研究ノート

回帰分析の性格

佐藤　保

(一)

計量分析で最もよく使われる方法は単純多元線型回帰の方法である。その形は

(1) $Y=a_0+a_1X_1+a_2X_2+\cdots\cdots+a_pX_p$

である。全体がこの形で通されることはなくても全体の計算の途中であらわれることもしばしばある。同時方程式体系の誘導型としてあらわれることもあるし，非線型の構造方程式体系の途中であらわれることもある。しかし最初からこの形で通すことも多い。筆者は先に種々の測定法の比較を試みたことがあるが，若干の続きを述べるということで，多元回帰の性質にふれておきたい。

ややそれるが，計量経済学による分析という点については二つの面があると考えられる。一つはその理論的側面といわれるものであって，主として計量分析に使われるモデルの数理統計学性質及びその開発といったもので，多くの計量経済学何々といった書物ではその大部分の頁はこのことのためについやされているといっていい。数理的発展は論理的発展といってもよく，つみ重ねであるから近年益々発展されて高度な（という意味は主として高度な数学を使う，あるいはその数学的証明がなされるということであろう）ものとなってきている。従って，そのような数学的知識のない者にとっては，その過程は理解しがたいものになるけれども，応用を主として考える者にとってはその結果について利用できればよいということになる。もう一つの面は実際の資料を使って計算を行うという面であるが，この段階となると比較的簡単な方法が用いられる。種々の測定法の比較について言われることであるが，複雑な方法を使う利点というのは，データーの中に含まれている情報量をより多く使っているということであろう。簡単な方法ほどそれを使っていないといえるが，見方を変えれば簡単な方法ほどそこから得られる結論というか，その論述に対して異論が含まれる余地は少なく，複雑な方法ほど異論をはさむ余地が大きくなるといえる。自然現象では複雑なというか，より精巧な装置ができるといままで不明瞭であったものが明瞭に誰にでもわかることがあるが，計量分析では装置が複雑になるとそれだけ多くの仮定が入ってくるので，すべての人をなっとくさせることはそれだけ困難になるといってもよい。単純多元回帰による最小自乗法が最も良く使われる理由がそこにもあるといってよいであろう。

統計分析の目的は，あるモデルを想定してその母集団からの標本が手元にあるものと考え，その標本から母集団のパラメーター，母数を推定することにある。そしてその推定値が真の母数を推定する上で安定的であることが望ましい。普通はこれは推定値の標準誤差で計られる。標準誤差が有効であるためには推定量の分布がわかっていなければならないが，もし推定量の分布が母数を含まず，標本統計量のみからなり，分布の形（密度函数）が正確にわかれば，正確な信頼区間が計測できるわけで最も望ましいことになる。単純な多元回帰が用いられる第二の大きな理由がここにある。このような理由で単純多元回帰が多く使われ，それにはそれで人をなっとくさせる理由があるが，実際上には更に制約が加わる。一般的に経済変量Yの動きは，

(2) $Y=a_0+a_1X_1+a_2X_2+\cdots+a_iX_i+\cdots\cdots+a_pX_p+u$

で示されるが，抽象的にはp個の独立変数によって示されるといってそれで終りとなるが，実際的にはこれをいくつかにきめなければならない。残差項uの存在については普通次のように説明される。従属変数Y，

例えば消費支出として，それに影響をあたえると思われる要因は抽象的には非常に多く考えられる。実際上はそれらの資料があるかどうかが問題でその面からチェックされるが，それ等がそろったとして，数十の要因が考えられたとしても，それ等の要因（考えられるすべての要因）を投入したとしても，それで完全にその動きをとらえられるわけではない（自由度は充分にあるものとする）。その動き自体（人間の行動自体に）確率的要素があり，それ故に残差項は残るものと考えられている。実際上はすべての要因を導入することはできないと考えられるので，両方の要因から残差項が残ると考えられるが，p個の要因といっても，たとえ自由度があったとしても事実上数十の要因を投入することはできない。普通何々モデルと呼ばれるものでも一つの方程式の中に含まれてる変数の数は5，6個というものであろう。元来(2)式を母集団の関係として考えれば，われわれの手元にある方程式はその推定式である。従って母集団において10個の要因を明確に考えることができるとすれば，標本式についても元来は10個の要因を導入すべきであるが，これを5個にしてしまえば，はじめから想定誤差（specification error）を犯すということになる。しかし元来(2)式をuがあるという意味において近似式と考えれば，想定誤差があっても更に近似式という意味と考えれば許容されるかもしれない。しかし逆に標本式で5個の独立変数を導入したということは，母集団においてもその関係式であるとみなしているのだということになるかもしれない。いずれにせよ，理論的にというか，想定上例えば8個の独立変数が考えられるとすれば，

(3) $Y=a_0+a_1X_1+a_2X_2+\cdots\cdots+a_8X_8+u$

が考えられ，これに対して母集団の係数をαで示せば

(4) $Y=\alpha_0+\alpha_1X_1+\alpha_2X_2+\cdots\cdots+\alpha_8X_8+u$

となってa_iはα_iの推定量となる。これがきまると次に実際の計算ということになるが，ここでこの計算結果がうまくいった，成功したというのと，失敗したというのはどんな場合をさしていうのであろうか。一般的には次のように言えよう。係数のa_iについてはあらかじめ何等かの想定があることが多い。第一は符号条件である。＋（プラス）であるか－（マイナス）であるか，言葉を変えていえば，$r_{YX_i}\cdots\cdots$，YとX_iとの偏相関が，あらかじめ予定された符号通りになっているかどうかということである。第二は，a_iのいくつかについてその範囲についての条件がある場合がある。例えばa_iについては0から1の範囲にあるべきという条件があればその範囲にないと困るということである。この二つの条件がみたされなければ，その計算は失敗したといってもよいであろう。ただし符号条件についてはYとX_iとの単相関がプラスであっても，XとX_iとの偏相関は理論的にマイナスになると考えられる場合があることに注意しておかなければならない。この二つの条件がみたされた後で次に問題になるのはa_iの標準誤差である。もちろんこれは小さいほど望ましいわけであるが，一つは係数との相対的比較であって，係数の大きさ自体はそれほど問題でないから元来相対論であろう。その比較はt（t分布）の値として示され，それが有意であることが望ましい。有意水準は元来主観的なものであるが，普通5％水準をとることが多いから，5％水準で有意であることが望ましいことになる。もし有意でないと判定されれば，それは偏相関が有意でないのであるから$r_{XY_i}\cdots$が0であるという仮説はすてられないことになり，X_iを導入すべきであるという積極的主張はうすらぐといってよいであろう。したがってすべてのa_iについて，それが有意であることが（a_0を除いて，$a_i=0$という仮説に対して）望ましい。そうであるとき成功であると考えられ，有意でないものがふえるにつれて失敗度が大きくなると考えられる。しかし一つのa_iについて有意でない場合，全体の式を失敗と考えるかどうかは分析者の判断にあるといってもよいであろう。それでは理論的には確かに関係があると思われる，例えば8つの変数を導入したにもかかわらず（ということは母集団においてはこの関係があると考えられるにもかかわらず），上記のような失敗が生じたとすれば，これは標本において生じる問題かもしれない（しかし事実上標本しかないのであるから，これをさけて通るわけにはゆかないのであるが）。それが多重共線性（multicollinearity）通常マルチコと呼ばれる問題である。先にあげた成功の条件として更に重相関係数が有意であること（例えば5％水準で），があげられよう。しかし経済変量をあつかう場合には，この条件はほとんどすべてみたされているといってもよいであろう。まれにみたされていない場合もあるが，これがみたされなくては問題にならないといってもよいであろう。もう一度並記してしるすと，

