(c) $E(\epsilon_i|W_i, u_i) = 0$

(d) $E(\epsilon_i\epsilon_i'|W_i, u_i) = \sigma_\epsilon^2 I_T$

(e) $E(u_i|W_i) = 0$

(f) $E(u_i^2|W_i) = \sigma_u^2$

(g) $E(u_i\epsilon_i|W_i) = 0$

以上の仮定の下で、$n \to \infty$ のとき、

(i) $\hat{\gamma}_{RE} \xrightarrow{p} \gamma_0$

(ii) $\sqrt{n}(\hat{\gamma}_{RE} - \gamma_0) \xrightarrow{d} N(0, T^{-1}V^{-1})$, $V = E(T^{-1}W_i'\Sigma^{-1}W_i)$

である。

実際には Σ は未知なので、この推定量は feasible でない。Σ に含まれる未知パラメータは σ_u^2, σ_ϵ^2 のみであるから、これらの一致推定量が得られればよい。そこで、以下のような２段階推定法が提案されている。

１）$y_{it} = w_{it}'\gamma + u_i + \epsilon_{it} = w_{it}'\gamma + \eta_{it}$ を pooled OLS 推定し、$\hat{\gamma}_{OLS}$ と残差 $\hat{\eta}_{it}$ を得る。それを用いて $E(\eta_{it}^2) = \sigma_\eta^2 = \sigma_u^2 + \sigma_\epsilon^2$ の推定量

$$\hat{\sigma}_\eta^2 = \frac{1}{nT-k-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\eta}_{it}^2$$

を構成することができる。また、$E(\eta_{it}\eta_{is}) = \sigma_u^2 (t \neq s)$ なので、$t \neq s$ なる $\hat{\eta}_{it}, \hat{\eta}_{is}$ の積和を集めて

$$\hat{\sigma}_u^2 = \frac{2}{nT(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{\eta}_{it}\hat{\eta}_{is}$$

によって σ_u^2 の推定値を得ることができる。これらから、

$$\hat{\sigma}_\epsilon^2 = \hat{\sigma}_\eta^2 - \hat{\sigma}_u^2$$

によって σ_ϵ^2 の推定値を得る。

２）上でえた $\hat{\sigma}_u^2, \hat{\sigma}_\epsilon^2$ を (76) に代入して $\hat{\Sigma}$、更にそれを対角に並べて $\hat{\Omega}$ を作り、そこから

$$\hat{\gamma}_{FRE} = (W'\hat{\Omega}^{-1}W)^{-1}W'\hat{\Omega}^{-1}Y$$

を計算する。

この推定量は Theorem 26 と同じ漸近特性を持つ。

8.2.3 固定効果モデルと変動効果モデルの再定義

上の２つの節で変動効果モデルと固定効果モデルを α_i が確率的であるか、非確率的であるかによって定義した。歴史的にはそのような分け方でパネル分析が始められた。しかし、近年はどちらの推定量を使うべきかでモデルを定義する方が主流である。8.2.2 節でふれたように、もし変動効果モデル（α_i が確率的である）で、α_i と x_{it} に相関がある場合は、RE 推定量は使えない。しかし、FE 推定量は β_0 の一致推定量を与える。すると、推定法に関して言えば、α_i が確率的であるか否かではなく、α_i と x_{it} に相関があるか否かで FE 推定を行うべきか、RE 推定を行うべきかが決まるわけである。その意味で、現在よく用いられる定義は、

- 固定効果モデル：α_i と x_{it} に相関があるパネルモデル
- 変動効果モデル：α_i と x_{it} に相関がないパネルモデル

というものである。

旧定義		推定法	新定義
FE model	α_i : 固定された定数	FE 推定	FE model
RE model	α_i : 確率変数、$cov(\alpha_i, x_{it}) \neq 0$		
	α_i : 確率変数、$cov(\alpha_i, x_{it}) = 0$	RE 推定	RE model

表 1: 固定効果モデル・変量効果モデルの定義

現在の計量経済学のテキストでは旧定義、新定義共に用いられているように見受けられるが、どちらかというと新定義を採用する方が多いようである。また、実際には α_i が非確率的な定数という状況は経済データにおいては適当でない場合が多いため（？）、その場合は除外して、単に $cov(\alpha_i, x_{it}) = 0$ が成立するか否かのみで FE model と RE model が定義されているのが普通である。

