R^2に関するノート

福井義高
青山学院大学大学院国際マネジメント研究科
〒150-8366　東京都渋谷区渋谷４－４－２５
fukui@gsim.aoyama.ac.jp

2004 年 10 月
2007 年 5 月 29 日改訂

要旨

会計実証分析において多用される R^2 と $\overline{R}^2$ を，母集団における当てはまりの指標及びモデル選択基準としての観点から，詳しく検討する．

R^2に関するノート

These measures of goodness of fit have a fatal attraction.
J. S. Cramer

１．推測統計における R^2 と $\overline{R}^2$

標本から母集団の振る舞いを推測する推測統計において，R^2 の意味は限定される．ところが，会計に限らず実証分析において，「インサイダーの間では意味がないことが理解されているにもかかわらず，高い値は論文の筆者にとって自慢と満足の元となっている．」[1]

ここでは，「アウトサイダー」にもわかりやすいように，R^2 及びその「修正版」である $\overline{R}^2$ の意味するところを，母集団における当てはまりの指標及びモデル選択基準としての観点から，先行研究に依拠[2]して，詳しく検討する．式の展開は飛躍がないようにし，順を追って読み進められるよう，式番号を付けず，必要な場合は何度でも再掲する．

なお，ここではベイジアンの立場に触れず，頻度論の立場での議論に限定する．会計分野の実証研究がほとんどの場合，頻度論を前提にしているという実際的理由からであって，頻度論がベイジアンより優れていると筆者が考えているからではない．

２．当てはまりの指標としての R^2：独立変数が固定された場合

まず，R^2 の導出から始める．標本数 n，従属変数 $\dot{y}=\begin{pmatrix}\dot{y}_1\\ \vdots \\ \dot{y}_n\end{pmatrix}$，確率変数でない独立変数 $\dot{x}_i'=\begin{pmatrix}x_{i1} & \cdots & x_{ik}\end{pmatrix}$ は k 個からなり $\dot{X}=\begin{pmatrix}\dot{x}_1'\\ \vdots \\ \dot{x}_n'\end{pmatrix}$ とする．それぞれ，・がついているのは，今後の式の展開上，置き換えた変数を中心に議論するためである．

---

[1] Cramer (1987, p. 253).
[2] 第 2・4 節は Cramer (1987)，第 3 節は Anderson (2003)，第 5 節は Hansen (2007)，第 6 節は Goldberger (1991; 1998)，そして第 7 節は McGuirk and Driscoll (1995)に拠っている．

回帰式は，定数項α，係数$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$， $c_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$，誤差項$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$とすると，

$$\dot{y} = \begin{pmatrix} c_n & \dot{X} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \varepsilon$$

となり， $Z = \begin{pmatrix} c_n & X \end{pmatrix}$， $\gamma = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$と置けば,

$$\dot{y} = Z\gamma + \varepsilon$$

となる．

以下，誤差項に関して $i.i.d.$正規分布 $\varepsilon \sim N(0, \sigma^2)$を仮定する．したがって，教科書どおりの最小二乗法が適用でき， $g = \begin{pmatrix} a \\ b \end{pmatrix}$は,

$$\begin{aligned} Z'e &= Z'(\dot{y} - Zg) = Z'\dot{y} - Z'Zg = 0 \\ g &= (Z'Z)^{-1} Z'\dot{y} \\ &= (Z'Z)^{-1} Z'(Z\gamma + \varepsilon) = \gamma + (Z'Z)^{-1} Z'\varepsilon \end{aligned}$$

したがって，推定値gの期待値は$E(g) = \gamma$となり，分散は

$$\begin{aligned} Var(g) &= E\left[(g-\gamma)(g-\gamma)'\right] = (Z'Z)^{-1} Z'E(\varepsilon\varepsilon')X(Z'Z)^{-1} \\ &= \sigma^2 (Z'Z)^{-1} \end{aligned}$$

