

“見せかけの相関”を克服するために：
2次偏相関係数のすすめ

日下部 真 一

広島大学大学院総合科学研究科

**Spurious Correlations Revisited:
A second-order partial correlation is useful
for the test of causal relation among social variables**

Shinichi KUSAKABE

Graduate School of Integrated Arts and Sciences, Hiroshima University

Abstract

Spurious correlations are often induced both in regression analyses using ratio data and in logarithms of these ratio data (Kusakabe 2002a, b, 2011). Path analysis in the present study proves the process for the induction of spurious correlations in regression analysis. Population size always affects social variables, and therefore a correlation in a regression always consists of both direct effects and indirect effects via the population size between the relevant variables. In this sense, neither a per capita value nor a logarithm of per capita can be used in OLS regressions. From these analyses, the following results were elucidated. ① In regression analyses using ratio data or logarithms of these, incorrect results came about in 20% of cases (Table 3). These results show that OLS regressions using ratio data artificially induce the incorrect spurious correlation due to indirect effects via the population size. ② Second-order partial correlation is equivalent to the direct effect of the 4-variable recursive model in the structural equation modeling and is highly useful for the test of causality of the variables in question.

As an example, the current research critically analyzes the paper by Per Pettersson-Lidbom (2012 in *Journal of Public Economics*) and the negative correlation found in that paper is shown to be induced by the spurious negative effect due to population size.

Finally, I present the Law of Power Function in a broad sense for preventing the fallacious causal relationships between the social variables in the cross-section analyses.

Keywords: Spurious correlation, Path analysis, Structural Equation Modeling, Second-order partial correlation, Causality

After decades of neglect of causal analysis in economics, a surge of interest seems to be in progress.

Judea Pearl 2009

1. はじめに

計量経済学を中心として、広く社会科学の統計分析にあたっては、一般最小2乗法 (ordinary least square method, OLS) による回帰分析が行われ、通常“人口あたり (以下、per capitaであらわす)”や“GDPあたり”などの数値やこの対数変換値が用いられて解析される。これは、ひとえに“人口効果 (またはGDP効果) を除くために”という、おそらくは“客観的な”、“科学的な”論考によったわけで、これを疑う人は皆無のようである。しかし、生物界や人間世界のさまざまな諸活動は相加的世界ではなく相乗的世界であるがゆえに人間生活にかかわるほとんどの社会現象は人口に対し相乗的な“ベキ関数的関係”で表される (日下部2011)。このような社会指標を“per capita”の数値で表現するとさまざまな偏考がもたらされ、回帰分析結果に“見せかけの相関”が多く生じることを警告した (日下部2011)。本論で取り扱ういわゆる“見せかけの相関”については“相関は因果関係をしめすものにあらず”という警句で常に片付けられ見過ごされて来た課題であるが、筆者の主張する社会指標がもつベキ関数の性質によって生じる“見せかけの相関”については、統計学始まって以来、また、計量経済学始まって以来、いまだ論じられてこなかった重要な課題なのである。

通常、計量経済学や社会科学一般では、実測値のper capitaやこの対数変換値を用いて、回帰分析が行われる。本論文では、これらのper capitaやこの対数変換値が人口効果を介入させ、見せかけの相関を引き起こし誤った分析結果を与えることを示す。そして、これを克服するには人口効果とGDPの効果を同時にのぞいた2次偏相関を見るのが良いことを示す。分析にあたっては、定量的解析を正確にするために日本の都道府県やアメリカ50州の社会指標を用いたクロス・セクション分析を行っているが、問題とする人口効果は人間の諸活動が関与するところには常に潜んでくる。

例えば、都市問題や過疎問題は今まで社会学の大きな問題であったがこのような問題の中においても定量的分析を行う時には大なり小なりper capitaで比較するというような状況が出現する。こういう意味で社会科学全般を超えて日常生活にまでおよぶ重要な課題なのである。

さらに、計量経済学における実例として *Journal of Public Economics* (Pettersson-Lidbom 2012, 269-278) にのっている“議員定数を増やせば行政支出が少なくすむ” (negative council size effect) と主張する論文が、実は、見せかけの相関に起因する可能性が強いことを、日本の都道府県議会議員数とアメリカの州議会議員数の分析を通して論じる。

2. 分析方法と資料

用いた資料は『現代アメリカ統計年鑑』と *NonProfit Almanach* からとっている。On lineでとったデータもある。国内データに付いては『日本社会統計年鑑』による。分析手法は一般的な単回帰・重回帰分析で、4変数以上にかかわる構造方程式分析にあたってはソフトパッケージ Amos (Arbuckle & Wothke 1999) を用いた。本論で展開したように2次偏相関 (Yule 1922, Yule and Kendall 1973; McNemar 1962, Ley 1972, Kerlinger and Pedhazur 1973) を実用に用いた例はほとんどなく、また現在、ほとんど解説した統計書がないので1次の偏相関を合わせ、理解を深めるために付録として解説しておく。

per capita (人口あたり) の相関分析では、実測値を人口で割った値どうしとこれらの対数変換値どうしの分析を行っている。他の全ての分析では実測値の対数変換値をそのままパス解析や2次偏相関分析に用いている。統計的有意性はすべてt値で示しているが、5%有意水準を*で示し、有意でない時はnsで表している。

3. ベキ関数モデルが予測する“見せかけの相関”の検証

per capitaやこの対数変換値で表したクロスセ

表1 アメリカの社会指標を人口に対して1次回帰した傾き(β)と決定係数(R^2), 相関係数(R)の値。
(Table 1. Regression analyses for the various log-transformed social indicators on population size. β is the slope of the regression line, and R , R^2 are the correlation coefficient and the coefficient of determinant, respectively.)

Table 1(a) Regression on population size

Dependent variables	β	r^2	r
NPO: Nonprofit organizations	0.841	0.908	0.953
Suicide	0.879	0.878	0.937
VOTE: Turnout in the house of representative election	0.895	0.960	0.980
Welfare: State expenditure for welfare	0.961	0.933	0.966
Education: State expenditure for education	0.976	0.978	0.989
White: Nonhispanic white residents	0.980	0.966	0.983
GDP	1.019	0.974	0.987
Income: Personal income	1.024	0.982	0.991
Unemployments	1.059	0.972	0.986
Public support: Public support for NPOs	1.094	0.949	0.974
Murder	1.292	0.839	0.916

Table 1(b) Regression on GDP

	β	r^2	r	$\sqrt{1-r^2}$
NPO	0.823	0.928	0.963	0.268
Suicide	0.824	0.867	0.931	0.365
VOTE	0.863	0.950	0.975	0.224
Welfare	0.924	0.919	0.959	0.285
Education	0.944	0.977	0.988	0.152
White	0.936	0.939	0.969	0.247
Income	0.997	0.994	0.997	0.077
Unemployments	1.009	0.939	0.969	0.247
Public support	1.069	0.967	0.983	0.182
Murder	1.256	0.845	0.919	0.394

クシオンデータ間の回帰分析で容易に見せかけの相関が生じることを簡単なベキ関数のモデルを用いて説明しよう。

社会現象を表す多くの統計指標は、人口のベキ関数で表現される(表1)。今、二つの統計指標値(y_1 と y_2)を人口 x のベキ関数として表現すると、

$$y_1 = a_1 * x^{\beta_1},$$

$$y_2 = a_2 * x^{\beta_2},$$

と表現される。 a と β は両側対数変換したときの切片(α)と傾き(β)に相当する係数である。したがって、per capitaの値は、

$$y_1/x = a_1 * x^{(\beta_1-1)}$$

となり、これを Y とし、 y_2/x を X とおけば、

$Y = c * X^\gamma$, $\gamma = (\beta_1-1)/(\beta_2-1)$, ($c = a_1/(a_2^\gamma)$ は定数)となる。

同様に、per capitaの対数值間の関係は、

$$Y = \gamma * X + \log c$$

となる。

これらの関係は因果関係の有無に関わらず生じる関数関係で見せかけの相関である。したがって、これらのモデルから予測される相関は、per capita間の相関でも、これらの対数変換値間の相関でも、①人口に対してもに逓増型($\beta_1 > 1$ かつ $\beta_2 > 1$)の指標間、または、人口に対してもに逓減型($\beta_1 < 1$ かつ $\beta_2 < 1$)の指標間では正の見せかけの相関、②人口に対し逓増型と逓減型の組み合わせ($\beta_1 > 1$ かつ $\beta_2 < 1$ か、 $\beta_1 < 1$ かつ $\beta_2 > 1$)の場合は負の見せかけの相関が検出されることが予測される。