8.2.4 固定効果モデル vs 変量効果モデルーハウスマン検定

実際のデータについてパネル分析を行う際には固定効果モデルと変量効果モデルのどちらが良いか（FE 推定と RE 推定のどちらが適切か）を定めなければならない。FE 推定はパラメータ数が多いために一般に効率性において劣るため、RE が適切ならそちらを使うのが望ましい。帰無仮説と対立仮説を

$$
\begin{aligned}
H_0: \ Cov(\alpha_i, x_{it}) &= 0 \\
H_1: \ Cov(\alpha_i, x_{it}) &\neq 0
\end{aligned}
$$

として、二つの推定量は下の表のような性質をもつ。

	H_0	H_1
$\hat{\beta}_{RE}$	Consistent & Eefficient	INconsistent
$\hat{\beta}_{FE}$	Consistent & INfficient	Consistent

表 2: RE vs FE のハウスマン検定

これは IV の妥当性の検定と同様の構造になっており、この場合ハウスマン検定が適用できる。検定統計量は

$$
n(\hat{\beta}_{FE} - \hat{\beta}_{RE})'\{V\hat{a}r(\hat{\beta}_{FE}) - V\hat{a}r(\hat{\beta}_{RE})\}^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \xrightarrow{d} \chi^2(k)
$$

である。

8.2.5 発展

発展としては、

- $N \to \infty, T \to \infty$ の漸近理論
- 動学的パネルモデル $y_{it} = \alpha_i + \gamma_0 y_{i,t-1} + x'_{it}\beta_0 + \epsilon_{it}$
- パネル単位根

9　ノンパラメトリック、セミパラメトリック法

パラメトリックモデルを想定し、その下で興味あるパラメータを推定するのは効率性の観点からは最もよい統計的分析である。具体的には、モデルが正しい時には一致性、漸近正規性、効率性のある推定が可能である。しかし、仮定したパラメトリックモデルが間違っているときには、以下に示すように一般には推定量の一致性すら失われる。つまり、仮定の強弱によって統計分析の効率性と頑健性（robusutness）のトレードオフがある。パラメトリック法に対して、できるだけ仮定を置かないようにするアプローチがノンパラメトリック法である。以下では、密度関数や回帰関数について、一定の滑らかさのみを仮定して推定を行う方法を紹介する。ノンパラメトリック法による推定量は非常に緩い仮定の下で一致性を有するが、通常収束のオーダーがパラメトリック法よりも遅い。一般にパラメトリック推定量は$n^{-1/2}$のオーダーを持つが、ノンパラメトリック推定量は$n^{-1/2+\delta}$のオーダーである。すなわち、収束のオーダーが遅い。

9.1　パラメトリックモデルの特定化の誤り

上に書いたように仮定したパラメトリックモデルが間違っているときには、一般には推定量の一致性すら失われる。単純な一例として、単回帰モデルを考える。真の構造が

$$y_i = x_i^2 + \epsilon_i,\ \ \epsilon_i | x_i \sim (0, \sigma_\epsilon^2)$$

で、更に$x_i \sim (0, \sigma_x^2)$、x_iの分布は左右対称であるとする。このモデルに誤って線形モデル

$$y_i = \beta x_i + u_i$$

をあてはめて OLS 推定を行ったとき、

$$\begin{aligned} \hat{\beta} &= (\sum x_i^2)^{-1} \sum x_i y_i = (\sum x_i^2)^{-1} \sum x_i^3 + (\sum x_i^2)^{-1} \sum x_i \epsilon_i \\ &= (\frac{1}{n} \sum x_i^2)^{-1} \frac{1}{n} \sum x_i^3 + (\frac{1}{n} \sum x_i^2)^{-1} \frac{1}{n} \sum x_i \epsilon_i \\ &\xrightarrow{p} \quad \sigma_x^{-2} \times 0 \quad + \quad \sigma_x^{-2} \times 0 = 0 \end{aligned}$$

となる。結果、xはyに影響を与えないという（間違った）結論を得ることになる。

　そこで、パラメトリックな仮定をおかない推定、検定法があればよいが、その第一ステップがノンパラメトリック密度推定である。

9.2　ノンパラメトリック密度推定

同時密度関数$f(x)$をもつd次元確率変数Xを考える。その分布から無作為標本$\{X_i\}, i = 1, \cdots, n$が得られたとき、密度関数を推定する問題を考える。これは統計学的に本質的な問題である。なぜなら、同時密度関数が推定されれば、その確率変数に関するどのような量（平均、分散、分位点、条件付き分布、条件付き期待値、その他）でも計算できるからである。最も初歩的な推定法はヒストグラムである。