1．重相関係数が有意であること（有意水準はすべて5％とする）。

2．係数の符号条件，範囲の条件をすべてみたしていること。

3．係数がすべて有意であること（仮説0に対して），

すなわち偏相関係数がすべて有意であること。

係数の分散（標準誤差）は小さい方が望ましい（この表現 仕方はあいまいであるが，感覚としてそういえるであろう）。

以上三つの条件がすべてみたされていれば，問題はないが，多くの場合は(1)と(2)の条件がみたされていればまず満足すべきで，(3)の条件についてはいくつかの偏相関係数が有意でなくてもまず仕方がないとするであろう。

いずれにせよ，この条件を満足せしめないようにする原因としてマルチコが考えられている（その他uの分布で自己相関がないこと，ダービンワトソン検定のことがあるが実際にはあまり重要視されない）。

（二）

多重共線性を論ずる場合，どのような状態になった場合についてそういえるのか，そしてそうであると判定されたときはどうすればよいか，これらの点を明確にしておかねばならないのであるが（それ等の点について竹内啓氏が論じたことがあるが[1]），実際上はなかなか明確にするのは困難なのである。一般に多重共線性で，完全な場合とそうでない場合とあり，完全な場合とは独立変数間に完全な一次関係がある場合をいう。三変数の場合についていえば

(5) $Y=a_0+a_1X_1+a_2X_2+u$

において，

(6) $X_1=a_0'+a_2'X_2$

という関係がある場合をいう。この場合は明確であって，偏回帰係数は計算不能になって二つの変数を分離することができない。係数の分散は無限大になってしまう。そのため実際上X_2を導入することはできないし，X_1あるいはX_2をどちらか一つ導入することになる。この場合は明確なのであるが，次にX_1，X_2間に完全な一次関係（完全相関）はないけれども，非常に高い相関がある場合も通常マルチコがあるという。これ以後は不明確となって非常に高いとはいったいどの程度をいうのか，ばくぜんとしている。例えば0.95以上であるとか（これだけでは標本の大きさが問題であるから，5％で有意というのも一つの規準となるかもしれない），いわれるが，明確性を欠く。そしてなぜ独立変数間に高い相関があるといけないかといえば，X_1の効果とX_2の効果を分離することが困難になる。その点からいえば，X_1，X_2が無相関であることが最も望ましいことになるが，$r_{X_1X_2}=0$であれば多元回帰の係数は単純回帰と同じものになって多元回帰をする意味はなくなるといってもよい。変数を標準形にすれば，正規方程式は$Y=1$，$X_1=2$，$X_2=3$とすれば，

(6) $$\begin{aligned} a_1r_{22}+a_2r_{23}&=r_{12}\\ a_1r_{32}+a_2r_{33}&=r_{13}\end{aligned}$$

と書けるから

(7) $$a_1=\frac{\begin{vmatrix} r_{12} & r_{23}\\ r_{13} & r_{33}\end{vmatrix}}{\begin{vmatrix} r_{22} & r_{23}\\ r_{32} & r_{33}\end{vmatrix}}=\frac{r_{12}-r_{13}r_{23}}{1-r_{23}{}^2}$$

$r_{23}=0$とすれば

(8) $a_1=r_{12}$

同様に

(9) $a_2=r_{13}$

となるから標準形の単純回帰$Y=r_{12}X_1$，$Y=r_{13}X_2$に等しいことになる。

従って$r_{23}\fallingdotseq 0$でなくても，むしろ$r_{23}\fallingdotseq 0$のときに多元回帰を用いる意味があるといえるが，あまり高くてはまずい（これについてはこれから述べるわけであるが），とすればどういうときに望ましくないかという規準の一つとしてL. R. Kleinが示したものがある[2]。それは

(10) $r^2{}_{X_iX_j}\geqq R^2{}_{Y,\ X_1,\ X_2,\ \cdots,\ X_a}$

$r^2{}_{X_iX_j}$は二つの説明（独立）変数X_iとX_jの単相関係数，R^2は重相関係数である。説明変数間の相関が重相関よりも高いのはよくないであろうということは感覚的にわかるが，理論的にはよくわからない点があろう。例はあとで示すが，実際の資料をあつかう場合はこのような場合は少ないであろう。しかし一つの規準として採用されるかもしれない。

次にP. E. FarrarとR. R. Glauber[3]によって示されたマルチコに対するχ^2（カイ）自乗検定がある。仮説としては標本における説明変数間には相関がない（$r_{X_iX_i}=1$ $r_{X_iX_j}=0$）とする。変数を変換して，

(11) $$\frac{X-\overline{X}}{\sqrt{n}\ S_{X_j}}$$

注(1) 竹内啓「多重共線性について」『理論経済学』1964, No. 2, p. 31.

(2) L. R. Klein, Introduction to Econometrics, Prentice-Hall International, London. p. 64, 101.

(3) D. E. Farrar and R. R. Glauber, Multicollinearty in Regression Analysis. Rev. Econ, and Statist vol. 49, 1967, p. 92-107.