以下の3つの例で、per capitaやこれの対数変換値を用いることによって人口を介した見せかけの相関が入り込むことを検証しよう。

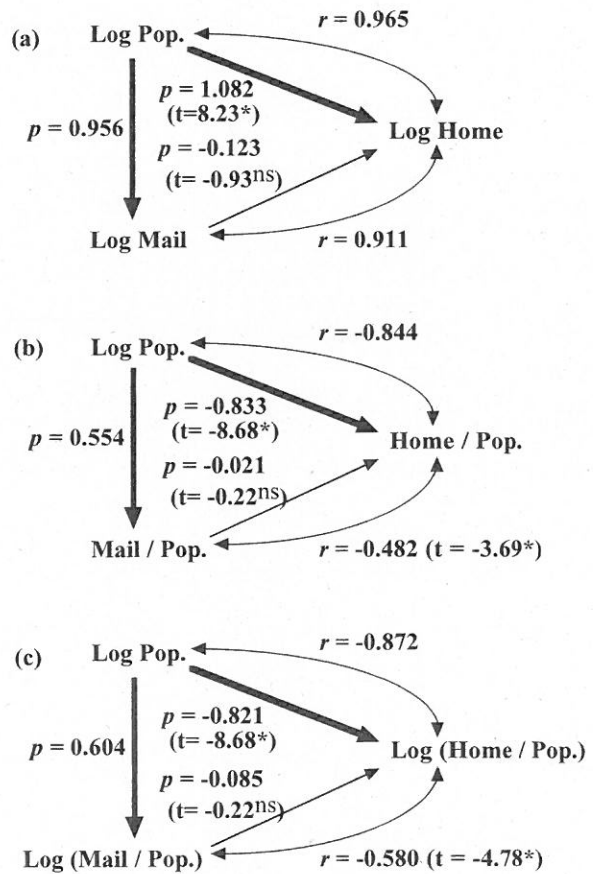


Figure 1. Path analyses decomposing the spurious correlations into the direct effect and the indirect effect among the variables: the population size, the number of mails and the number of the homes for the aged. (a) Path analysis among the logarithm of the observed values. (b) Path analysis among the per capita of the observed values. (c) Path analysis among the logarithm of the per capita values.

(図1 日本の都道府県郵便物数(Mail)と老人介護施設数(Home)とのパス解析。(a)実測値の対数変換値間でのパス解析。(b)per capita間でのパス解析。(c)per capitaの対数変換値間でのパス解析。)

①例1 都道府県郵便物数と老人介護施設数との相関 (Figure 1)

日本の都道府県の郵便物数($\beta_2 = 1.3052$)と老人介護施設数($\beta_1 = 0.6736$)の相関をパス図を用いて分析しながら問題の所在を検討していこう。上述のベキ関数モデルによる予測に従うとすれば、 $\gamma = -1.07$ と推定されるから強い負の見せかけの相関が検出されてもおかしくない。これらの変数と人口の対数値を用いてパス解析を行うとFigure 1(a)に見られるような3変数の関係を得る。郵便物数(Mail)と施設数(Home)との直接効果(パス係数: 標準偏回帰係数)は $-0.123(t=-0.93^{ns})$ であり、まったく有意ではない。人口効果を除く1次偏相関係数の値は、 $-0.138(t=-0.93^{ns})$ 、人口とGDPの効果をともに除いた2次偏相関係数は $-0.104(t=-0.67^{ns})$ で有意な効果は検出されなかった。

これら二つの変数のper capitaの値をとって同様なパス解析を行うとper capita間の相関では、 $-0.482(t=-3.69^*)$ となりFigure 1(b)の結果を得る。これは直接効果の $-0.021(t=-0.22^{ns})$ に人口を経由した間接効果(見せかけの相関) $-0.833 \times 0.554 = -0.461$ が上積みされた値であることがわかる。

per capitaの対数変換値をとって同様なパス解析を行うとFigure 1(c)の結果を得る。たしかに郵便物数と施設数のper capitaの対数変換値間の相関は、予測していたように $-0.580(t=-4.78^*)$ となるが、これは直接効果の $-0.085(t=-0.22^{ns})$ に人口を経由した間接効果(見せかけの相関) $-0.821 \times 0.604 = -0.496$ が上積みされた値であることがわかる。

この例は、本来有意な負の相関が検出されるべきではないのにper capitaまたはこの対数変換値を用いたことによって引き起こされた見せかけの相関である。つまり、per capita間、または、これらの対数変換値をとって回帰分析を行っていることは、人口効果を除いているのではなく、人口で割ることによって違った形で人口効果を埋め込んでいることを示している。これが今まで見過ごされてきた“人口あたりの数値”で回帰分析することの問題点である。

②例2 都道府県県民所得と失業者数との相関 (Figure 2)

日本の都道府県県民所得(Income, $\beta_2 = 1.122$)と完

全失業者数(Unemployment, $\beta_1 = 1.027$)の間では、ベキ関数の法則に従えば有意になるかならないか位の正の相関であっても良いのであるが、Figure 2に示すように、相関係数を計算するとper capita間では、 $-0.392(t=-2.86^*)$ 、これの対数変換値間では、 $-0.433(t=-3.22^*)$ と有意な負の値となる。予測に反した負の相関が検出されるということは、実際にこれらの間に見せかけの正の相関をしのぐ強い負の効果が働いていると推測でき①と同様な計算で負の強い直接効果が人口を介した正の効果によって打ち消されていることがわかる。この場合は、統計的有意性の判断には影響がなかったが県民所得と失業者数との直接効果(-0.773)が人口を介した相関($0.340=0.600 \times 0.568$)によって有意性が過小評価された例に相当する。

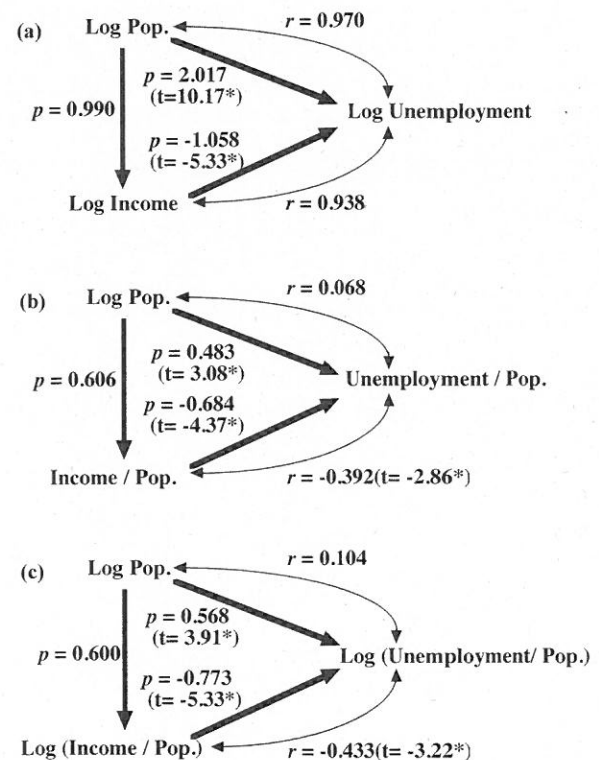


Figure 2. Path analyses decomposing the spurious correlations into the direct effect and the indirect effect among the variables: the population size, the personal income and the unemployment. (a) Path analysis among the logarithm of the observed values. (b) Path analysis among the per capita of the observed values. (c) Path analysis among the logarithm of the per capita values.

(図2 日本の県民所得 (Income) と失業者数 (Unemployment) とのパス解析。

(a) 実測値の対数変換値間でのパス解析. (b) per capita間でのパス解析. (c) per capitaの対数変換値間でのパス解析.)

③例3 アメリカの殺人数と強姦数との相関 (Figure 3)

アメリカ FBI の犯罪記録は7-8種の項目に分類されている。この中で強姦(Forcible rape)だけは人口に対して逡減型の増加を示すが、他種の犯罪はかなり強い逡増型の増加を示す。すると例えば殺人数(Murder)と強姦数との相関は強い負の見せかけの相関が検出されても良いはずである。しかし、各種犯罪に共通する強い直接的効果が働きかけているならば、結果として無相関かやや正の相関が検出されるくらいであろうことが推測される。

パス解析結果をFigure 3に示す。予想通り per capita間(0.123, $t=0.95^{ns}$, Fig.3(b))でも、対数変換値間(0.161, $t=1.13^{ns}$, Fig.3(c))でも有意な相関とは

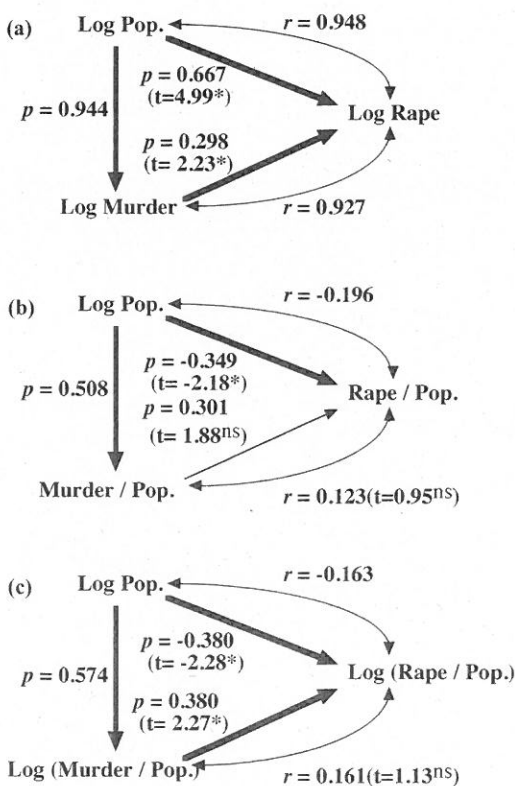


Figure 3. Path analyses decomposing the spurious correlations into the direct effect and the indirect effect among the variables: the population size, the number of murders and the number of the forcible rapes. (a) Path analysis among the logarithm of the observed values. (b) Path analysis among the per capita of the observed values. (c) Path analysis among the logarithm of the per capita values.