9.2.1　ヒストグラム

単純なケースとして$d = 1$の場合を考える。ヒストグラムの始点x_0、バンド幅（ビン幅）$h > 0$、ビンの数Jを適当に決めて、

$$[x_0, x_0 + h),\ [x_0 + h, x_0 + 2h), \cdots, [x_0 + (J-1)h, x_0 + Jh)$$

の各区間の中に含まれるデータ数を数えて棒グラフにしたものをヒストグラムという。式で書くと次のようになる。I_k を k 番目の区間 $[x_0+(k-1)h, x_0+kh)$ とする。そのとき、区間 I_k に含まれる点 x における密度 $f(x)$ のヒストグラム推定値は

$$\hat{f}(x) = \frac{\{\# \, of \, observations \, in \, I_k\}}{nh}$$

である。ヒストグラムは直観的で簡単であるため、データの分布を概観するには良い手法である。しかし、以下のような問題点がある。

(1) 始点 x_0 の取り方によって大きく印象が変わることがある。

(2) ３変量以上では描けない。

(3) 各区間の端に近い点での推定値はよくないかもしれない。

(4) 連続でない（微分できない；微分はほとんど至る所０になる）。

9.2.2 Naive estimator (NE)

ヒストグラムの問題点 (1)、(3) に対処した推定量が NE である。それは、ヒストグラムと違ってビンを固定せず、

$$\hat{f}(x) = \frac{\{\# \, of \, observations \, in \, [x-\frac{h}{2}, x+\frac{h}{2})\}}{nh}$$

によって定義される。この推定量は、x の左右 $h/2$ の区間に入っている観測値の数を数えて密度の推定値を作っているので、h を小さくすると x のすぐ近くの観測値のみ用いることになり、逆に h を大きくすると x から離れた観測値も用いることになる。大きい h では、当然使われるデータ数が増えるので分散が小さくなるが、x から離れた値を $f(x)$ の推定に用いることになってしまい、それがバイアスの増加となって現れる。逆に h を小さくとると、バイアスは小さくなるがデータ数が減るために分散が上昇する。つまり、h の大小によって分散ーバイアスのトレードオフが生ずる。

ヒストグラムでは先にビンを決めてしまうが、NE は x に対応させて ”ビン ”(区間) を決める。この推定量は、以下のように書き換えられる。

$$\begin{aligned}
\hat{f}(x) &= \frac{1}{n}\sum_{i=1}^{n} I(x-\frac{h}{2} \le X_i \le x+\frac{h}{2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} I(-\frac{1}{2} \le \frac{X_i - x}{h} \le \frac{1}{2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} I(|\frac{X_i - x}{h}| \le \frac{1}{2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} w(\frac{x - X_i}{h})
\end{aligned}$$

ただし

$$w(u) = \begin{cases} 1 & if \ |u| \le \frac{1}{2} \\ 0 & if \ |u| > \frac{1}{2} \end{cases}$$

この推定量は (1), (3) の問題点の解決にはなっているが、(4) は残っている。それを解決しようとするのがカーネル推定量である。

9.2.3 カーネル密度推定量

$w(u)$ を一般化して、

$$\int K(u)du = 1, \quad K(u) = K(-u)$$

を満たす関数 $K(u)$ を用いて、

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h}) \tag{77}$$

をカーネル密度推定量という。h はバンド幅、平滑化パラメータなどと呼ばれ、$K(u)$ はカーネル関数と呼ばれる。カーネル推定量は原理的に NE と同じであるため、h の選択によって NE と全く同様のバイアスー分散のトレードオフが起こる。

多変量のカーネル密度推定量は (77) の自然な拡張として

$$\hat{f}(x_1, x_2, \cdots, x_d) = \frac{1}{nh_1h_2\cdots h_d}\sum_{i=1}^{n} K(\frac{x_1 - X_{1i}}{h_1}, \frac{x_2 - X_{2i}}{h_2}, \cdots, \frac{x_d - X_{di}}{h_d})$$