S_{X_j}はXの標準偏差とすると，Xの変動，共変動行列は三変数の場合

(12) $$\begin{bmatrix} \sum X_1^2 & \sum X_1X_2 & \sum X_1X_3 \\ \sum X_1X_2 & \sum X_2^2 & \sum X_2X_3 \\ \sum X_1X_3 & \sum X_2X_3 & \sum X_3^2 \end{bmatrix}$$

(13) $$\begin{bmatrix} 1 & r_{X_1X_2} & r_{X_1X_3} \\ r_{X_1X_2} & 1 & r_{X_2X_3} \\ r_{X_1X_3} & r_{X_2X_3} & 1 \end{bmatrix}$$

で示される。説明変数間の相関がすべて0の場合，行列式を考えれば

(14) $$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1$$

となり，完全なマルチコの場合で$r_{X_iX_j}=1$なら

(15) $$\begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0$$

となる。実際はこの中間にあるわけであるが仮説の下で，Glanber と Farrar は次の関係を示した。

(16) $$*\chi^2 = -\left[n-1-\frac{1}{6}(2K+5)\right]\log e \begin{bmatrix}\text{標準化さ}\\\text{れた行列}\\\text{式の値}\end{bmatrix}$$

$*\chi^2$ は標本から計算された値，n＝標本の大きさ，K＝説明変数の数，仮説の下で $*\chi^2$ は自由度$r=\frac{1}{2}K(K-1)$ の χ^2（カイ）自乗分布をする。従ってこの値を計算して，有意水準を定めておけば，例えば5％として，その値を越えれば $r_{X_iX_j}=0$（直交性の仮定）は棄却され，マルチコの危険性があると判断される。有意でなければ直交性の仮定が棄却されないのでマルチコはそれほどないと判断される。

第三の規準として，説明変数が2個の場合はX_1とX_2の相関が高ければ問題とされるが，説明変数が3個以上の場合は個々の二変数間の相関が低くても，この3個の説明変数間の一次関係が高いことがありうる，Dhrymes(4)の著書の中の例では，独立変数間の相関行列として

(16) $$\begin{bmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & -0.68 \\ 0.4 & -0.68 & 1 \end{bmatrix}$$

となった場合，その行列式を計算すると，

(17) $$\begin{vmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & -0.68 \\ 0.4 & -0.68 & 1 \end{vmatrix} = 1-(0.68)^2 -0.8(0.672)=0.$$

二変数間にあまり高い相関はないにもかかわらず，相関行列は実際上0となり計算不能になるわけで，マルチコが存在するものと考えられる。これに対して相関行列が

(18) $$\begin{bmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{bmatrix}$$

と各変数間に高い相関がある場合を計算すると

(19) $$\begin{vmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{vmatrix} = (1-0.81)-0.9(0.9-081) +0.9(0.81-0.9)=0.028$$

となって，X_1, X_2, X_3の間には一次関係はなく相関行列式も0ではない。しかし0に近いとも言えるわけで，このへんは分析者の判断ということになろう。

そこで説明変数が3個以上の場合，この説明変数間の重相関係数を計算して，それが有意でないことが望ましいことになる。Glauber と Farrar によって示された式ではk個の説明変数に対して（$R^2{}_{X_1\cdot X_2X_3\cdots\cdots X_k}$，$R^2{}_{X_2\cdot X_1X_3\cdots\cdots X_k}$，一般に$R^2{}_{X_i\cdot X_1X_2\cdots\cdots X_k}$）を計算して重相関に対する検定

(20) $$F=\frac{(R^2{}_{X_i\cdot X_1X_2\cdots X_k})/(k-1)}{(1-R^2{}_{X_i\cdot X_1X_2\cdots X_k})/(n-k)}$$

n：標本の大きさ

k：説明変数の数

自由度 $r_1=k-1$ $r_2=(n-k)$ のF分布の表より例えば5％水準で検定する。有意ならマルチコが存在するとし，有意でなければそういえないと判断する。

第4の規準として，X_iX_j間の偏相関を計算してその有意性を検定する。

(21) $$t=\frac{(r_{X_iX_j\cdot X_1X_2\cdots X_k})\sqrt{n-k}}{\sqrt{1-r^2{}_{X_iX_j\cdot X_1X_2\cdots X_k}}}$$

自由度（$n-k$）のt分布の表から検定を行う。もし，有意であればマルチコの危険性があると判断し，有意でなければその危険性がないと判断する。

Glauber と Farrar の例では

(22) $$\log Y=b_0+b_1\log X_1+b_2\log X_2+\cdots +b_7\log X_7+\log u$$

船の推持費に関するもので$n=96$　$k=7$

変数全部では8で，結果は

$R^2=0.80$　$F=56$　自由度7と88

5％点の値に2.13ぐらいであるからもちろん有意である。説明変数間の重相関は，

(23) $R_{X_1\cdot X_2X_3\cdots\cdots X_7}=0.30$　　$F_{X_1}=6.3$

注(4) Phoebus J. Dhrymes Introductory Econometrics Springer-Verlag New York Heidelbery Berlin, p. 191.

$R^2 x_2 \cdot x_1 x_3 \cdots\cdots x_7 = 0.57$　　$F_{X_2} = 19.9$

$R_2 x_3 \cdot x_1 x_2 \cdots\cdots x_7 = 0.30$　　$F_{X_3} = 6.3$

$R^2 x_4 \cdot x_1 x_2 \cdots\cdots x_7 = 0.76$　　$F_{X_4} = 47.5$

$R^2 x_5 \cdot x_1 x_2 \cdots\cdots x_7 = 0.76$　　$F_{X_5} = 47.1$

$R^2 x_6 \cdot x_1 x_2 \cdots\cdots x_7 = 0.46$　　$F_{X_6} = 12.7$

$R^2 x_7 \cdot x_1 x_2 \cdots\cdots x_7 = 0.24$　　$F_{X_7} = 4.8$

自由度6,87のF分布の5％点は2.22ぐらいであるからすべて有意であるが，とりわけX_4とX_5が値が大きいのでX_4とX_5についてマルチコの影響があるのではないかとしている。この辺にも問題があろう。

係数とその標準誤差については

	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\hat{b}_6$	b_7
(24)	0.34	0.40	−0.79	0.05	−0.03	0.11	−0.16
$S(\hat{b}_i)$	(0.03)	(0.08)	(0.09)	(0.10)	(0.09)	(0.04)	(0.06)

自由度88のt分布の5％点は1.99ぐらいであるから$\hat{b}_4$と$\hat{b}_5$を除いていちぢるしく有意である。この点からみればまずまずの成功といってよいであろう。