(図3 アメリカの州の殺人数 (Murder) と強姦 (Rape) とのパス解析。(a) 実測値の対数変換値間でのパス解析。(b) per capita 間でのパス解析。(c) per capita の対数変換値間でのパス解析。)

なっていない。人口効果が介入した結果、有意な正の直接効果が相殺されて有意でなくなったことが明瞭にわかる。人口とGDPの効果をとともに除いた2次偏相関係数は $0.306(t=2.16^*)$ となる。殺人と強姦の対数値をとってパス解析するとFigure 3(a)に示すように、殺人と強姦との間に有意な効果($0.298, t=2.23^*$)が検出される。殺人率、強姦率の対数変換値をとってパス解析するとFigure 3(c)に示す結果になる。直接効果は有意な $0.380(t=2.27^*)$ であるが、人口を介する間接効果($-0.380 \times 0.574 = -0.218$)によって相殺され有意ではない相関 $0.161(t=1.13^{ns})$ になったことがわかる。有意性を検出すべきところが間接効果によって検出されなくなったのである。

以上の例を見ると、per capita間、またはこの対数変換値間でベキ関数の予測に合う統計的検定の有意性が検出された時は、相関係数の値にして2-3割増しくらいの間接効果が加算された見せかけの相関の有意性を判断していることになる。逆に、ベキ関数の予測にしたがわず、無相関が検出されたり、逆方向の相関が検出される場合は間接効果に抗した強い直接効果が働いていることの証拠となるのである。

これらの分析から見せかけの相関がまぎれこむ過程が理解でき、人口効果がper capitaという値を用いることによって埋め込まれることを確認でき、回帰分析でper capitaの値を用いる危険性が理解されたのではないだろうか。つまり、per capitaやこの対数変換値を従属変数や独立変数に用いた回帰分析はほとんど意味がないことになる。

それでは実際に、①このような人為的に間違った判断がどれ位の頻度で起こっているのか、そして、②これを克服する統計的手法があるのかということを考えてみよう。

実は、通常これらの検討はなかなか難しいのだが、アメリカのソーシャル・キャピタルの因果関係を定量、検証、考察している過程でいくつかの有効な知見を得たので次に紹介する。アメリカの社会指標を検討対象としたのは途中、南北戦争という大きな変動があったとはいえ、アメリカ各州間の地域特性がかなり安定的に社会指標に反映

されているように感じられ(日下部 2012), 因果関係の検証に用いる社会指標として有効であると感じられたからである。

4. 因果関係を検証する2次偏相関係数とその検出力

(1) 見せかけの相関を克服する2次偏相関係数
 一般に, このような見せかけの相関を除くためには第3の変数の効果を除く“偏相関係数”の統計的有意性を考える。Figure 4(a)に示すように, 2変数 X_1, X_2 は第3の変数 X_3 の直接的効果を受けている。この時, 2変数 X_1, X_2 の正味の相関を見るにはそれぞれの残差 U_1 と U_2 の相関を検討しなければならない。これが変数 X_3 の効果を除いた X_1, X_2 の1次偏相関である。したがって, 人口効果を除くためには人口に対する回帰偏差(日下部 2002, Ley 1972 のいう partial score または residual score), U_1 と U_2 間の相関を測らねばならない。

3変数を X_1, X_2, X_3 とし, これらの相関係数を

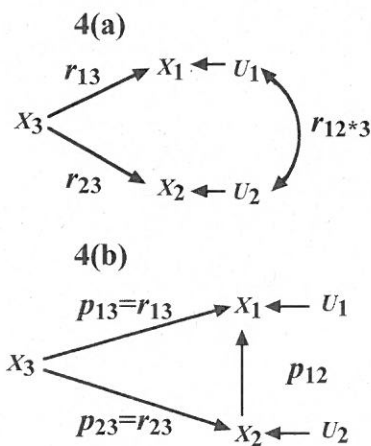


Figure 4. Path diagrams expressing the first-order partial correlation (a) and the direct effect between variables (b).
 (図4 (a) 1次偏相関係数を表すパス図。(b) 直接効果を表すパス図。)

r_{12}, r_{23}, r_{13} とすれば, X_3 の効果をのぞいた X_1, X_2 の偏相関係 r_{12*3} は,

$$r_{12*3} = \frac{r_{12} - r_{13} \times r_{23}}{\sqrt{1 - r_{13}^2} \times \sqrt{1 - r_{23}^2}}$$

となる。

Figure 1~3で検討してきた見せかけの相関が介入してくる過程を例示したモデルはFigure 4(b)である。2変数 X_1 と X_2 の間に見られる相関 r_{12} は, 直接効果を示すパス係数 p_{12} と変数 X_3 から2変数 X_1, X_2 への直接効果 p_{13} と p_{23} との積 $p_{13} \times p_{23}$ との和, $r_{12} = p_{12} + p_{13} \times p_{23}$ となる。つまり, 例えば X_3 を人口とすれば絶えず人口を介した効果 $p_{13} \times p_{23}$ が介入してくることになる。これが見せかけの相関を引き起こす。よく間違いが起こるので注意しておきたいが, パス係数(p)は標準偏回帰係数を意味しているのでその絶対値はかならずしも1以内になるわけではなく, 1を越える場合もある(例えばFigure 2(a))。

多くの社会指標には人口と同時に産業活動を表すGDPが影響を与えており, 人口と同様にGDPの対数値へ回帰するとすべて高い相関を示す(表1)。したがって, 社会指標間の因果関係の検証としては, 人口効果だけを除いて考えるだけでなく, 同時にGDPの効果も除いて考えた方がよいと判断される。これはFigure 5(a)に示す X_1, X_2, X_3, X_4 の4変数のパス図として表現される。2変数 X_3 と X_4 の効果を除いて2変数 X_1, X_2 の相関を検出するためには U_1 と U_2 間の相関(r_{12*34})を見なければならない。これが2次の偏相関係数である。これは付録の解説で示すように次の式で表現される。1行目の式は, 付録のFigure A2から誘導される1次偏相関係数から成る式であり, 2行目の式はFigure 5(a)から誘導される通常の2変数相関係数から成る式で, やや煩雑ではあるが同値であることが証明される。

$$r_{12*34} = \frac{r_{12*3} - r_{14*3} \times r_{24*3}}{\sqrt{1 - r_{14*3}^2} \times \sqrt{1 - r_{24*3}^2}}$$

$$= \frac{r_{12} - r_{12}r_{34}^2 - r_{13}r_{23} - r_{14}r_{24} + r_{14}r_{23}r_{34} + r_{13}r_{24}r_{34}}{\sqrt{(1 - r_{34}^2) - (r_{13}^2 + r_{14}^2 - 2r_{13}r_{14}r_{34})} \times \sqrt{(1 - r_{34}^2) - (r_{23}^2 + r_{24}^2 - 2r_{23}r_{24}r_{34})}}$$

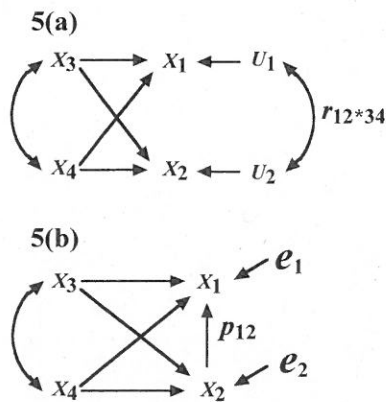


Figure 5. Path diagrams expressing the second-order partial correlation (a) and the direct effect between variables in the recursive model in the structural equation modeling (b).
(図5 (a) 2次偏相関係数を表すパス図 (b) 4変数逐次モデルの構造方程式分析における直接効果を表すパス図.)