によって与えられる。多変量カーネル関数を 1 変量カーネル関数の積で置き換えることも可能で、

$$\hat{f}(x_1, x_2, \cdots, x_d) = \frac{1}{nh_1h_2\cdots h_d}\sum_{i=1}^{n} K(\frac{x_1 - X_{1i}}{h_1})K(\frac{x_2 - X_{2i}}{h_2})\cdots K(\frac{x_d - X_{di}}{h_d})$$

としてもよい。

9.2.4 カーネル密度推定量の漸近的性質

証明に以下の結果を用いる。

Lemma 5. *Dominated convergence theorem*

$g_n(x)$ を S 上で定義された関数とし、$g_n(x) \to g(x)$ であるとする。また、$x \in S$ に対して $|g_n(x)| \le m(x)$、$\int_S m(x)dx < \infty$ を満たす関数 $m(x)$ があるとき、

$$\lim_{n\to\infty}\int_S g_n(x)dx = \int_S g(x)dx$$

が成り立つ。

Theorem 27. *Consistency*

(i)$f(x)$ は 2 回連続微分可能である。

(ii) カーネル関数 K は有界なサポート $[-c, c]$ をもち、$0 \le K(u) \le C$, $K(u) = K(-u)$, $\int K(u)du = 1$ である。 また、$\mu_2 = \int u^2 K(u)du$,
$\kappa = \int K(u)^2 du$ とする。

(iii) バンド幅 $h > 0$ について、$n \to \infty$ のとき $h \to 0$, $nh \to \infty$ である。

(i)-(iii) が満たされているとき、x を $f(x)$ のサポートの内点として、$n \to \infty$ のとき

$$E\{\hat{f}(x) - f(x)\}^2 = \frac{1}{4}h^4\mu_2^2 f''(x)^2 + \frac{\kappa f(x)}{nh} + o(h^4 + \frac{1}{nh})$$

が成立する。したがって、$\hat{f}(x) \xrightarrow{p} f(x)$ が成立する。

（証明） まず、$\hat{f}(x)$ の期待値と分散を評価する。iid の仮定と $K(u)$ の対称性を用いて、$\lambda \in [0,1]$ について

$$\begin{aligned}
E\{\hat{f}(x)\} &= E\{\frac{1}{h}K(\frac{x-X_1}{h})\} = E\{\frac{1}{h}K(\frac{X_1-x}{h})\} = \int \frac{1}{h}K(\frac{y-x}{h})f(y)dy \\
&= \int K(u)f(x+hu)du \\
&= \int K(u)\{f(x)+huf'(x)+\frac{(hu)^2}{2}f''(x+\lambda hu)\}du \\
&= f(x)+\frac{h^2}{2}\int u^2K(u)f''(x+\lambda hu)du
\end{aligned}$$

と書ける。x を固定された点と見て、仮定 (i)、(ii) より $-c \le u \le c$ において十分大きな n について $|f''(x+\lambda hu)| < M < \infty$ が成立し、従って

$$|u^2K(u)f''(x+\lambda hu)| \le Mu^2K(u)$$

となり、また $u^2K(u)$ は $K(u)$ のサポートの有界性より可積分である。更に、

$$u^2K(u)f''(x+\lambda hu) \to u^2K(u)f''(x)$$

であるから、Proposition 10 より

$$\int u^2K(u)f''(x+\lambda hu)du \to \int u^2K(u)f''(x)du = \mu_2 f''(x) \tag{78}$$

となる。従って、

$$E\{\hat{f}(x)\} = f(x)+\frac{h^2}{2}\mu_2 f''(x)+o(h^2) \tag{79}$$

を得る。次に分散を評価する。iid の仮定より、

$$V\{\hat{f}(x)\} = \frac{1}{n}\left[E\{\frac{1}{h}K(\frac{x-X_1}{h})\}^2 - [E\{\hat{f}(x)\}]^2\right] \tag{80}$$

で、右辺 [ ] の中の第二項は (79) の 2 乗である。第一項は

$$\begin{aligned}
E\{\frac{1}{h}K(\frac{x-X_1}{h})\}^2 &= \frac{1}{h^2}\int K(\frac{y-x}{h})^2 f(y)dy \\
&= \frac{1}{h}\int K(u)^2 f(x+hu)du \\
&= \frac{1}{h}\int K(u)^2\{f(x)+huf'(x)+\frac{(hu)^2}{2}f''(x+\lambda hu)\}du \\
&= \frac{1}{h}\{\kappa f(x)+\frac{h^2}{2}\int u^2K(u)^2 f''(x+\lambda hu)du\} \\
&= \frac{1}{h}\kappa f(x)+O(h) \qquad (81)
\end{aligned}$$