次に偏相関の計算をみると，

表1

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	—						
X_2	$r_{12\cdot}=0.13$ (1.27)	—					
X_3	$r_{13\cdot}=0.21$ (2.01)	$r_{23\cdot}=0.27$ (2.60)	—				
X_4	$r_{14\cdot}=-0.35$ (−3.57)	$r_{24\cdot}=0.34$ (3.44)	$r_{34\cdot}=0.09$ (0.82)	—			
X_5	$r_{15\cdot}=-0.27$ (1.27)	$r_{25\cdot}=-0.21$ (−2.03)	$r_{25\cdot}=0.12$ (1.15)	$r_{45\cdot}=-0.68$ (−8.77)	—		
X_6	$r_{16\cdot}=0.45$ (4.72)	$r_{26\cdot}=0.08$ (0.78)	$r_{36\cdot}=-0.35$ (−3.51)	$r_{46\cdot}=0.31$ (3.13)	$r_{56\cdot}=0.50$ (5.51)	—	
X_7	$r_{17\cdot}=0.34$ (3·38)	$r_{27\cdot}=-0.13$ (−1.27)	$r_{37\cdot}=0.06$ (0.59)	$r_{47\cdot}=0.40$ (4.06)	$r_{57\cdot}=0.27$ (2.68)	$r_{67\cdot}=-0.26$ (−2.55)	—

（ ）内の値はtの値である。これをみると全部で21個の値があるが，そのうち14個が有意である。この辺をどうみるかということであろう。X_4とX_5についてはX_4については6個中5個が有意，X_5については6個中4個が有意でX_5については全体の中で特に多いというわけではない。X_4とX_5が偏相関の値として最も高く，有意度も高い。

以上の結果からみれば，X_4とX_5について最も問題があり，これについて再考されるべきであるとの考えもでてくるであろう。

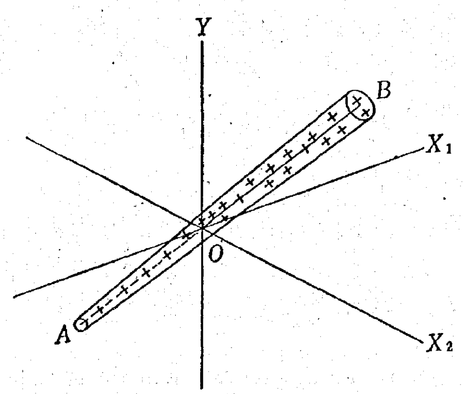
なおマルチコの第5の規準として，これは主として小標本の場合についていわれることであるが，標本のいくつか（数は少ない）を除いたり，つけ加えたりした場合係数が大きく変る場合は，要するに係数が不安定であるからマルチコが存在するという考え方もある。

以上5つの規準をすべてみたすことはほとんど考えられないから，むしろその程度と，程度をどう考えるかということになろう。

（三）

2つの説明変数間に高い相関があるときの図がWaltersの著書の中にある。(5)

図1



これをみれば平面をきめようとしてもきめにくくな

注(5)　A. A. ウォルターズ，神保一郎訳，「入門計量経済学，上」東洋経済新報社，p. 133.

ると説明されている。また X_1 と X_3 を切りはなして動かせることが困難になると説明されている。また r_{12}^2 が0.98から0.99に変化すると，偏回帰係数を示す式で他の事情にして等しい限り，その分母が0.02から0.01になり，偏回帰係数の値は2倍になってしまうから，r_{12}^2 の値が大きい場合は推定値はかなり不安定になると述べられているが，事実上分子も変化するので一般的にはいえないように思われる。つまり不安定というのは係数の分散が大きくなるということであるが，これについて述べてゆこう。二変数の場合

(25) $Y=a_0+a_1X+u$

とすれば a_1 の分数は

(26) $S_{a_1}^2=\frac{\sigma_u^2}{\sum(X-\bar{X})^2}$

で示される。σ_u^2 は母集団回帰における誤差項 u の分散である。三変数以上になると，

(27) $Y=a_0+a_1X_1+a_2X_2+\cdots\cdots+a_pX_p+u$

で

$$S_{a_1}^2=\frac{\sigma_u^2}{\sum(X_1-\hat{X}_1)^2}$$

X_1 は他の X についての X_1 の回帰，すなわち

(28) $X_1=b_0+b_2X_2+b_3X_3+\cdots\cdots+b_pX_p+v$

X_1 に対する他の X との重回帰における誤差の変動が分母になるということである[(6)]。全部が三変数で考えれば

(29) $\hat{X}_1=b_0+b_2X_2$

であるから，

(30) $\sum(X_1-\hat{X}_1)^2=\sum x_1^2(1-r_{12}^2)$　小文字は平均からの偏差

と書くことができ，四変数以上の場合は

(31) $\sum(X_1-\hat{X}_1)^2=\sum x_1^2(1-R_2)$　R^2は重相関説明変数のみ

と書くことができる。従って a_1 の分散は三変数のときは，$X_1=2$ $X_2=3$ とおけば

(32) $S^2_{a_1}=\frac{\sigma_u^2}{\sum x_1^2(1-r_{23}^2)}$

(33) $S_{a_1}^2=\frac{\sigma_u^2}{\sum x_1^2(1-R^2)}$

となり，(32)よりみれば $r_{23}^2=0$ X_1 と X_2 の相関が0のとき最も小さく，r_{23}^2 大きくなるほど分散は大きくなり $r_{23}^2=1$ のとき無限大となる。三変数以上の場合は X_1 を X_2 以下で説明できればできるほど，説明変数間の重相関が高ければ高いほど a_1 の分散は大きくなる。σ_u^2 は母集団の分散であるから常数と考えられるから分母が小さくなればなるほど全体は大きくなる。この様子は[(7)]，

表　2

r_{23} の値	a_1 の分数
0	$\frac{\sigma^2}{\sum x_1^2}$
0.50	$(1.33)\frac{\sigma^2}{\sum x_1^2}$
0.70	$(1.96)\frac{\sigma^2}{\sum x_1^2}$
0.80	$(2.78)\frac{\sigma^2}{\sum x_1^2}$
0.90	$(5.26)\frac{\sigma^2}{\sum x_1^2}$
0.95	$(10.26)\frac{\sigma_2}{\sum x_1^2}$
1.00	∞

r_{23} が増大すると a_1 の分散は増大し，$r=0.8$ のときは無相関のときの2.8倍になり，$r=0.95$ となれば10倍以上となる。三変数以上ときは X_1 に対する重相関が高ければ高いほど a_1 の分散は大きくなるが，そのなり方は r_{23}^2 を R^2 でおきかえたものに等しいことになる。