(2) パス解析の直接効果と2次偏相関係数の有意性検出力の比較

2次偏相関係数が実際どれほどの検出力があるかを示すために、アメリカの州毎の社会指標10項目(表1の10項目:NPO数, 下院議員選挙投票者数, 殺人数, 公的NPO支援費, 個人所得, 地方行政の教育費支出, 地方行政の福祉費支出, 自殺者数, 失業者数, 非ヒスパニック系白人居住者数)

間の45の組み合わせで相関分析を行った。同時に因果関係の検証法として用いられる構造方程式分析 (Arbuckle & Wothke 1989, Bollen 1989, Schipley 2000, Pearl 2009) による直接効果の推定と有意性を解析した。構造方程式分析のパスモデルは人口とGDPと分析対象の2変数の4変数逐次モデルでFigure 5(b)に相当し、モデルの構造上、この2変数間の直接効果(p_{12})の有意性は2次偏相関係数(r_{12*34} , Figure 5(a))の有意性とほぼ等価であることが推測される。

分析結果を表2に示す。1行目のゴシック体の数字が2次偏相関係数で2行目のイタリック体が構造方程式分析における直接効果である。これらの中にはある程度のばらつきが見られるが()内に示したt値はほぼ同じである。また、1例を除いてすべてで2次偏相関係数のt値がやや小さめになっている。影をつけた二つの相関(白人居住者数と殺人, および白人居住者数と自殺)で5%有意水準 $t = 2.01$ を挟んでいるがほぼ等しいと考えて良い。つまり、構造方程式分析における直接効果の有意性と2次の偏相関係数の有意性はほとんど同じと考えてよい。クロスセクションデータ分析において2変数間の直接的効果を検証する方法としておそらく2次偏相関係数を用いるのは適

表2 アメリカの各種社会指標間の相関分析。1行目は2次偏相関, 2行目は共分散構造分析における直接効果。()内はt値。

(Table 2. Correlation matrix among the social indicators in America. In each cell, the second-order partial correlations are expressed in bold numerals, and those of the direct effects in the structural equation model are in italics. The shaded cells indicate the discrepancies between them. The t values are expressed in the parentheses.)

Vote	Murder	Pub. Sup.	Income	Edu.	Welfare	Suicide	Unemployment	White	NPO
0.725 (7.05) <i>1.014 (7.36)</i>	-0.562 (-4.56) <i>-0.680 (-4.76)</i>	0.741 (7.40) <i>0.870 (7.73)</i>	0.523 (5.07) <i>0.150 (4.30)</i>	0.449 (3.37) <i>0.189 (3.52)</i>	0.452 (3.40) <i>0.505 (3.55)</i>	-0.217 (-1.49) <i>-0.235 (-1.56)</i>	0.050 (0.33) <i>0.082 (0.35)</i>	0.291 (2.04) <i>0.422 (2.13)</i>	
	-0.443 (-3.31) <i>-0.749 (-3.46)</i>	0.463 (3.50) <i>0.760 (3.65)</i>	0.459 (3.46) <i>1.141 (3.62)</i>	0.212 (1.45) <i>0.360 (1.52)</i>	0.220 (1.51) <i>0.275 (1.58)</i>	0.048 (0.33) <i>0.063 (0.34)</i>	0.149 (1.01) <i>0.126 (1.05)</i>	0.422 (3.12) <i>0.437 (3.25)</i>	
		-0.212 (-1.45) <i>-0.218 (-1.52)</i>	-0.427 (-3.16) <i>-0.102 (-3.30)</i>	-0.118 (-0.80) <i>-0.041 (-0.83)</i>	-0.096 (-0.65) <i>-0.071 (-0.68)</i>	0.228 (1.57) <i>0.175 (1.64)</i>	0.229 (1.57) <i>0.114 (1.64)</i>	-0.277 (-1.94) <i>-0.158 (-2.02)</i>	
			0.453 (3.41) <i>0.111 (3.56)</i>	0.241 (1.67) <i>0.674 (1.74)</i>	0.577 (4.73) <i>0.755 (4.94)</i>	-0.270 (-1.88) <i>-0.213 (-1.96)</i>	0.124 (0.84) <i>0.241 (0.88)</i>	0.030 (0.20) <i>0.052 (0.21)</i>	
				0.137 (0.93) <i>0.200 (0.97)</i>	0.132 (0.89) <i>0.410 (0.93)</i>	-0.351 (-2.52) <i>-0.109 (-2.63)</i>	-0.310 (-2.19) <i>-0.148 (-2.28)</i>	0.165 (1.13) <i>0.069 (1.17)</i>	
					0.473 (3.60) <i>0.222 (3.75)</i>	-0.246 (-1.71) <i>-0.112 (-1.78)</i>	0.240 (1.66) <i>0.167 (1.73)</i>	0.159 (1.08) <i>0.097 (1.12)</i>	
						-0.440 (-3.29) <i>-0.425 (-3.43)</i>	0.243 (1.68) <i>0.360 (1.75)</i>	0.018 (0.12) <i>0.023 (0.12)</i>	
							0.113 (0.76) <i>0.073 (0.79)</i>	0.279 (1.95) <i>0.208 (2.04)</i>	
								0.179 (1.22) <i>0.205 (1.27)</i>	
									Unemployment

切であると判断できる。

(3) per capitaの値とこれの対数変換値の有意性検出力の比較

因果関係を検証する手法として構造方程式解析と同じように2次偏相関係数が可能であることを示した。そこで、(2)で分析した45のすべての組み合わせで、2次偏相関係数値と、独立変数と従属変数とともにper capitaの値を使ったときの相関係数値とこの対数変換値を用いたときの相関係数値およびその有意性を評価する。

結果は表3の各セルの2行目にper capitaの相関、3行目にこれの対数值間の相関を示す。影をつけた9つの組み合わせ(9/45, 20%)で有意性判定についての不一致が見られる。違いを見るとあきらかにper capitaとこの対数変換値が同調して2次偏相関係数からずれている。ということは、人口効果が介入した間接効果が見せかけの相関を引き起こしている可能性が強い。Per capitaやこ

の対数変換値を用いた回帰分析は、独立変数の効果を判断する検出法としては全く使えないことがはっきりとわかる。

先に例③で少し触れたが、アメリカの刑犯罪は7-8種の項目で記録されている。それらの7項目間で同様の分析を行って表4に示す。2次偏相関係数は21の全ての組み合わせで高い相関を示している。つまりアメリカ50州のそれぞれの地域社会に犯罪が起こり易い土壌とでも表現される因子があると考えられる結果である。しかし、per capitaやこの対数変換値を用いた場合は強姦との組み合わせの3つの場合(3/21)で不一致が起こっている。これらは、すでに例③で明らかにした人口効果の介入によって生じた見せかけの相関である。

以上の分析は、per capita やこの対数変換値を用いた回帰分析では独立変数の効果を判断している2割ほどの場合で間違った判断をしている可能性が高いということになる。また、因果関係の検

表3 アメリカの各種社会指標間の相関分析。1行目は2次偏相関、2行目はper capita間の相関係数、3行目はこの対数変換値間の相関係数。()内はそれぞれのt値である。

(Table 3. Correlation matrix among the social indicators in America. In each cell, the second-order partial correlations are expressed in bold numerals, and those of the per capita data (the second line in each cell) and the logarithm(the third line in each cell) are in italics. The shaded cells indicate the discrepancies among them. The t values are expressed in the parentheses.)

Vote	Murder	Pub. Sup.	Income	Education	Welfare	Suicide	Unemployment	White	
0.725(7.05)	-0.562(-4.56)	0.741(7.40)	0.523(5.07)	0.449(3.37)	0.452(3.40)	-0.217(-1.49)	0.050(0.33)	0.291(2.04)	NPO
<i>0.768(8.31)</i>	<i>-0.649(-5.91)</i>	<i>0.492(3.92)</i>	<i>0.338(2.49)</i>	<i>0.562(4.71)</i>	<i>0.478(3.77)</i>	<i>0.085(0.59)</i>	<i>-0.172(-1.21)</i>	<i>0.331(2.43)</i>	
<i>0.784(8.13)</i>	<i>-0.674(-6.31)</i>	<i>0.599(5.18)</i>	<i>0.418(3.19)</i>	<i>0.565(4.74)</i>	<i>0.456(3.55)</i>	<i>-0.031(-0.21)</i>	<i>-0.157(-1.10)</i>	<i>0.369(2.75)</i>	
-0.443(-3.31)	0.463(3.50)	0.469(3.46)	0.212(1.45)	0.220(1.51)	0.048(0.33)	0.149(1.01)	0.422(3.12)		Vote
<i>-0.607(-5.30)</i>	<i>0.249(1.78)</i>	<i>0.175(1.23)</i>	<i>0.280(2.02)</i>	<i>0.256(1.83)</i>	<i>0.249(1.78)</i>	<i>0.219(1.53)</i>	<i>-0.054(-0.38)</i>	<i>0.483(3.82)</i>	
<i>-0.601(-5.20)</i>	<i>0.321(2.34)</i>	<i>0.231(1.64)</i>	<i>0.293(2.13)</i>	<i>0.259(1.86)</i>	<i>0.216(1.53)</i>	<i>-0.044(-0.30)</i>	<i>0.552(4.59)</i>		
-0.212(-1.45)	-0.427(-3.16)	-0.118(-0.80)	-0.096(-0.65)	0.228(1.57)	0.229(1.57)	-0.277(-1.94)			Murder
<i>-0.161(-1.13)</i>	<i>-0.189(-1.33)</i>	<i>-0.233(-1.66)</i>	<i>-0.174(-1.22)</i>	<i>-0.055(-0.38)</i>	<i>0.314(2.29)</i>	<i>-0.551(-4.58)</i>			
<i>-0.097(-0.67)</i>	<i>-0.150(-1.05)</i>	<i>-0.207(-1.47)</i>	<i>-0.154(-1.07)</i>	<i>-0.071(-0.49)</i>	<i>0.359(2.67)</i>	<i>-0.416(-3.17)</i>			
	0.453(3.41)	0.241(1.67)	0.577(4.73)	-0.270(-1.86)	0.124(0.84)	0.030(0.20)			Public support
	<i>0.662(6.12)</i>	<i>0.455(3.54)</i>	<i>0.563(4.72)</i>	<i>-0.326(-2.39)</i>	<i>0.090(0.63)</i>	<i>-0.070(-0.49)</i>			
	<i>0.676(6.36)</i>	<i>0.449(3.48)</i>	<i>0.522(4.24)</i>	<i>-0.422(-3.23)</i>	<i>0.081(0.56)</i>	<i>0.006(0.04)</i>			
		0.137(0.93)	0.132(0.89)	-0.351(-2.52)	-0.310(-2.10)	0.165(1.13)			Income
		<i>0.472(3.71)</i>	<i>0.193(1.37)</i>	<i>-0.447(-3.46)</i>	<i>-0.208(-1.48)</i>	<i>-0.082(-0.57)</i>			
		<i>0.481(3.80)</i>	<i>0.155(1.09)</i>	<i>-0.505(-4.05)</i>	<i>-0.202(-1.43)</i>	<i>-0.051(-0.36)</i>			
			0.473(3.60)	-0.246(-1.71)	0.240(1.66)	0.159(1.08)			Education
			<i>0.493(3.93)</i>	<i>-0.168(-1.18)</i>	<i>0.071(0.49)</i>	<i>0.041(0.29)</i>			
			<i>0.467(3.66)</i>	<i>-0.274(-1.98)</i>	<i>0.069(0.48)</i>	<i>0.121(0.85)</i>			
				-0.440(-3.29)	0.243(1.68)	0.018(0.12)			Welfare
				<i>-0.278(-2.00)</i>	<i>0.171(1.20)</i>	<i>0.020(0.14)</i>			
				<i>-0.353(-2.62)</i>	<i>0.176(1.23)</i>	<i>0.109(0.76)</i>			
					0.113(0.76)	0.279(1.95)			Suicide
					<i>-0.025(-0.18)</i>	<i>0.324(2.38)</i>			
					<i>-0.011(-0.08)</i>	<i>0.204(1.44)</i>			
						0.179(1.22)			Unemployment
						<i>0.005(0.04)</i>			
						<i>0.145(1.01)</i>			

