<研究ノート>

同時方程式モデルにおける二段階最小二乗推定量の厳密分布および2つの近似分布の精度の比較

川　副　延　生

1. はじめに

同時方程式モデルにおいて，2つの内生変数を含む方程式での TSLS 推定量の厳密分布を高速計算機を使って数値評価(*)を行い，それをもとにしてGram-Charlier 型および Monte Carlo 法による近似分布の精度を詳しく調べる。

2. 導入

同時方程式モデルの中で，2つの内生変数を含む方程式の係数を推定する場合の TSLS 推定量の厳密分布は，Anderson and Sawa [3] 1973 により無限二重級数を用いて表されたが，さらに Anderson and Sawa [4] 1979 では厳密分布と Gram-Charlier 型の近似について次の2つの事が述べられている。即ち，

① 二重級数をそのアルゴリズムにより計算して厳密分布を数値評価する場合に，非心パラメータの値が大きいと十分な精度を得る為には多大な計算時間がかかり，実際に計算したのは非心パラメータ値 δ^2 が50まで，標準化された係数パラメータ α が5までの場合である。

② ①の計算によって得られた厳密分布について，Gram-Charlier 型の近似分布による誤差は計算したパラメータ値の範囲内で最も好ましくない場合でさえも，最大で0.001である。

以上の2点を述べて，[4] では広範なパラメータ値についてGram-Charlier

＊ 本稿での計算にあたり，一橋大学情報処理センターの方々にいろいろ御指導いただいたことを感謝申し上げます。

型の近似による多数の Table が示されている。

ところで，1983年に東大大型計算機センターにベクトルプロセッサHITAC S-810-20が導入され，ベクトル処理に適するプログラムについては高速で計算することが可能となった。そこで本稿では［4］では計算されなかった比較的大きなパラメータ値の場合について，無限二重数列の項を十分に多く加算することにより厳密分布を十分な精度で数値評価し,この数値をもとにしてGram-Charlier 型および Monte Carlo 法による近似の精度を詳しく調べ，2つの近似方法の適切さの程度を検討する。

3. モデル，定義(1)および厳密分布の式

方程式

(1) $Y_1=\beta Y_2+Z_1\gamma_1+u$

で，β の TSLS 推定量を b とする。Y_1,Y_2 は2つの内生変数のT次元観測値ベクトルで，Z_1 は K_1 個の外生変数の観測値から成る $T\times K_1$ 行列，β はスカラー，r_1 は K_1 次元パラメータベクトル，u はT次元撹乱項ベクトルである。

同時方程式モデルの誘導形は，

(2) $(Y_1Y_2)=(Z_1Z_2)\begin{pmatrix}\Pi_{11}&\Pi_{12}\\\Pi_{21}&\Pi_{22}\end{pmatrix}+(v_1v_2)$

という式を含む。ここで Z_2 は方程式(1)から除外されている K_2 個の外生変数の $T\times K_2$ 観測値行列，$\Pi ij(i=1,2;j=1,2)$ はそれぞれ誘導形係数のベクトル，(v_1v_2) は $T\times 2$ の誘導形〝撹乱項行列〃である。

以下の3つの仮定をおく。

(仮定1) (v_1v_2) の各行は互いに独立に平均ベクトル0，分散行列 Ω の2変量正規分布に従う。但し，

$$\Omega=\begin{pmatrix}\omega_{11}&\omega_{12}\\\omega_{12}&\omega_{22}\end{pmatrix}$$

とする。

(仮定2) (Z_1Z_2) は既知の $T\times K$ 行列で，$T>K$ でその階数は K である。

(仮定3) $(\Pi_{21}\Pi_{22})$ は階数が1で，Π_{22} は少なくとも1つの0でない要素を含む。

TSLS 推定量 b について，

(3) $\dfrac{\delta\sqrt{\omega_{22}}}{\sigma}(\mathrm{b}-\beta)$

という推定量を考える。[2] ここで，

(4) $\delta^2=\dfrac{\Pi'_{22}(Z'_2Z_2-Z'_2Z_1(Z'_1Z_1)^{-1}Z'_1Z_2)\Pi_{22}}{\omega_{22}}$