分散を小さくする方法として，r，R が一定なら $\sum x_1^2$ は標本の大きさが大きくなるにつれて増大する傾向があるといえるから，標本の大きさを増すことが考えられる。それについて，Michael S. Commonのモンテカルロ実験の結果を示す[(8)]と，彼の記号に従えば

(34) $Y_t=\beta_0+\beta_1X_{1t}+\beta_2X_{2t}+\varepsilon_t:\varepsilon_t\sim N\ (0,\ \sigma^2)$

$X_{1t}=K_0+K_2X_{2t}+u_t$

このモデルで

(35) $Y_t=1.0+0.5X_{1t}+2.0X_{2t}+\varepsilon_t:\varepsilon_t\sim N(0,\ 1)$

$X_{2t}=20.0+4.0X_{1t}+u_t:u_t\sim N\ (0,\ 0.25)$

とする。このようにすれば X_1 と X_2 とはかなり高い相関が得られるであろう。その結果は表2となった。T＝標本の大きさ，回数は100回（の平均），一般的に言えば，標本の大きさが大きくなるほどよくなっている

注(6) Harry H. Kelejian, Wallace E. Oates, Introduction to Econometrics, Principle and Appslications. Harper International Edition, p. 186.

(7) Damodar Gujarati Basic Econometrics McGraw-Hill Book Company p. 176.

(8) Michael S. Common, Basic Econometrics, an introductory text for econometrics, Longman p. 363-366.

表 2

	1	2	3	4	5	6	7	8	9	10	11
	$E[\hat{\beta}_0]$	$VAR[\hat{\beta}_0]$	$E[\hat{\beta}_1]$	$VAR[\hat{\beta}_1]$	$E[\hat{\beta}_2]$	$VAR[\hat{\beta}_2]$	Reject $H_0:\beta_0=0$	Reject $H_0:\beta_1=0$	Reject $H_0:\beta_2=0$ *and*	Reject $H_0:\beta_1=0$ $H^0:\beta_2=0$	$\bar{R}^2$
$T=10$	0.6767	288.320	0.3875	11.3509	2.0228	0.7145	4	5	59	2	0.9987
20	1.8934	117.147	0.6854	4.6599	1.9551	0.2920	5	7	89	4	0.9986
30	−0.2063	60.6675	0.2589	2.3908	2.0602	0.1500	8	7	100	7	0.9985
40	0.9137	28.6503	0.4940	1.1396	2.0033	0.0711	5	4	100	4	0.9985
50	0.9772	33.3821	0.4864	1.3358	2.0024	0.0832	5	6	100	6	0.9984
60	0.6599	20.3115	0.4272	0.7962	2.0179	0.0500	5	5	100	5	0.9984
70	0.9601	25.0701	0.4980	1.0058	2.0014	0.0623	5	6	100	6	0.9983
80	1.2683	20.7338	0.5487	0.8344	1.9878	0.0518	6	11	100	11	0.9984
90	0.8863	15.9197	0.4717	0.6170	2.0067	0.0390	2	5	100	5	0.9984
100	0.9793	16.1222	0.4964	0.6574	2.0012	0.0404	7	10	100	10	0.9984

といえるであろう。これを，

(36) $X_{2t}=20.0+4.0X_{1t}+u_t$: $u_t \sim N\ (0,25.0)$

の実験，X_1 と X_2 の相関の低い場合と比較してみると，その結果は，表3のごとくなり，これをみればこの方が良い結果を示しているのは明瞭である。

しかし上記の場合は母分散既知の場合といえるが，実際上は母分数は未知であるから，これを $\hat{\sigma}_u{}^2$ で置きかえなければならない。従って，前の記号に戻れば，

三変数なら

(36) $S_{a_1}{}^2=\dfrac{\hat{\sigma}_u{}^2}{\sum x_1{}^2(1-r_{23}{}^2)}$

となり，

(37) $\hat{\sigma}_u{}^2=\dfrac{\sum y^2(1-R^2)}{n-2}$　　R^2 全体の重相関

となり，変数の標準形を考えれば

表 3

	1	2	3	4	5	6	7	8	9	10	11
	$E[\hat{\beta}_0]$	$VAR[\hat{\beta}_0]$	$E[\hat{\beta}_1]$	$VAR[\hat{\beta}_1]$	$E[\hat{\beta}_2]$	$VAR[\hat{\beta}_2]$	Reject $H_0:\beta_0=0$	Reject $H_0:\beta_1=0$	Reject $H_0:\beta_2=0$ *and*	Reject $H_0:\beta_1=0$ $H_0:\beta_2=0$	$\bar{R}^2$
$T=10$	1.0867	3.5097	0.4695	0.1154	2.0023	0.0071	11	24	100	24	0.9986
20	1.0856	1.3650	0.5238	0.0504	1.9955	0.0029	21	67	100	67	0.9988
30	0.8876	0.7787	0.4757	0.0269	2.0060	0.0015	20	75	100	75	0.9987
40	0.9733	0.3880	0.5059	0.0140	2.0003	0.0007	28	93	100	93	0.9987
50	1.0204	0.4224	0.4950	0.0156	2.0002	0.0008	33	97	100	97	0.9986
60	0.9826	0.2802	0.4917	0.0091	2.0018	0.0005	44	100	100	100	0.9986
70	0.9848	0.3247	0.5029	0.0128	2.0001	0.0006	47	100	100	100	0.9986
80	1.0486	0.2631	0.5047	0.0104	1.9988	0.0005	54	100	100	100	0.9986
90	1.0060	0.2280	0.4956	0.0069	2.0007	0.0004	56	100	100	100	0.9986
100	1.0010	0.1937	0.5008	0.0087	2.0001	0.0004	61	100	100	100	0.9986

(38) $S_{a_1}{}^2=\frac{(1-R^2)}{(n-2)(1-r_{23}{}^2)}$

となる。$r_{23}{}^2$が動けばR^2も動くことになり，$r_{23}{}^2$が大きくなれば一方的に$S_{a_1}{}^2$が大きくなるとはいえないし，回帰係数との比較tの値も一方的に減少してゆくとも言えないことになる。