表4 アメリカ50州の刑犯罪7種についての相関分析。1行目は2次偏相関、2行目はper capita 間の相関係数、3行目はこの対数変換値間の相関係数。()内はt値である。
 (Table 4. Correlation matrix among the crimes in America. In each cell, the second-order partial correlations are expressed in the bold numerals and those of the per capita data and the logarithm are in the Italics. The t values are expressed in the parentheses.)

Rape	Robbery	Assault	Burglary	Larceny-theft	Motor vehicle theft	
0.306 (2.16)	0.757 (7.76)	0.690 (6.39)	0.562 (4.56)	0.308 (2.17)	0.423 (3.13)	Murder
<i>0.123 (0.95)</i>	<i>0.754 (7.95)</i>	<i>0.667 (6.21)</i>	<i>0.623 (5.51)</i>	<i>0.268 (1.93)</i>	<i>0.428 (3.28)</i>	
<i>0.161 (1.13)</i>	<i>0.799 (9.21)</i>	<i>0.731 (7.42)</i>	<i>0.622 (5.51)</i>	<i>0.332 (2.44)</i>	<i>0.547 (4.53)</i>	
	0.333 (2.37)	0.500 (3.87)	0.528 (4.17)	0.418 (3.08)	0.385 (2.80)	Rape
	<i>-0.034 (-0.23)</i>	<i>0.445 (3.45)</i>	<i>0.357 (2.64)</i>	<i>0.326 (2.39)</i>	<i>0.203 (1.44)</i>	
	<i>0.071 (0.49)</i>	<i>0.388 (2.91)</i>	<i>0.431 (3.31)</i>	<i>0.389 (2.93)</i>	<i>0.267 (1.92)</i>	
		0.611 (5.18)	0.638 (5.56)	0.394 (2.88)	0.577 (4.74)	Robbery
		<i>0.566 (4.75)</i>	<i>0.431 (3.31)</i>	<i>0.162 (1.14)</i>	<i>0.511 (4.12)</i>	
		<i>0.657 (6.03)</i>	<i>0.552 (4.58)</i>	<i>0.314 (2.29)</i>	<i>0.669 (6.24)</i>	
			0.666 (5.99)	0.528 (4.17)	0.507 (3.95)	Assault
			<i>0.622 (5.51)</i>	<i>0.438 (3.37)</i>	<i>0.371 (2.76)</i>	
			<i>0.655 (6.00)</i>	<i>0.515 (4.17)</i>	<i>0.594 (5.11)</i>	
				0.736 (7.30)	0.672 (6.09)	Burglary
				<i>0.687 (6.55)</i>	<i>0.522 (4.24)</i>	
				<i>0.733 (7.46)</i>	<i>0.656 (6.02)</i>	
					0.656 (5.84)	Larceny-theft
					<i>0.506 (4.06)</i>	
					<i>0.621 (5.49)</i>	

証法として2次偏相関係数はきわめて有効であることを強く示している。

5. 実例：地方議会議員定数と地方行政支出の相関

社会科学の中で最もよくOLS回帰分析が用いられるのは計量経済学であろう。最近、英文誌 *Journal of Public Economics* に掲載された Per Pettersson-Lidbom (2012) の論文を検討してみる。取り扱われている問題は、地方議会議員数と地方行政支出の関係で、「一般には、「議員数が増えると行政支出も増える」ように考えるだろうが、フィンランドとスウェーデンの分析から得られた結果は逆で、「議員数が増えると行政支出は減る」、つまり「負の有意な相関が検出された」というものである。この分析結果から、このような効果もたらされる可能性について議論を展開している。

この論文の最大の欠陥は“どのような測定値に

対して回帰分析を行っているか”についての問題意識が全くないことである。つまり、per capita やこの対数変換値を回帰分析に用いることの人為的効果についての配慮が全くなく従属変数や独立変数として使われている。表(Table 3-6 in Pettersson-Lidbom 2012)の下にNotesとして記載されているのは独立変数にper capita spending の対数値、独立変数にCouncil size (議員数) の対数値と記載されている。

一般に、社会指標のクロスセクションデータでは上述したように“ベキ関数の法則”(日下部2011)が経験的に成り立つ。後の日本とアメリカの分析で示すように地方議会議員数も地方行政支出もほぼ人口のベキ関数として表現できる。今、人口を x 、支出を y_1 、議員数を y_2 、とすれば、 $y_1 = a_1 * x^{\beta_1}$ 、 $y_2 = a_2 * x^{\beta_2}$ 、と表現される。Pettersson-Lidbom が従属変数(Y)とした log per capita spending は $\log(y_1/x) = \log(a_1 * x^{\beta_1 - 1}) = \log a_1 + (\beta_1 - 1) \log x$ と表現され、独立変数(X)とした log council size は $\log y_2 = \log a_2 + \beta_2 \log x$ と表現される。人口の対数値を

独立変数($X = \log x$)とすれば、 $Y = \log \alpha_1 + (\beta_1 - 1) * X$ の人口対数値との直線関係で表される。また、 \log council sizeを独立変数($X = \log y_2$)とすれば、 $Y = \gamma * X - \log \alpha_1 + \log \alpha_2$, ($\gamma = (\beta_1 - 1) / \beta_2$), となり、議員数の対数値との関係は、おおむね係数 γ の直線として表現され、因果関係の有無に関わらず有意な相関が検出されてもおかしくないのである。一般に、行政支出は人口にたいし逓減型の増加を示すので β_1 は1より小さく、したがって直線の傾き γ は負になることが予測される。フィンランドとスウェーデンのデータが手に入らないので直接検証することが難しいがPettersson-Lidbomの分析結果は単なる見せかけの相関である可能性が強い。日本とアメリカの地方議員数の例を分析しながらこれを示す。

(1) 日本の地方議員数と地方行政支出との相関

日本の都道府県議会議員数と行政支出を対数変換して人口の対数値に回帰するとおおむね直線回帰できる(表5)。これらの対数変換値を用いてパス解析を行うとFigure 6(a)に示すパス図を得る。都道府県人口とGDP, 地方議会議員数, 行政支出との間の4変数逐次モデルを設定し、AMOSでパス解析すると地方議会議員数と行政支出との間に正の有意な直接効果(0.275, $t=2.52^*$)が検出できる。人口とGDPの効果をとものにのぞいた2次の偏相関係数($r_{12 \cdot 34}$)は0.299($t=2.23^*$)となり、直接効果よりやや大きい。つまり“議員数が増えれば

行政支出は増える”ということが明らかである。

以上のような、議員数と行政支出との関係を、Pettersson-Lidbom (2012)が分析した手法にならって行ってみるとどうなるであろうか？まず、単純に人口で割ったper capitaどうしを比較してみる。Figure 7(a)で示すように“正の有意な相関”を示す。これは“ベキ関数の法則”によって予測できる当然の結果であり、当然、見せかけの相関が疑われる。ここで注意したいのは(支出/人口)の値である。行政支出は人口に対する逓減型の増加関数だから人口で割った値にすると無条件に人口効果が埋め込まれる。これは人口の対数値を横軸にとって表したFigure 7(b)に示されている。この傾向は、Pettersson-Lidbom (2012)が従属変数として用いた(支出/人口)の対数値をとっても変化しない(Figure 7(c))。Per capitaの支出の対数値は人口の対数値に対し、 $Y = \log \alpha_1 + (\beta_1 - 1) * X$ の直線関係で表現されたので傾きを計算すると $\beta_1 - 1 = 0.912 - 1 = -0.088$ となりFigure 7(c)の回帰直線の傾き

表5 日本とアメリカの地方議会議員数と地方歳出についての人口にたいする回帰分析。

(Table 5. The regression analysis for the log-transformed council size and the expenditure both in Japan and America. β is the slope of the regression line, and R is the correlation coefficient.)