(5) $\sigma^2=\omega_{11}-2\beta\omega_{12}+\beta^2\omega_{22}$

とする。

さらに，

(6) $\alpha=\dfrac{\omega_{22}}{\sqrt{|\Omega|}}\left(\beta-\dfrac{\omega_{12}}{\omega_{22}}\right)$

とする。

［3］，［4］によれば，

(7) $\mathrm{P_r}\left\{\dfrac{\delta\sqrt{\omega_{22}}}{\sigma}(\mathrm{b}-\beta)\leqq x\right\}=\mathrm{P}_r\left\{\mathrm{F}''\leqq 1+2r\sqrt{r^2+1}+2r^2\right\}$

$$=\mathrm{e}^{-\frac{1}{2}(1+\alpha^2)\delta^2}\sum_{i,j=0}^{\infty}\frac{\delta^{2(i+j)}\zeta_1^i\zeta_2^j}{2^{i+j}\,i!\,j!}\mathrm{I_g}\left(\frac{\mathrm{K}_2}{2}+i,\ \frac{\mathrm{K}_2}{2}+j\right)$$

が成り立つ。ここで，

(8) $r=\alpha+\dfrac{\sqrt{1+\alpha^2}}{\delta}x$

であり，F″ は非心パラメータがζ_1, ζ_2で自由度が K_2 と K_2 である二重非心F分布に従う確率変数を表す。但し，

(9) $\zeta_1=\dfrac{1}{2}\left[1+\alpha^2+\dfrac{2\alpha-r+r\alpha^2}{\sqrt{r^2+1}}\right]\delta^2$

(10) $\zeta_2=\dfrac{1}{2}\left[1+\alpha^2 \text{〃}-\text{〃};\dfrac{2\alpha-r+r\alpha^2}{\sqrt{r^2+1}}\right]\delta^2$

とする。また，

(11) $\mathrm{I_g}(\mathrm{p},q)=\displaystyle\int_0^{\mathrm{g}}\frac{1}{\mathrm{B}(\mathrm{p},q)}x^{\mathrm{p}-1}(1-x)^{q-1}\mathrm{d}x$

(12) $\mathrm{g}=\dfrac{1}{2}+\dfrac{r}{2\sqrt{\text{〃}r^{2}\text{〃}+1}}$

とする。[3]

本稿では除外外生変数の数 K_2 は，計算時間の短縮の為に偶数に限定する。

4. 厳密分布の数値評価および近似分布との比較

(13) $$p=\frac{K_2}{2}+i$$

(14) $$q=\frac{K_2}{2}+j$$

(15) $$a_{ij}=e^{-\frac{1}{2}(1+\alpha^2)\delta^2}\frac{\delta^{2(i+j)}\zeta_1^i\zeta_2^j(p+q-1)!}{2^{i+j}i!j!(p-1)!(q-1)!}$$
$$\times\left[\frac{1}{p}g^p(1-g)^{q-1}+\sum_{k=1}^{q-1}\frac{(p-1)!(q-1)!}{(p+k)!(q-k-1)!}g^{p+k}(1-g)^{q-k-1}\right]$$

とすると，(11)式で部分積分を繰り返すことにより，

(16) $$P_r\left\{\frac{\delta\sqrt{\omega_{22}}}{\sigma}(b-\beta)\leqq x\right\}=\sum_{i,j=0}^{\infty}a_{ij}$$

が成り立つ。

この無限二重級数について，数列 $\{b_L\}(L=0,1,2,\cdots)$ を次の様に定義する。

(17) $$b_L=\begin{cases}\sum_{j=0}^{L}a_{Lj}+\sum_{i=0}^{L-1}a_{iL} & (L\geqq 1)\\ a_{00} & (L=0)\end{cases}$$