三変数の場合についてみると，再び $Y=1$ $X_1=2$ $X_2=3$と記号すると，

(39) $R^2=\frac{r_{12}{}^2+r_{13}{}^2-2r_{12}r_{13}r_{23}}{1-r_{27}{}^2}$

である。今$r_{12}=r_{13}$の場合について考えると，まとめてr_{12}と書くと，

(40) $R^2=\frac{r_{12}{}^2+r_{12}{}^2-2r_{12}r_{12}r_{23}}{1-r_{23}{}^2}=\frac{2r_{12}{}^2-2r_{12}{}^2r_{23}}{1-r_{23}{}^2}$

$$=\frac{2r_{12}{}^2(1-r_{23})}{1-r_{23}{}^2}=\frac{2r_{12}{}^2(1-r_{23})}{(1+r_{23})(1-r_{23})}=\frac{2r_{12}{}^2}{1+r_{23}}$$

(41) $S_{a_1}{}^2=\frac{1-R^2}{(n-2)(1-r_{23}{}^2)}=\frac{1-\frac{2r_{12}{}^2}{1+r_{23}}}{(n-2)(1-r_{23}{}^2)}$

$$=\frac{\frac{1+r_{23}-2r_{12}{}^2}{1+r_{23}}}{(n+2)(1-r_{23}{}^2)}=\frac{1+r_{23}-2r_{12}{}^2}{(n-2)(1+r_{23})(1-r_{23}{}^2)}$$

(40)よりr_{23}が大となるにつれてR^2は一方的に減少してゆくことがわかる。$\{r_{23}\geq 0\}$ $r_{23}=0$のとき，

(41) $R^2=2r_{12}{}^2$ （r_{12}がとれる範囲は限られている）

$R^2=1$のときは

(42) $\frac{2r_{12}{}^2}{1+r_{23}}=1$　$2r_{12}{}^2=1+r_{23}$　$r_{23}=2r_{12}{}^2-1$

例えば$r_{12}=r_{13}=0.7$とすれば，$2r_{12}{}^2=0.98$で$R=1$となり，そのときの$r_{23}=0.98-1=-0.02$となり，この値より+方向へr_{23}が増大するにつれてRは減少し，係数の分散は一方的に増大してゆくことがわかる。$r_{23}=2r_{12}{}^2-1$を分散の式に導入すれば

(43) $S_{a_1}{}^2=\frac{1+(2r_{12}{}^2-1)-2r_{12}{}^2}{(n-2)(1+2r_{12}{}^2-1)(1-[2r_{12}{}^2-1)^2])}=0$

重相関が1なのであるから，当然係数の分散は0となるが，実際はそのようなことはほとんど起り得ないであろう。

(44) $a_1=\frac{r_{12}-r_{13}r_{23}}{1-r_{23}{}^2}$　$r_{12}=r_{13}$なら

(45) $a_1=\frac{r_{12}-r_{12}r_{23}}{1-r_{23}{}^2}=\frac{r_{12}(1-r_{23})}{1-r_{23}{}^2}$

$$=\frac{r_{12}(1-r_{23})}{(1+r_{23})(1-r_{23})}=\frac{r_{12}}{1+r_{23}}$$

(46) $t=\frac{a_1}{S_{a_1}}=\frac{r_{12}}{1+r_{23}}\times\sqrt{\frac{(n-2)(1+r_{23})(1-r_{23}{}^2)}{(1+r_{23}-2r_{12}{}^2)}}$

$$=\frac{\sqrt{(n-2)}\,r_{12}\sqrt{(1-r_{23})}}{\sqrt{(1+r_{23}-2r_{12}{}^2)}}$$

となり，r_{23}の最小値（マイナスを含む）からr_{23}が大となるにつれてtは一方的に減少してゆくことがわかる。

これより三変数において$r_{12}=r_{13}$，各説明変数の従属変数に対する相関が等しい，単独での影響度が等しい場合には，説明変数間の相関が大きくなればなるほど，分散は大きくなり，tの値は小さくなり（=偏相関の値が小さくなり），望ましくない結果を与える。したがってこのような場合は説明変数間の相関が大きいほどマルチコが存在するといえることになる。この場合当然ながら，二つの偏回帰係数の値は等しく，回帰係数の値はr_{23}が増大すれば一方的に減少し（説明力がおちる），標準誤差は一方的に増大するからtの値は一方的に減少することになる。

しかしこの時注意すべきことは，$r_{12}\neq r_{13}$のときそのように一方的にならないことである。R^2については，これをYとおき，r_{23}をXと書けば，

(46) $Y=\frac{r_{12}{}^2+r_{13}{}^2-2r_{12}r_{13}X}{1-X^2}$

(47) $\frac{dY}{dX}=\frac{-2r_{12}r_{13}(1-X^2)-(r_{12}{}^2+r_{13}{}^2-2r_{12}r_{13}X)(-2X)}{(1-X^2)^2}=0$

$$X=\frac{r_{12}{}^2+r_{13}{}^2\pm\sqrt{(r_{12}{}^2+r_{13}{}^2)^2-4(r_{13}r_{12})^2}}{2r_{12}r_{13}}$$

$$=\frac{r_{12}}{r_{13}}\text{ または }\frac{r_{13}}{r_{12}}$$

$R^2=1$のときは

(48) $1=\frac{r_{12}{}^2+r_{12}{}^2-2r_{12}r_{13}X}{1-X^2}$

(49) $X=\frac{2r_{12}r_{13}\pm\sqrt{(2r_{12}r_{13})^2-4(r_{12}{}^2+r_{13}-1)}}{2}$

で，これよりr_{23}のとりうる範囲がきめられる。偏回帰係数の式より，$r_{12}<r_{13}$とすれば（$r_{12}>0$，$r_{13}>0$）$r_{23}=r_{12}/r_{13}$を代入すれば，この値よりr_{23}の値が大となればa_1の符号はプラスよりマイナスに変ることがわかる。そしてこの値よりR^2はそれまでの減少から増加に変るわけである。a_1についてr_{12}/r_{13}の値で0まで減少するが，それより増加に転ずる。tについても同様である。係数の標準誤差については一義的に言えない。問題は符号の変化ということであって，理論的にプラスであるものがマイナスに転ずるのであれば好ましくないことであり，r_{23}の増加は望ましいことではないが，マイナスであることが望ましいのであれば，$r_{23}=r_{12}/r_{13}$を越えたところであることが望ま

しいわけで，r_{23}が一方的に増加することが望ましくないとはいえないことになる。

4変数以上になると更に複雑となって一義的にいえないことになるが，Foxのあげた例よりみれば，$r_{12}=r_{13}=r_{14}$，$r_{23}=r_{23}$として，r_{34}の変化をみてゆく場合において，a_2，a_3の値は等しくなるが，r_{34}の増加