となる。最後の等号は、$K(u)$ の有界性と (78) を用いた。(80) に (79)、(81) を代入して

$$V\{\hat{f}(x)\} = \frac{\kappa f(x)}{nh}+o(\frac{1}{nh})$$

を得る。従って、

$$MSE(\hat{f}(x)) = E\{\hat{f}(x)-f(x)\}^2 = \frac{1}{4}h^4\mu_2^2 f''(x)^2+\frac{\kappa f(x)}{nh}+o(h^4+\frac{1}{nh}) \tag{82}$$

となる。また、(iii) より $MSE(\hat{f}(x)) \to 0$ なので、

$$\hat{f}(x) \xrightarrow{p} f(x)$$

が成立する。

この定理は各点での一致性を保証しているが、条件を少し強めることによってある閉区間 D 上の uniform consistency

$$\sup_{x \in D} |\hat{f}(x) - f(x)| \xrightarrow{p} 0$$

が証明される（詳細は例えば Li and Racine (2005) Nonparametric Econometrics, p.31, Theorem 1.4, 1.5 を参照）。

次に、カーネル推定量に関する中心極限定理を証明なしに述べる（詳細は例えば Li and Racine (2005) Nonparametric Econometrics, p.29, Theorem 1.3 を参照）。

Theorem 28. *Asymptotic normality*
(i)$f(x)$ は３回連続微分可能である。
(ii) カーネル関数 K は有界なサポート $[-c, c]$ をもち、$0 \le K(u) \le C$, $K(u) = K(-u)$, $\int K(u)du = 1$ である。また、$\mu_2 = \int u^2 K(u)du$ とする。
(iii) バンド幅 $h > 0$ について、$n \to \infty$ のとき $nh \to \infty$, $nh^5 \to 0$ である。

(i)-(iii) が満たされているとき、x を $f(x)$ のサポートの内点として、$n \to \infty$ のとき

$$\sqrt{nh}\{\hat{f}(x) - f(x)\} \xrightarrow{d} N(0, \kappa f(x))$$

である。

バンド幅の選択は実際にカーネル推定を行う時に厄介な問題である。ひとつの考え方は MSE が小さくなるように h を選ぶやり方である。(82) におけるオーダーの大きな項

$$\frac{1}{4}h^4\mu_2^2 f''(x)^2 + \frac{\kappa f(x)}{nh}$$

を h に関して最小にするように

$$h(x)^* = c(x)n^{-\frac{1}{5}},\ \ c(x) = \left\{\frac{\kappa f(x)}{\mu_2^2 f''(x)^2}\right\}^{\frac{1}{5}}$$

とすることが考えられる。また、これはある特定の点 x の推定において良い選択であって、密度関数全体としては良いかどうかわからない。そこで、大域的に見るために MISE(mean integrated squared error)$\int E\{\hat{f}(x) - f(x)\}^2 dx$ のオーダーの大きい部分

$$\int \{\frac{1}{4}h^4\mu_2^2 f''(x)^2 + \frac{\kappa f(x)}{nh}\}dx$$

を小さくするように

$$h^* = c_1 n^{-\frac{1}{5}},\ \ c_1 = \left\{\frac{\kappa}{\mu_2^2 \int f''(x)^2 dx}\right\}^{\frac{1}{5}}$$

と選ぶことも考えられる。いずれにしても、これらは未知の関数 f を含む表現であるため、そのままでは実現可能でないため、色々な手法が提案されている。

9.3 ノンパラメトリック回帰推定

回帰モデル $y = m(x) + \epsilon$ から無作為標本 $(y_i, x_i),\ i = 1, \cdots, n$ が得られたとする。

$$y_i = m(x_i) + \epsilon_i,\ i = 1, \cdots, n$$

パラメトリックなアプローチでは、例えば $m(x) = \beta' x$ といった線形関係を想定して β を LS 推定する。しかし、特定化に誤りがあると推定は全く意味をもたなくなる。そこで、ノンパラメトリック法では $m(x)$ の形を特定化せずに関数そのものを推定することを考える。(y, x) の同時密度関数を $f(y, x)$、x の密度関数を $f(x)$ とすると、回帰関数の定義は

$$m(x) = \int y f(y|x) dy = \frac{1}{f(x)} \int y f(y, x) dy$$

である。$f(x)$, $f(y, x)$ のノンパラメトリック密度推定量を

$$\hat{f}(x) = \frac{1}{nh_2} \sum_{i=1}^{n} K(\frac{x - X_i}{h_2})$$

$$\hat{f}(y, x) = \frac{1}{nh_1 h_2} \sum_{i=1}^{n} K(\frac{y - y_i}{h_1}) K(\frac{x - x_i}{h_2})$$