Dependent	β	Correlation (r)
JAPAN		
Council Size	0.435	0.9774
Expenditure	0.912	0.9863
AMERICA		
Council Size	0.163	0.3387
Expenditure	0.982	0.9852

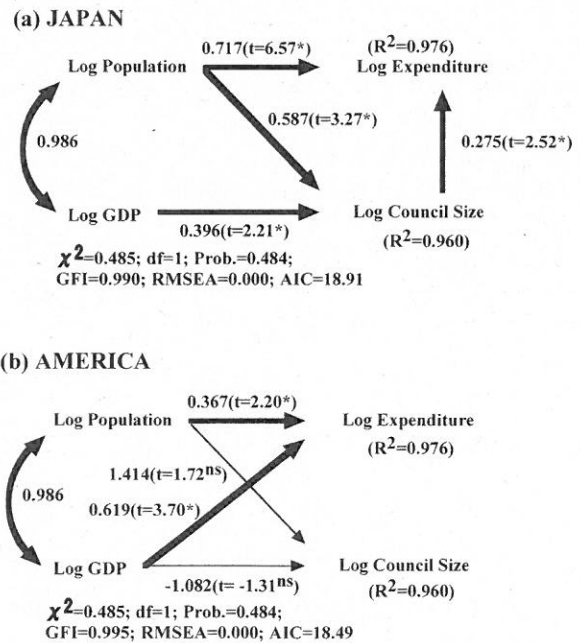


Figure 6 Estimated causal model for council size and expenditure both in Japan and America.

Figure 6. Path analysis showing the direct positive effect between the council size and the expenditure in the Japanese local governments (a). In America, the effect cannot be detected (b).

(図6 日本(a)とアメリカ(b)における地方議会議員数と地方歳出との共分散構造分析。)

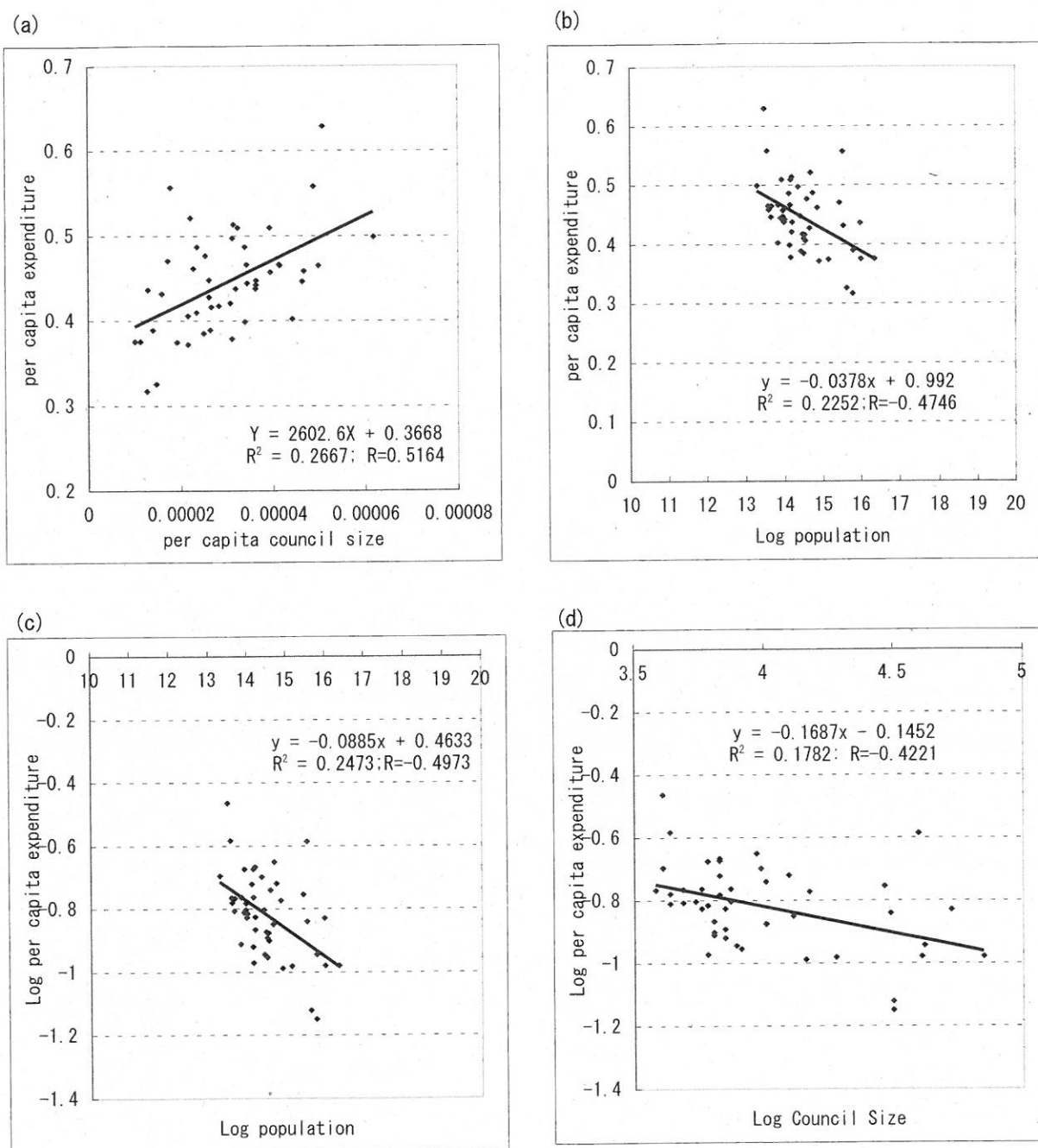


Figure 7. The regression analyses between the council size and the expenditure. (a) The regression between the per capita observed values. (b) The regression of the per capita expenditure on the population size. (c) The regression of the logarithm of the per capita expenditure on the logarithm of the per capita expenditure. (d) The regression of the logarithm of the per capita expenditure on the logarithm of the council size.

(図7 日本の地方議員数と地方歳出との回帰分析。(a) per capita 間の回帰分析。(b)人口の対数値にたいする per capitaの地方歳出の回帰分析。(c)人口の対数値にたいする per capitaの地方歳出の対数値の回帰分析。(d) 議員数の対数値にたいする per capitaの地方歳出の対数値の回帰分析。)

-0.0885 にほぼ一致している。また独立変数を議員数の対数値とすると予測される回帰直線の傾きは $\gamma = (0.912-1)/0.435 = -0.2023$ となり Figure 7(d) の傾き -0.1687 に近い。先にベキ関数モデルを用

いて演繹したように、因果関係の有無にかかわらず負の関数関係が表われるのである。

これを Pettersson-Lidbom (2012) が回帰分析したと想定されるように、人口の対数値、議員数の対

数値と per capita spending (行政支出) の間でパス解析を行って Figure 8 に示す。Pettersson-Lidbom (2012) が検出した負の相関は確かに $-0.422(t=-3.23)$, Figure 8(b)) となるが, これはすでに詳述したように直接効果 1.431 と人口を介した効果 $0.977 \times (-1.896)$ の和によってもたらされた見せかけの相関であることが明瞭に理解できる。おそらく, Pettersson-Lidbom (2012) はフィンランドやスウェーデンの町村の分析で同じような結果を真の直接的負の効果と判断して議論をすすめたのであろう。

(2) アメリカの地方議員数と地方行政支出との相関

アメリカの議員数は人口に対し強い逓減型の増加を示し, 対数変換値の相関は低い (0.339^* , 表 5)。つまり, アメリカでは議員数はあまり人口に依存しない, つまり, 都市人口の大小に関わらず議員

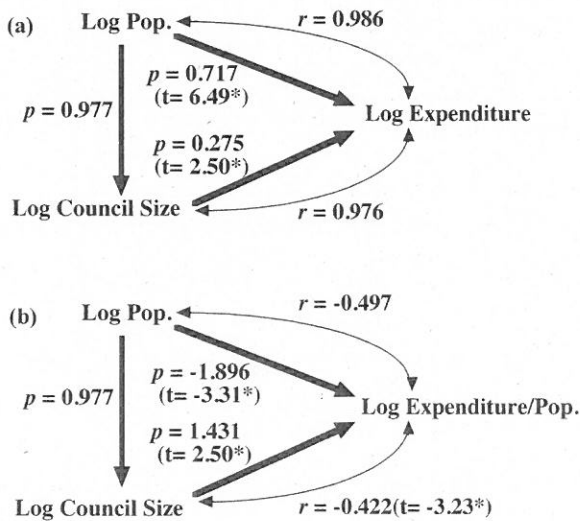


Figure 8. Path analyses decomposing the spurious correlations between the council size and the expenditure into the direct effect and the indirect effect via the population size. (a) Path analysis among the logarithm of the observed values: the population size, the council size and the expenditure. (b) Path analysis among the variables: the logarithms of the population size, the council size and the logarithm of the per capita expenditure. Note that the negative spurious correlation ($r=-0.422$) is contradictory to the positive direct effect ($p=1.431$) between the log council size and the log per capita expenditure.