この数列について，適当な整数 L_1, L_2 を選んで，

(18) $$\sum_{L=L_1}^{L_2}b_L$$

を計算することにより，厳密分布の数値評価を十分な精度で行う。[4]

なお選択したパラメータ値は，

(19) $\alpha=0.6,\ 1,\ 2$

(20) $\delta^2=10,\ 50,\ 100,\ 300,\ 500$

(21) $K_2=4,\ 10,\ 30$

の組合せからなる42組[5]であり，それぞれの場合について，厳密分布，Gram-Charlier 型近似分布[6]および Monte Carlo 法による近似分布[7]の3つについて，

(22) $x=-3.0\ (0.5)\ 3.0$

での値を求め比較をした。表1は〝x〟の結果をまとめたものである。

δ^2 ＼ α ＼ K_2		4	10	30
10	0.6	0.0031 10^{-4}	0.0039 10^{-5}	0.0043 10^{-6}
	1	0.0027 10^{-4}	0.0059 10^{-4}	0.0028 10^{-6}
	2	−0.0017 10^{-3}	0.0034 10^{-4}	0.0046 10^{-5}
50	0.6	0.0047 10^{-4}	0.0024 10^{-5}	−0.0021 10^{-5}
	1	0.0044 10^{-4}	−0.0033 10^{-4}	0.0028 10^{-5}
	2	0.0048 10^{-4}	−0.0052 10^{-4}	−0.0021 10^{-5}
100	0.6	0.0036 10^{-4}	−0.0020 10^{-5}	−0.0035 10^{-5}
	1	0.0034 10^{-4}	0.0026 10^{-5}	−0.0027 10^{-5}
	2	0.0033 10^{-4}	−0.0029 10^{-5}	−0.0041 10^{-5}
300	0.6	0.0034 10^{-5}	−0.0037 10^{-5}	−0.0053 10^{-6}
	1	0.0032 10^{-4}	−0.0034 10^{-5}	−0.0046 10^{-6}
	2	0.0030 10^{-5}	−0.0033 10^{-5}	−0.0047 10^{-6}
500	0.6	0.0043 10^{-5}	−0.0033 10^{-5}	−0.0042 10^{-6}
	1	0.0036 10^{-5}	−0.0039 10^{-5}	−0.0053 10^{-6}
	2			

表　1

各ブロックの上側の数字は Monte Carlo 法による誤差の最大値であり，下側の数字は Gram-Charlier 型による誤差の最大の数字について0でない数字が初めて現れる小数位を表している(8)。なお Monte Carlo 法は20000回の繰り返しによっている。

5. 結 論

比較的重要と考えられるパラメータ値のケースでは，Gram-Charlier 型の近似の精度は非常によく，誤差の最大値は $\alpha=2$, $\delta^2=10$, $K_2=4$ のケースで 0.0012 となるだけで他の全てのケースでは 0.001 未満であった。これに対して本稿で行った方法での Monte Carlo 法による近似は，2万回の実験のもとでは Gram-Charlier 型よりもかなり精度が劣り，また $\alpha=1$, $\delta^2=300$ といういわゆる typical value に近いパラメータ値の場合での100万回の実験のもとでも，誤差の最大値を 0.001 未満にしようとすれば少なくとも50万回以上は必要となっている。