表 4 $r_{12}=0.9$，$r_{13}=0.7$，$N=20$の場合

r_{23}	$R^2_{1.23}$	$\beta_{12.3}$	$\beta_{13.2}$	$r_{12.3}$	$r_{13.2}$	S_β[a]	tの値 $\beta_{12.3}$	tの値 $\beta_{13.2}$
0.3187[b]	1.0000	0.7535	0.4599	1.0000	1.0000	0	—	—
0.4000	0.9488	0.7381	0.4048	0.9473	0.8511	0.0599	12.3222	6.7579
0.4500	0.9191	0.7335	0.3695	0.9172	0.7578	0.0772	9.5013	4.7863
0.5000	0.8933	0.7333	0.3333	0.8892	0.6623	0.0917	7.9967	3.6347
0.5500	0.8703	0.7384	0.2939	0.8635	0.5632	0.1046	7.0593	2.8098
0.6000	0.8500	0.7500	0.2500	0.8402	0.4588	0.1175	6.3830	2.1277
0.6500	0.8329	0.7706	0.1991	0.8200	0.3472	0.1305	5.9050	1.5257
0.7000	0.8196	0.8039	0.1373	0.8039	0.2249	0.1442	5.5749	0.9521
0.7500	0.8114	0.8571	0.0571	0.7938	0.0867	0.1592	5.3838	0.3587
0.8000	0.8111	0.9444	−0.0556	0.7935	−0.0765	0.1758	5.3720	−0.3163
0.8500	0.8252	1.0991	−0.2342	0.8107	−0.2831	0.1925	5.7096	−1.2166
0.9000	0.8737	1.4211	−0.5789	0.8673	−0.5789	0.1977	7.1882	−2.9282
0.9413[c]	1.0000	2.1149	−1.2912	1.0000	−1.0000	0	—	—

(a) 各βに対して同一。
(b) r_{23}の可能なもっとも低い値。
(c) r_{23}の可能なもっとも高い値。

$r_{12}=r_{13}=r_{14}=0.7$，$r_{23}=r_{24}=0.5$，$N=20$の場合

r_{34}	$\beta_{12.34}$	$\beta_{13.24}$	$\beta_{14.23}$	S_β[a]	tの値 $\beta_{12.34}$	tの値 $\beta_{13.24}$	tの値 $\beta_{14.23}$
−0.0196[b]	−0.0286	0.7286	0.7286	0.3572	−0.0800	2.0396	2.0396
0	0	0.7000	0.7000	0.3500	0	2.0000	2.0000
0.1000	0.1167	0.5833	0.5833	0.3224	0.3618	1.8091	1.8091
0.2000	0.2000	0.5000	0.5000	0.3062	0.6532	1.6330	1.6330
0.2500	0.2333	0.4667	0.4667	0.3012	0.7746	1.5492	1.5492
0.3000	0.2625	0.4375	0.4375	0.2981	0.8806	1.4676	1.4676
0.4000	0.3111	0.3889	0.3889	0.2970	1.0474	1.3093	1.3093
0.5000	0.3500	0.3500	0.3500	0.3031	1.1547	1.1547	1.1547
0.6000	0.3818	0.3182	0.3182	0.3182	1.2000	1.0000	1.0000
0.7000	0.4083	0.2917	0.2917	0.3472	1.1762	0.8402	0.8402
0.8000	0.4308	0.2692	0.2692	0.4038	1.0667	0.6667	0.6667
0.9000	0.4500	0.2500	0.2500	0.5449	0.8259	0.4588	0.4588
0.9500	0.4586	0.2414	0.2414	0.7538	0.6084	0.3202	0.3202
0.9800	0.4635	0.2365	0.2365	1.1763	0.3940	0.2010	0.2010
1.0000[c]	—	—	—	—	—	—	—

(a) 各βに対して同一。
(b) r_{34}の可能なもっとも低い値。
(c) r_{34}の可能なもっとも高い値。

と共に $a_2 a_3$ の値は一方的に減少し，標準誤差は一方的に増大し，t の値は一方的に減少してゆく。その意味において

(49) $Y=a_0+a_1X_1+a_2X_2+a_3X_3$

において X_2 と X_3 の相関が増加するほど，a_2，a_3 を推定する上で望ましくないことになるが a_1 については一義的にいえないことになる。ただしこれは上記の条件付のことであるから実際上はほとんどそのような条件にめぐりあうことはないであろう。Foxの例をあげておく（表4）。

Foxの記号のままあげたが，$\beta_{12,3}=a_1$，$\beta_{12,2}=a_2r_{12,3}$，$r_{12,1}$ はそれぞれ偏相関，$S_\beta{}^2=S_a{}^2$ と考えればよい。

4変数の場合も同様である。Foxはこれ以外にも同様の計算を行っている。

(四)

前に筆者はいろいろな計算比較の例を試みたが，マルチコの例にあてはめてみよう。

資料は，表5のようであり，相関行列は表6（標本の大きさ，10の場合と12の場合）で与えられる。Y_1 をすべての X で考えるという場合をとりあげてみる。

(50) $Y_1=a_0+a_1X_1+a_2X_2+aX_3+u$

これまでに述べた計算結果は「計量分析の方法」の

表　5

	Y_1	Y_2	X_1	X_2	X_3
昭和27年	7096	8330	1163.5	100.0	9241
28	8741	8250	1505.1	95.3	10575
29	10640	8000	1507.6	90.8	12834
30	10519	7697	1469.0	88.1	16230
31	12969	6497	2044.9	90.8	17514
32	15107	7000	2684.3	97.6	18696
33	14904	6700	2665.3	95.7	21408
34	17169	6300	3263.2	91.4	23148
35	22425	6700	4339.9	88.4	29020
36	24484	6000	5985.2	86.6	31200
37	28662	6350	6534.6	84.9	35568
38	29766	6275	7083.5	82.2	43200

Y_1＝セメントの生産量，Y_2＝セメントと価格，X_1＝総投資量，
X_2＝石炭価格指数，X_3＝セメントの容量。

標本10のときの相関行列

	X_1	X_2	X_3	Y_1	X^1
Y_1	.10000000 E +01	−.54398195 E +00	.94679313 E +00	.97339518 E +00	.80120823 E +00
X_2	−.54398195 E +00	.10000000 E +01	−.62157664 E +00	−.59230707 E +00	.48521478 E +00
X_3	.94679313 E +00	−.62157664 E +00	.10000000 E +01	.98568452 E +00	−.88180014 E +00
Y_1	.97339518 E +00	−.59230707 E +00	.98568452 E +00	.10000000 E +01	−.85186380 E +00
Y_2	.80120823 E +00	.48521478 E +00	−.88180014 E +00	−.85186380 E +00	.10000000 E +01