(図 8 日本の地方議会議員数と地方歳出とのパス解析。(a) 対数値間でのパス解析。(b) per capita 歳出の対数値と人口, 議員数の対数値とのパス解析.)

数の上限が見られるのである。したがってパス解析を行うと Figure 6(b) で示したように議員数に対する人口と GDP の効果が検出されず議員数の独立性が示される。当然, 議員数と行政支出との間に直接効果は検出されない。日本と対照的な結果である。

6. 考察

回帰分析において従属変数や独立変数に per capita やこの対数変換値などを用いた場合, “見せかけの相関”が生じることは理解できたであろう。したがって, 従来“人口効果を除くために”といって人口で割っていた操作は実は“人口効果を埋め込む”ことになっていたのである。これが今日まで, 特に計量経済学を中心として慣習的に踏襲されてきた回帰分析手法における不適切な, かつ間違った点なのである。この問題点を指摘したものは今まで洋の東西を問わず誰もいない。それほど“常識的な慣習”だったのである。

本論では典型的なクロス・セクションデータを用いて見せかけの相関が生じてくる過程を分析してきたが, これは定量的評価を正確にしたいためであった。本論で示したような per capita の値が持つ問題は, 多かれ少なかれ, 人間活動のおよぶ所では常に見られるのである。例えば, 都市問題や過疎問題を論じる時に当然, 定量的考察が必要となる。この時, per capita で比較するような状況は広く社会科学全般から日常生活にまで見られる。このような時にも問題によっては対象となる二つの現象の背後に人口を介した効果が多く紛れ込んでいる可能性を考える必要がある。パス解析では,

$$r_{12} = p_{12} + p_{13} \times p_{23}$$

と表現され, 「二つの変数の相関は互いの直接効果に, 背後の人口を介した二つの効果の積を加えたものになる」ことを覚えていた方がよい。

普通, 計量経済では独立変数や従属変数に per capita やこの対数変換値を用いた分析が広く行われてきた。実例として検討した Pettersson-Lidbom の論文は正にその典型で, 独立変数の council size の効果が有意な負の相関として検定されている。

これがどれ位の信頼度があるかはこの有意性を信じるわけにはいかないし、補正することもできない。なぜなら、回帰分析モデルが悪いのではなく、見せかけの効果を介入させる不適切な値、per capitaやこの対数変換値を用いていることが間違いなのである。従来、見せかけの相関は“時として起こる不可解な現象”として見過ごされてきて不問にされてきたが、本論で分析したようにper capitaやこの対数変換値を分析に用いることで、たえず起こっていることがわかる。このような不適切な値を統計的手法として用いることが問題なのである。Per capitaという値は、ある量の実測値を人口という実測値で割った値であり、これがどのような確率分布をするのかについては誰にもわからないという意味においても回帰分析に用いることは不適切である。実際にこの分布を見るといかにきかないかがわかる。

このような見せかけの相関が生じることを防ぐには、2次の偏相関係数を推定すれば良いことを示した。アメリカの刑犯罪各種間の因果関係がこれによって適切に検証できることは、2次偏相関の有効性と検出力の確かさを明らかにしている。

7. まとめ：「ベキ関数の法則」について

社会科学分野においては、特に、計量経済学を中心としてper capitaやper GDPを用いた回帰分析が日常的に行われ、回帰係数の統計的有意性や相関の度合いにもとづいて独立変数の効果が判断され、有意性があればあたかも因果関係があるかのように解釈されて政策論議にまで展開されるようなことがおもうように行われてきた。しかし、本論で考察してきたようにこのような回帰分析手法は必然的に見せかけの相関を生じさせており、統計学的手法としては不適切かつ間違った手法である。過去数十年、慣習的に踏襲されてきた回帰分析結果にもとづく計量経済学の多くの課題（例えばBarro 1997など）は再検討されなければならない。

社会科学分野における因果関係についてのあいまいな論議をさけるためにも経験則として以下の

3つの事項を「広義のベキ関数の法則」としてまとめおきたい。

- ①社会指標の多くは人口にたいしてベキ関数関係にあり、したがって、任意の社会指標はたがいにベキ関数で表現される。
- ②このような任意の社会指標間の回帰分析にper capitaやper GDPで表した社会指標値や、これを対数変換した値を用いると必然的にベキ関数関係を生じ、因果関係の有無に関わらず見せかけの相関が常に引き起こされる。
- ③このような見せかけの相関をさげ、直接的効果や因果関係を検証するためには、人口とGDPの効果をともに除いた二つの社会指標間の2次の偏相関係数($r_{12\cdot34}$)が統計的に有意となることを基準とするのがよい。これは、人口とGDPを変数として組み入れた4変数逐次モデルの構造方程式分析によって有意な直接的効果を検証する手法と等価である。

最終稿に目を通していただいた山田俊弘博士、および、英文校閲をお願いしたJoseph Lauerさんに感謝いたします。

参考文献

- 『現代アメリカデータ総覧』 柘風舎
『日本統計年鑑』 日本統計協会
日下部眞一(2002a) 回帰偏差値の考案とその効用：地域間格差を相対評価する偏差値。広島大学総合科学部紀要IV 理系編, 第28巻, 109-126。
日下部眞一(2002b) NPOの規模をはかる回帰偏差値。“NPO指数”の考案-NPO指数を通して見えてきた地域格差 - *The Nonprofit Review*, 2: 177-185。
日下部眞一(2011) 経済世界のアロメトリー“見せかけの相関はなぜ生じるか?” 環境科学研究(総合科学研究科紀要II) 第6巻, 1-7。
日下部眞一(2012) ソーシャル・キャピタル論の陥穽“ソーシャル・キャピタル指数は何を測っているのか?” 環境科学研究(総合科学研究科紀要II) 第7巻, 13-41。
ザイデル(2005)『数字で語るー社会統計学入門』 新曜

- 社
スネデカー, コ克蘭 (1972) 『統計的方法 原書6版』
岩波書店
- 藤田英典(2005) 「第4章 階層・階級」 宮島喬編『現代社会学 (改訂版)』(有斐閣) 93ページ.
- Agresti, A. (1997) *Statistical Methods for the Social Science*, 3rd ed. Prentice Hall, Inc. NJ. P.417.
- Arbuckle, J. L., & Wothke, W. (1999) *Amos 4.0 User's Guide*. Small-Waters, Chicago.
- Barro, R. J. (1997) *Determinants of Economic Growth. A Cross-Country Empirical Study*. MIT press.
- Bollen, K. A. (1989) *Structural Equations with Latent Variables*. John Wiley & Sons, Inc. New York.
- Duncan, O. D. (1975) *Introduction to Structural Equation Models*. Academic Press.
- Ezekiel, M. (1941) *Methods of Correlation Analysis*, 2nd ed. John Wiley & Sons, Inc.
- Kenny, D. A. (1979) *Correlation and Causality*. John Wiley & Sons, New York.
- Kerlinger, F. N. & Pedhazur, E. J. (1973) *Multiple Regression in Behavioral Research*. Holt, Rinehart and Winston, Inc. 1973.
- Ley, P. (1972) *Quantitative Aspect of Psychological assessment*. Banes & Noble, New York,
- Li, C. C. (1975) *Path Analysis – a primer*. The Boxwood Press, Pacific Grove, CA.
- Morgan, S. L. (2013) *Handbook of Causal Analysis for Social Research*. Springer.
- McNemar, Q. (1969) *Psychological Statistics*. 4th ed. John Wiley and Sons, NY.
- Pearl, J. (2009) *Causality*, 2nd ed. Cambridge University Press, UK.
- Pettersson-Lidbom, P. (2012) Does the size of the legislature affect the size of government? Evidence from two natural experiments. *Journal of Public Economics* 96, 269-278.
- Shipley, B. (2000) *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference*. Cambridge University Press, New York.
- Wright, S. (1921) Correlation and causation. *Jour. Agric. Res.* 20: 557-85.
- Wright, S. (1968) *Evolution and the Genetics of Populations. Volume 1 Genetics and biometric foundations*. University of Chicago Press, Chicago.
- Yule, U. (1922) *An Introduction to the Theory of Statistics*. Charles Griffin, Lond.
- Yule, U. & Kendall, M. G. (1950) *An Introduction to the Theory of Statistics*. 14th ed. Charles Griffin, Lond.