付論. 厳密分布計算のプログラム(9)

```
      PARAMETER (K2=4,L1=1,L2=60)
      DIMENSION TSLS(13),F(0:K2+2*L2),
     ?          B(0:L2,0:L2),B4(K2+L2),B5(K2+L2)
      REAL*8 A,D,G,M,R,X,Z1,Z2,F,P,Q,AI,
     ?          TSLS,B,B1,B2,B3,B4,B5
      A=2
      D=SQRT(10.0D0)
      M=SQRT(1+A*A)*D
      F(0)=0
      DO 1 I=1,K2+2*L2
      AI=I
    1 F(I)=F(I-1)+LOG(AI)
      DO 10 IX=1,13
      X=-3.5+0.5*IX
      R=(1+A*A)/M*X+A
      G=0.5*(1+R/SQRT(1+R*R))
      Z1=0.5*(1+A*A+(2*A-R+R*A*A)/SQRT(1+R*R))
      Z2=0.5*(1+A*A-(2*A-R+R*A*A)/SQRT(1+R*R))
      B1=0
      DO 20 L=L1,L2
      B2=0
      P=0.5*K2+L
      DO 30 J=0,L
      Q=0.5*K2+J
   30 B(L,J)=(2*(L+J)*LOG(D)+L*LOG(Z1)+J*LOG(Z2)
     ?        +F(P+Q-1)+P*LOG(G)+(Q-1)*LOG(1-G)+174)
     ?       -((L+J)*LOG(2.0D0)+F(L)+F(J)+F(P-1)
     ?        +F(Q-1)+0.5*M*M+LOG(P))
      DO 40 J=0,L
      B3=0
```

```
      Q=0.5*K2+J
      DO 41  K=1,Q-1
 41   B4(K)=2*(L+J)*LOG(D)+L*LOG(Z1)+J*LOG(Z2)
    ?        +F(P+Q-1)+(P+K)*LOG(G)
      DO 42  K=1,Q-1
 42   B5(K)=((Q-K-1)*LOG(1-G)+174)
    ?       -((L+J)*LOG(2.0D0)+F(L)+F(J)+0.5*M*M
    ?        +F(P+K)+F(Q-K-1))
      DO 43  K=1,Q-1
 43   B3=B3+EXP(B4(K)+B5(K))
 40   B2=B2+EXP(B(L,J))+B3
      Q=0.5*K2+L
      DO 50  I=0,L-1
      P=0.5*K2+I
 50   B(I,L)=(2*(I+L)*LOG(D)+I*LOG(Z1)+L*LOG(Z2)
    ?         +F(P+Q-1)+P*LOG(G)+(Q-1)*LOG(1-G)+174)
    ?        -((I+L)*LOG(2.0D0)+F(I)+F(L)+F(P-1)
    ?         +F(Q-1)+0.5*M*M+LOG(P))
      DO 60  I=0,L-1
      B3=0
      P=0.5*K2+I
      DO 61  K=1,Q-1
 61   B4(K)=2*(I+L)*LOG(D)+I*LOG(Z1)+L*LOG(Z2)
    ?        +F(P+Q-1)+(P+K)*LOG(G)
      DO 62  K=1,Q-1
 62   B5(K)=((Q-K-1)*LOG(1-G)+174)
    ?       -((I+L)*LOG(2.0D0)+F(I)+F(L)+0.5*M*M
    ?        +F(P+K)+F(Q-K-1))
      DO 63  K=1,Q-1
 63   B3=B3+EXP(B4(K)+B5(K))
 60   B2=B2+EXP(B(I,L))+B3
 20   B1=B1+B2
 10   TSLS(IX)=B1*EXP(-174.0D0)
      WRITE(6,100) TSLS
100   FORMAT(1H ,D18.10)
      STOP
      END
```

（注）

(1)　この節のモデルと定義は全て［3］,［4］によっている。

(2)　(3)式の漸近分布は標準正規分布である。

(3)　B(p, q)はベータ関数である。

(4)　本稿で考察するパラメータベクトル（α, δ^2, K_2）の値に限れば，対応して得られる数列 $\{b_L\}$ の集合は2つのグループに分類される。1つは常に単調減少と

なっている数列のグループであり，もう1つは初めは単調に増加するが途中から単調減少に変わる数列のグループである。どちらにしても1つの数列 $\{b_L\}$ について，