標本12のときの相関行列

	X_1	X_2	X_3	Y_1	Y_2
X_1	1.00000000	−0.76442904*	0.97297765*	0.98798193*	−0.77494435*
X_2	−0.76442904*	1.00000000	−0.80528125*	−0.78426861*	0.59097228**
X_3	0.97297765*	−0.80528125*	1.00000000	0.98561912*	−0.81261437*
Y_1	0.98798193*	−0.78426861*	0.98561912*	1.00000000	−0.81273610*
Y_2	−0.77494435*	0.59097228**	−0.81261437*	−0.81273610*	1.00000000

（注）＊印は1％で有意，＊＊印は5％で有意。

中に述べられているが，標本10の場合は

(51) $Y_1=794+\underset{(0.54)}{1.437X_1}+\underset{(77)}{6.69X_2}+\underset{(0.121)}{0.4828X_3}$

となり，標本12の場合は，

(52) $Y_1=3114+\underset{(0.64)}{1.980X_1}-\underset{(98)}{3.998X_2}+\underset{(0.14)}{0.3342X_3}$

（ ）内は標準誤差である。これよりみると符号その他かなりの変化がみられ，マルチコでないという一つの条件はみたしていないと思われる。あと(51)式について独立（説明）変数間の重相関，偏相関，直交性の検定の問題を考えてみよう。相関行列式の値は，

(51) $$\begin{vmatrix} 1.00000000 & -0.54398195 & 0.94679313 \\ -0.54398195 & 1.00000000 & -0.62157664 \\ 0.94679313 & -0.62157664 & 1.00000000 \end{vmatrix}$$

$$=0.06158053$$

(52) $\chi^2=-[10-1-\frac{1}{6}(2\times3+5)]\log e\,0.06158053$
$=19.976491$

自由度$\frac{1}{2}3(3-1)=\frac{6}{2}=3$

自由度3の χ^2（カイ）自乗分布の5％点は7.81であるから有意である。直交性の仮定はすてられることになる。次に重相関をみてみる。

X_1を従属とするときを$R^2{}_{X_1}$以下$R^2{}_{X_2}$，$R^2{}_{X_3}$を計算する。

(53) $$R^2{}_{X_1}=\frac{(-0.54398195)^2+(0.94679313)^2-2((-0.54398195)\times0.94679313\times(-0.62157664))}{1-(-0.62157664)^2}$$

$=0.89964754 \quad R=0.94849752$

$$R^2{}_{X_2}=\frac{(-0.54398195)^2+(-0.62157664)^2-2((-0.54398195\times0.94679313\times(-0.62157664))}{1-(0.94679313)^2}$$

$=0.40549447 \quad R=0.63678848$

$$R^2{}_{X_3}=\frac{(0.94679313)^2+(-0.62157664)^2-2[1-0.54398195\times0.94679313\times(-0.62157664)]}{1-(-0.54398195)^2}$$

$=0.91253805 \quad R=0.95526857$

重相関の検定はF分布を使うが，検定表よりみれば，三変数で自由度10−3=7のときの5％点は0.758であるから，この点からすればX_2みが有意でないことになる。

次に偏相関をみる。三変数の式で$r_{12.3}$は

(54) $r_{12.3}=\dfrac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}{}^2}\sqrt{1-r^2{}_{23}}}$

$X_1=1 \quad X_2=2 \quad X_3=3$として，

(55) $$r_{12.3}=\frac{(-0.54398195)-(0.94679313)(-0.62157664)}{\sqrt{1-(0.94679313)^2}\sqrt{1-(-0.62157664)^2}}=0.17659503$$

$$r_{13.2}=\frac{(0.94679313)-(-0.54398195)(-0.62157664)}{\sqrt{1-(-0.54398195)^2}\sqrt{1-(-0.62157664)}}=0.92599721$$

$$r_{23.3}=\frac{(-0.62157664)-(-0.54398195)(0.94679313)}{\sqrt{1-(-0.54398915)^2}\sqrt{1-(0.94679313)^2}}=-0.39450249$$

自由度7の二変数の5％点は0.666であるから，$r_{X_1X_3}$が有意であるが他は有意でない。この結果からみるかぎりX_2という変数は他との相関が低く，その意味でマルチコのない変数となっているが，これは同時に従属変数に対しても低い相関をもっているので，最初に示した重回帰の式においてはいわば最も悪い変数となってしまっているのである。ここにも一つの問題があるといってもよいであろう。Farrar と Glauber の式ではすべての単相関が示されていないのでこれが示されていたらまた別の問題が生じていたかもしれない。

ちなみに大きさ12標本の場合を計算してみると，$\chi^2_3=36.633573$と更に大きな値となっており，$R_{X_1}=0.97127400$，$R_{X_2}=0.79160490$，$R_{X_3}=0.97177425$，となって自由9の5％点は0.697となるのでいずれも有意となる。偏相関は，$r_{12.3}=0.13945812$，$r_{13.2}=0.9328587$，$r_{23.1}=-0.41244447$，自由度9の5％は0.602で，$r_{12.3}$と$r_{23.1}$は有意でない。この点は前の結果と変りないが，χ^2と重相関の点でマルチコの度合は更に増大していると考えられる。二つの相関行列をみると，r_{13}の値はほぼ等しいが，r_{12}，r_{23}の値が標本12の場合の方が，いずれも大きくなっており，この点が影響したものと思われる。最初に示した全変数の重回帰からもその点がうかがわれるが，標本の大きさを増せばマルチコの危険度がうすらぐといってもこの程度のことではほぼ同じ，更に悪くなる場合もあるのは当然のことであろう。もちろん標本を増せばよいというのは全体の標本抽出のモデルが変らないことを前

提としている。

いずれにしても，マルチコでないための条件をすべてみたすということは実際上不可能に近いことであろう。「計量分析の方法」の中で示された所得と消費の相関は0.98～0.99台にあり，消費函数を考える場合所得以外にもいろいろな要因（変数）が考えられるが，統計的にはすでに一変数だけでこのように高い相関があればそれで充分なのであり，他の変数が導入される余地は極めて少ないといってよいであろう。ここにむしろなやみがあるといってもよいであろう。

変数除去の問題はここであつかわない。(9)

注(9) A. Koutsoyiannis, Theory of Econometrics An Introductory of Econometrics Methods, Second Edition The Macmillan Pres LTD. p. 253.