とすると、$m(x)$ の自然な推定量は

$$\hat{m}(x) \quad = \quad \frac{1}{\hat{f}(x)} \int y \hat{f}(y, x) dy$$

である。ここで、

$$\begin{aligned}
\int y \hat{f}(y, x) dy &= \int y \frac{1}{nh_1 h_2} \sum_{i=1}^{n} K(\frac{y - y_i}{h_1}) K(\frac{x - x_i}{h_2}) dy \\
&= \frac{1}{nh_2} \sum_{i=1}^{n} \int (y_i + h_1 u) K(u) K(\frac{x - x_i}{h_2}) du \\
&= \frac{1}{nh_2} \sum_{i=1}^{n} K(\frac{x - x_i}{h_2}) y_i
\end{aligned}$$

なので、

$$\hat{m}(x) \quad = \quad \frac{\sum_{i=1}^{n} K(\frac{x - x_i}{h_2}) y_i}{\sum_{i=1}^{n} K(\frac{x - x_i}{h_2})}$$

となる。これを回帰関数の Nadaraya-Watson(NW) カーネル推定量という。一定の条件のもとで、一致性と漸近正規性が証明される。

9.4 セミパラメトリック回帰推定

部分的にパラメータで特徴づけられる統計モデルをセミパラメトリックモデルという。以下にその代表的なものを記す。セミパラメトリックモデルは、パラメトリックな部分とノンパラメトリックな部分の両方を含む統計モデルであるが、そのパラメトリック部分に興味があり、ノンパラメトリック部分は nuisance parameter(局外母数) である場合が多い。多くのセミパラメトリック推定量は、収束が遅いノンパラメトリック部分の推定量を含むにも関わらず $n^{-1/2}$ のオーダーで収束するものが多く、当初は驚くべき結果と考えられていた。

9.4.1 部分線形回帰モデル (Robinson (1988), Econometrica)

これはサンプルセレクションモデル等を含むモデルで、$g(\cdot)$ を未知の関数として

$$y_i = \beta' x_i + g(z_i) + \epsilon_i, \quad i = 1, \cdots, n \tag{83}$$

によって定義される。言うまでもなく、回帰関数

$$E(y_i|x_i, z_i) = \beta' x_i + g(z_i) \tag{84}$$

は x_i については線形である。興味対象となるパラメータは β であるとする。z_i を条件づけて (83) の期待値を取って (83) から引くと、$g(z_i)$ が消えて、

$$y_i - E(y_i|z_i) = \beta' \{x_i - E(x_i|z_i)\} + \epsilon_i$$

を得る。これは線形回帰モデルの形になっており、$E(y_i|z_i)$ と $E(x_i|z_i)$ がわかれば OLS によって β が推定できる。それらの関数はノンパラメトリック回帰 $\hat{E}(y_i|z_i)$、$\hat{E}(x_i|z_i)$ により推定できるため、それらで置き換えて、β を

$$\hat{\beta} = [\sum_{i=1}^{n} \{x_i - \hat{E}(x_i|z_i)\}\{x_i - \hat{E}(x_i|z_i)\}']^{-1} \sum_{i=1}^{n} \{x_i - \hat{E}(x_i|z_i)\}\{y_i - \hat{E}(y_i|z_i)\}$$

によって推定することができる。(y_i, x_i, z_i) が iid で、ϵ_i は x_i, z_i と独立であるとき、この推定量は $\sqrt{n}$– 一致性を持つ推定量であることが示されている。すなわち、ある正値定符号な V に対して