(23) $10^{-8} < b_L$

を満たす項の全てから成る集合Bが一意的に対応し，そして，

(24) B“C” $\{b_L \mid L_1 \leqq L \leqq L_2\}$

を満たす様な整教の組（L_1, L_2）をその数列に対応して選ぶことができる。

なお，数列 $\{b_L\}$ はLが十分大きい値のときには単調減少となっているが，その単調性はかなり強いものであり，(23)式を満たさない項を加算しなくても精度には影響を与えないことを確信しているが，あくまでも $\{b_L\}$ の有限個の項の“値の”減少具合から判断したものであり厳密に証明している訳ではない。

また実際に計算するときに用いた（L_1, L_2）は，(24)式で集合Bよりも包含の意味で少しだけ大きい集合を与える様な値を適当に選んだものであり，用いた値を以下の表に示す。なお，この様にして選んだ（L_1, L_2）はK_2の値には余り依存しないでもっぱら（$1+\alpha^2)\delta^2$ の値に依存する。

δ^2 \ α	0.6	1	2
10	(0 , 30)	(0 , 30)	(1 , 60)
50	(1 , 60)	(1 , 90)	(1 , 200)
100	(1 , 100)	(1 , 140)	(1 , 360)
300	(1 , 240)	(160 , 360)	(500 , 900)
500	(160 , 360)	(300 , 600)	

（L_1, L_2） の値

(5) パラメータ値の選択は［2］,［4］を参考にした。特に［2］で，typical value として $\alpha=0.9, \frac{\delta^2}{TK_2}=1.03$ としていることを考慮した。

また，$\alpha=2$, $\delta^2=500$ で $K_2=4, 10, 30$ の3つの場合については，計算時間が多くなり主に予算制約の面から本稿では計算を省略した。

(6) ［4］p. 181 による近似式。

(7) ［1］，および［5］第2節に示されている方法による。即ち，平均ベクトル0，分散行列 $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ の2変量正規分布からサンプルサイズ K_2の標本を取り，それを用いて TSLS 推定量を作り，これを20000回繰り返して近似分布を得る方法である。

本稿では正規乱数を以下の様に作成した。即ち，乗算型合同法（FACOM, SSL II, RANU 2, 初期値＝584287）により$2K_2$個の一様乱数を発生させ，順に逆関数法（FACOM, SSL II, INDF）により$2K_2$個の標準正規乱数を作り，この$2K_2$個の正規乱数を，

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \end{pmatrix}, \cdots, \begin{pmatrix} 2K_2-1 \\ 2K_2 \end{pmatrix}$$

の順に並べて K_2 個のベクトルを作り，これを2変量正規分布からの標本と見なした。

この様な正規乱数ベクトルの作り方は非常に特殊であり，従って上記の様な方法による Monte Carlo 法での近似分布も特殊なものとなっている訳で，ただちに Gram-Charlier 型の近似分布との比較をして結論を出すのは不適切である。

そこで一様乱数を発生させる段階での他の方法として，RANU2 を用いるかわりに乗算型合同法シャッフル型（FACOM，SSLII，RANU3初期値＝584287）を用いて同様にシミュレーションを行い，2つの Monte Carlo法による結果を比較し，RANU2を用いた Monte Carlo 法の結果の特殊性を緩和できるかどうか検討した。

① まず20000回の実験による場合では，表1で考察したパラメータ値の各ケースについて，誤差の最大値のオーダーは全て同じであり，また誤差の最大値の大小からみた2つの方法の優劣は各ケースに応じて立場を変え，どちらの方法が勝れているとは一概に判断できなかった。

② さらに，いわゆる typical value と呼ばれる値に近いパラメータ値である

$$\alpha=1,\ \delta^2=300$$

について $K_2=4$，10，30 とした3つの場合について，それぞれ100万回のシミュレーションを行い（2万回単位に連続させて50回行い，2万回増えるごとに $x=-3.0$（0.5）3.0 の範囲での誤差の最大値を調べた），以下の表を得た。

	0.003	0.002	0.001
$K_2=4$	6万回 6万回	10万回 10万回	68万回 68万回
$K_2=10$	4万回 2万回	10万回 18万回	84万回 56万回
$K_2=30$	4万回 16万回	14万回 22万回	62万回 62万回