【校正時の注記】

「標準偏回帰係数 (パス係数) の絶対値が往々にして1を超えることがある」という事実は常識であるためか英文書籍ではAgresti (1997)だけにしか記載されていない。国内のパス解析についての解説では藤田英典(2005)のほか「パス係数は相関係数とおなじようにその絶対値は1以内である」というような間違った紹介が散見される。

付 録

(1) 1 次偏相関係数と部分相関係数について

二つの変数の関係に第三の変数が関与している場合がある。このような時に第三の変数の効果を取り除いた (partial out) 二つの変数の実質上の相関を検出することが大切である。このような時に用いられるのが偏相関 (partial correlation) の概念である。

Figure A1 を見ながら偏相関係数の考えと式の誘導について解説する。

2 変数 X_1 と X_2 があって、ともに X_3 の影響を受けている。その効果はパス係数、 p_{13} と p_{23} で表される。変数 X_1 と X_2 は X_3 以外に未定項 U_1 と U_2 の効果をそれぞれ受けている。この U_1 と U_2 の相関が X_3 の効果を除いた X_1 と X_2 の偏相関係数 r_{12*3} を意味する。 X_1 と U_2 の相関、および X_2 と U_1 の相関は part correlation, または semi-partial correlation と呼ばれている部分相関を意味する。本文でもふれたが、パス係数は標準偏回帰係数であり、相関係数とは異なり、その絶対値が 1 を越えることもある。

パス図から互いの相関関係で表現される式を導き出すには二つの方法がある。

①パス解析の基本規則から誘導する方法

【規則】パス図における 2 変数間の相関は、2 つの変数がたどるパスすべてのパス係数の積和になる。

ただし、二つ以上のパスを同時に受ける変数 ($\rightarrow X \leftarrow$), いわゆる合流点 (Collider, Pearl 2009) を途中に含むパスは除く。

この規則に従えば、図 A1 において、 $X_1 \rightarrow X_3 \rightarrow X_2$ のパスで $r_{13} \times r_{23} (= p_{13} \times p_{23})$, もう一つのパス、 $X_1 \rightarrow U_1 \rightarrow r_{12*3} \rightarrow U_2 \rightarrow X_2$ のパスで $u_1 \times u_2 \times r_{12*3}$ 。したがって、

$$r_{12} = r_{13} \times r_{23} + u_1 \times u_2 \times r_{12*3}$$

が、ただちに誘導される。

②構造方程式から相関関係を誘導する方法

パス図から X_1 と X_2 についての二つの同時方程式が得られる。

$$X_1 = p_{13} \times X_3 + u_1 \times U_1$$

$$X_2 = p_{23} \times X_3 + u_2 \times U_2$$

これらの式は操作変数法により互いの相関係数で表される連立方程式に変換される。かかわる変数、 X_1, X_2, X_3, U_1, U_2 を上記二つの方程式に掛けて期待値をとると、10 個の連立方程式が得られる。うち 6 つの方程式から、

$$r_{X_1, U_1} = u_1 = \sqrt{1 - r_{13}^2}$$

$$r_{X_2, U_2} = u_2 = \sqrt{1 - r_{23}^2}$$

$$r_{13} = p_{13}$$

$$r_{23} = p_{23}$$

Figure A1 r_{12*3} : a first-order partial correlation

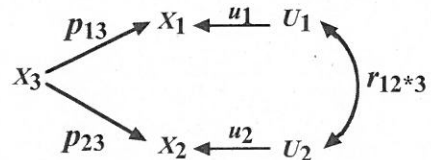


Figure A2 r_{12*34} : a second-order partial correlation defined as a partial correlation between $X_1 \times 3$ and $X_2 \times 3$

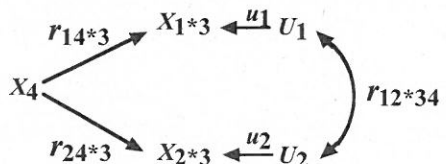
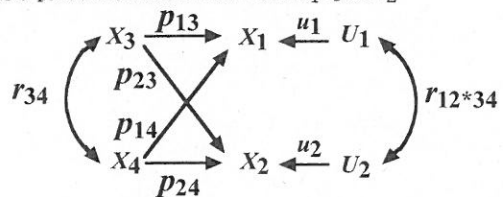


Figure A3 r_{12*34} : a second-order partial correlation defined as a partial correlation between X_1 and X_2



A formula for a second-order partial correlation, r_{12*34} between X_1 and X_2

$$r_{12*34} = \frac{r_{12*3} - r_{14*3} \times r_{24*3}}{\sqrt{1 - r_{14*3}^2} \times \sqrt{1 - r_{24*3}^2}}$$

Figure A1 expresses the first-order partial correlation. A2 is the second-order partial correlation expressed using the first-order correlations. A3 is the second-order partial correlation expressed using the zero-order correlations.

(図 A1 1 次偏相関係数を表す。A2 1 次偏相関係数で表される 2 次偏相関係数。A3 通常の相関係数で表される 2 次偏相関係数。)

が得られ、残りの二つから偏相関係数を持つ式

$$r_{12} = r_{13} \times r_{23} + u_1 \times u_2 \times r_{12*3}$$

が得られる。この式から偏相関係数を表す一般式が求まる。

$$r_{12*3} = \frac{r_{12} - r_{13} \times r_{23}}{\sqrt{1 - r_{13}^2} \times \sqrt{1 - r_{23}^2}}$$

残る二つの式は

$$r_{X_1 U_1} = u_1 \times r_{12*3}$$

$$r_{X_2 U_2} = u_2 \times r_{12*3}$$

であり、さきに述べた部分相関 (part correlation または semi-partial correlation) に相当する式である。

(2) 2次偏相関係数について

2次偏相関係数については、Yule (1922), Yule and Kendall (1950), Ezekiel (1941), McNemar (1969), Karlinger and Pedhazur (1973), Ley (1972) などが記載しているが式の誘導は Yule & Kendall (1950) に演繹的に記載されているのみで煩雑である。ここではすでに述べた1次偏相関係数にならってパス図A1-A3で説明する。本文図4(a)でも示した1次偏相関係数を表す図A1における2変数 X_1 と X_2 をそれぞれ X_{1*3} , X_{2*3} でおきかえて図A2のパス図がえられる。 X_{1*3} , X_{2*3} は partial score (Ley 1972) で、第3の変数の効果を除いた残差を意味し回帰偏差 (日下部2002) に相当している。図A1での U_1 , U_2 に相当する変数である。図A2は、これらの partial score から第4の変数 X_4 の効果を除くことを表現したパス図である。このパス図から1次偏相関係数と同様な計算を行えば2次偏相関係数 (r_{12*34}) を表す式が導かれる。もちろん変数の組み合わせによって他の5つの場合の2次偏相関係数 (r_{13*24} , r_{23*14} など) も考えることができる。

この2次偏相関係数の式は1次偏相関係数の関数で表現されているが、式を誘導して行くと通常の2変数相関係数で表現される式が導かれる。

$$r_{12*34} = \frac{r_{12*3} - r_{14*3} \times r_{24*3}}{\sqrt{1 - r_{14*3}^2} \times \sqrt{1 - r_{24*3}^2}}$$

$$= \frac{r_{12} - r_{12}r_{34}^2 - r_{13}r_{23} - r_{14}r_{24} + r_{14}r_{23}r_{34} + r_{13}r_{24}r_{34}}{\sqrt{(1 - r_{34}^2) - (r_{13}^2 + r_{14}^2 - 2r_{13}r_{14}r_{34})} \times \sqrt{(1 - r_{34}^2) - (r_{23}^2 + r_{24}^2 - 2r_{23}r_{24}r_{34})}}$$

Figure A2のパス図は、Figure A3で示した逐次4変数のパス図と同等で、Figure A3のパス図から同時方程式をたてて導いて行けば、煩雑ではあるが上記の2変数相関係数で表現される同じ式が導かれる。

同じ誘導を繰り返せば高次の偏相関係数が得られる。

例えば、3次の偏相関係数は、

$$r_{12*345} = \frac{r_{12*34} - r_{15*34} \times r_{25*34}}{\sqrt{1 - r_{15*34}^2} \times \sqrt{1 - r_{25*34}^2}}$$

n次の偏相関係数は、

$$r_{12*34\dots(n+2)} = \frac{r_{12*34\dots(n+1)} - r_{1(n+2)*34\dots(n+1)} \times r_{2(n+2)*3\dots(n+1)}}{\sqrt{1 - r_{1(n+2)*34\dots(n+1)}^2} \times \sqrt{1 - r_{2(n+2)*34\dots(n+1)}^2}}$$

で表わされる。