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$$

が導かれる。

9.4.2 インデックスモデル

y が x の線形結合の未知関数 $G(\cdot)$ に依存する回帰モデル

$$y_i = G(\beta^\tau x_i) + \epsilon_i$$

を（シングル）インデックスモデルという。例えば Tobit, Probit, Logit などはこのモデルの特殊ケースである。興味対象のパラメータは β である。このモデルに対して、$\hat{E}(y|\beta^\tau x)$ を" ノンパラメトリック回帰推定量"

$$\hat{E}(y|\beta^\tau x) = \frac{\sum_{j=1}^{n} K(\frac{\beta^\tau x - \beta^\tau x_j}{h}) y_j}{\sum_{j=1}^{n} K(\frac{\beta^\tau x - \beta^\tau x_j}{h})}$$

として

$$\hat{\beta} = \min_{\beta} \sum_{i=1}^{n} \{y_i - \hat{E}(y_i|\beta^\tau x_i)\}^2$$

によって推定することが考えられる。これをインデックスモデルのセミパラメトリック LS 推定量という (Ichimura (1993), Journal of Econometrics)。この目的関数は β に関して非線形なので数値計算が必要になる。

別のより簡便な推定法として Averaged derivatives 推定量 (Härdle & Stoker (1989), JASA) が提案されている。$g(x) = G(\beta^\tau x)$ とおくと

$$g'(x) = G'(\beta^\tau x)\beta$$

なので、

$$E\{g'(x)\} = E\{G'(\beta^\tau x)\}\beta = c\beta$$

となる。c はある未知定数である。従って、$c\beta$ の推定は $E\{g'(x)\}$ の推定と同じである。x の密度を f として、$g(u)f(u)$ が裾で 0 に収束すると仮定すると、

$$\begin{aligned} E\{g'(x)\} &= \int g'(u)f(u)du = [g(u)f(u)]_{-\infty}^{\infty} - \int g(u)f'(u)du \\ &= -\int g(u)\frac{f'(u)}{f(u)}f(u)du \\ &= -E\{g(x)\frac{f'(x)}{f(x)}\} \\ &= -E\{y\frac{f'(x)}{f(x)}\} \end{aligned}$$

f のノンパラメトリック推定量を $\hat{f}$ とすると、これは

$$-\frac{1}{n}\sum_{i=1}^{n} y_i \frac{\hat{f}'(x_i)}{\hat{f}(x_i)}$$

によって推定できる。これらの推定量も、適当な仮定の下で $\sqrt{n}-$ 一致性を持つ推定量であることが示されている。

9.4.3 未知の分散不均一性を持つ回帰 (Robinson (1987), Econometrica)

誤差項の分散が不均一な回帰モデル

$$y_i = \beta' x_i + \epsilon_i, \quad E(\epsilon_i|x_i) = 0, \; V(\epsilon_i|x_i) = \sigma^2(x_i)$$

を考え、$\sigma^2(x)$ は未知関数であるとする。そのとき、次の 2 ステップ FGLS 推定によって、β の効率的な推定量が得られる。

(i) OLS 回帰を行い、残差 $\hat{\epsilon}_i = y_i - x_i'\hat{\beta}_{OLS}$ を計算する。$\sigma^2(x_i) = E(u_i^2|x_i)$ なので、これをノンパラメトリック回帰とみなし、更に u_i を $\hat{u}_i$ で置き換えて、

$$\hat{\sigma}^2(x) = \frac{1}{\hat{f(x)}}\frac{1}{nh^p}\sum_{i=1}^{n} K(\frac{x-x_i}{h})\hat{u}_i^2$$

によって $\sigma^2(x_i)$ の推定量を構成する。

(ii) 上の結果を用いて、以下の FGLS を行う。

$$\hat{\beta} = \{\sum_{i=1}^{n}\frac{(x_i-\bar{x})(x_i-\bar{x})'}{\hat{\sigma}^2(x_i)}\}^{-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})(y_i-\bar{y})}{\hat{\sigma}^2(x_i)}$$

この推定量も $\sqrt{n}-$ 一致性を有する推定量であることが示される。

計量経済学の分野では、1980 年代の終わりころから様々なセミパラメトリックモデルが提案され、その推定法が開発されてきた。その一つの理由は、経済理論は完全にパラメトリックな統計モデルを与えることはできないため、興味ある部分のみをパラメトリックに表現し、それ以外を特定しないモデルが計量経済学は適切であると考えられることである。また、ノンパラメトリックモデルは非常に緩い仮定の下でよい性質を持つが、収束のオーダーが遅いため、経済データのサンプル数が通常多くないことを考慮すると、パラメトリック法と同じ収束オーダーをもつセミパラメトリック法は経済分析に適切であると考えられる。