この表は次の様に作られている。たとえば $K_2=10$ と 0.002 から決まるブロックで，上側の10万回という数字は RANU2 を用いた場合の誤差の最大値の絶対値が，2万回を単位として，10万回以降100万回までは全て0.002未満だが，1単位まえの8万回の場合には0.002以上になっていることを示し，また下側の18万回という数字は RANU3 を用いた場合について同様の意味を示す。

以上の2点から，RANU2 と RUNU3 との違いによる Monte Carlo 法の優劣は明らかではないと判断でき，本稿での Monte Carlo 法の特殊性が少し緩和されたと判断した。

そこで，RANU2 を用いた特殊な方法による結果を使っているという限界のもとに，本稿では Gram-Charlier 型と Monte Carlo 法による近似の精度を比較している。

(8) 上側の数字は，$x=-3.0$（0.5）3.0 の各値のときに，Monte Carlo 法による近似から厳密分布の値を引いて得られる13個の数字のなかで，その絶対値が最

大となる数字の小数第5位を切り上げた値を示す。下側の数字も同様に，Gram-Charlier 型の近似から厳密分布の値を引いて得られる13個の数字のなかで，その絶対値が最大となる数字についてのオーダーを示したものである。

(9)① このプログラムは $L_1 \geqq 1$ の場合のものである。

② 計算途中でのオーバーフローとアンダーフローのエラーを避けること，およびベクトルプロセッサ向きに DO ループを細かく分割していることの為に少しプログラムが不自然になっている。

③ パラメータ値により異なるが，プログラムのベクトル化率は90%～95%の間である。

④ CPU の計算時間のデータを表に示す。付論のプログラムで DO 10 ループを省き，$x=-3.0$ （0.5） 3.0 の13回の計算をするかわりに $x=0$ として1回の計算をして CPU TIME を得ている。

同じプログラムを以下の3種類の計算機を使って計算しているが，M-360 では最適化レベル3で，M-280H ではコンパイルオプションに opt(3)のみを指定しやはり最適化レベル3で実行している。

一橋大学情報処理センター，FACOM M-360
東大大型計算機センター　，HITAC M-280H
〃　　　　　　　　　　　，HITAC S-810-20

なおパラメータ値は，

$\alpha=2$，$K_2=30$

で $\delta^2=10$，50，100，300，500 の5ケースである。

CPU TIME　（単位 SECOND）

δ^2	(L_1, L_2)	M-360	M-280H	S-810
$\delta^2=10$	(1, 60)	7.44S	2.89S	1.31S
$\delta^2=50$	(1, 200)	132.32S	33.52S	3.72S
$\delta^2=100$	(1, 360)	665.69S	171.90S	12.80S
$\delta^2=300$	(500, 900)	7546.21S	1937.82S	116.35S
$\delta^2=500$	(950, 1400)	23460.12S		338.54S

(参考文献)

〔1〕 Anderson, T.W. "An asymptotic expansion of the distribution of the Limited Information Maximum Likelihood estimate of a coefficient in a simultaneous equation system", *Journal of the American Statistical Association* 69, 1974, pp. 565—573

〔2〕 Anderson, T.W., K. Morimune and T. Sawa, "The numerical values of some key parameters in econometric models", *Journal of Econometrics* 21, 1983, pp. 229—243

〔3〕 Anderson, T.W. and T.Sawa, "Distributions of estimates of co "e"ffients of a single equation in a simultaneous system and their asymptotic expansions", *Econometrica* 41, 1973, pp. 683—714

〔4〕 Anderson, T.W. and T.Sawa, "Evaluation of the distribution function of the Two-Stage Least Square estimate", *Econometrica* 47, 1979, pp. 163—182

〔5〕 森棟公夫 「同時方程式推定量の小標本特性」
竹内啓編「計量経済学の新展開」第２章所収，東大出版会 1983，

（筆者の住所：三鷹市下連雀 3-43-9　山本アパート）