となる．

次に，今後の展開に便利なように変数を変換する．具体的には，変数を平均からの乖離に定義しなおす．そのため， $J = I_n - \frac{c_n c_n'}{n}$を用いて， $X = J\dot{X}, y = J\dot{y}$とする．したがって,

$$\dot{y} = Zg + e$$

は

$$\begin{aligned} y = J\dot{y} &= JZg + Je = J\begin{pmatrix} c_n & \dot{X} \end{pmatrix} g + e = \begin{pmatrix} 0 & X \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + e \\ &= Xb \end{aligned}$$

になる．

$$\begin{aligned} y'y &= (Xb + e)'(Xb + e) = b'X'Xb + e'e \\ \hat{y}'\hat{y} &= b'X'Xb \end{aligned}$$

なので，この回帰式により「説明」される従属変数の割合である R^2 は

$$R^2=\frac{\hat{y}'\hat{y}}{y'y}=1-\frac{e'e}{y'y}=\frac{b'X'Xb}{b'X'Xb+e'e}$$

となる.「説明」されない割合は

$$1-R^2=\frac{e'e}{b'X'Xb+e'e}$$

である.

次に X が確率変数でない場合の R^2 の漸近的性質を確認する．X が固定されているので，n 個からなる標本 X を p 回繰り返し抽出すると考え，$N = np\to\infty$ とすれば,

$$\frac{X_N'X_N}{N}=\frac{pX'X}{N}\to_p\frac{X'X}{n},\ \frac{e_N'e_N}{N}\to_p E(e_i^2)=\sigma^2$$

となる．さらに,

$$b\sim N\left(\beta,\sigma^2(X'X)^{-1}\right)$$

なので,

$$b\to_p\beta$$

一方，R^2 の真の値は

$$\rho^2=\frac{\beta'X'X\beta}{\beta'X'X\beta+n\sigma^2}=\frac{\beta'\dfrac{X'X}{n}\beta}{\beta'\dfrac{X'X}{n}\beta+\sigma^2}$$

である．結局,

$$R_N^2=\frac{b'\dfrac{X_N'X_N}{N}b}{b'\dfrac{X_N'X_N}{N}b+\dfrac{e_N'e_N}{N}}=\frac{b'\dfrac{pX'X}{N}b}{b'\dfrac{pX'X}{N}b+\dfrac{e_N'e_N}{N}}\to_p\frac{\beta'\dfrac{X'X}{n}\beta}{\beta'\dfrac{X'X}{n}\beta+\sigma^2}=\rho^2$$

なので，R^2 は真の値 ρ^2 の一致推定量になっている.

次に,

$$\hat{\lambda}=n\frac{\beta'X'X\beta}{\sigma^2}$$

とすると,

$$\rho^2=\frac{\hat{\lambda}}{\hat{\lambda}+n},\ \hat{\lambda}=\frac{n\rho^2}{1-\rho^2}$$

となっている. なお, ここでは X を固定して考えているので, $\hat{\lambda}$ は確率変数ではない.

さて，$n \times n$ 行列

$$M = I_n - Z(Z'Z)^{-1}Z'$$

は，

$$M = M' = M^p,\ MZ = 0$$

という性質を持っている．ゆえに，

$$e = \dot{y} - Zg = \dot{y} - Z(Z'Z)^{-1}Z\dot{y} = M\dot{y} = M(Z\gamma + \varepsilon) = M\varepsilon$$

であることがわかる．M は対称行列なので，

$$G'G = GG' = I \Rightarrow G' = G^{-1}$$

となる行列 G で対角化できる．さらに，

$$M = G\Lambda G^{-1} = G\Lambda G',$$
$$M^p = G\Lambda G' \cdots G\Lambda G' = G\Lambda^p G'$$

となり，

$$M = M^p$$

なので，

$$M^p = G\Lambda^p G' = G\Lambda G' = M$$

つまり，

$$\Lambda^p = \Lambda$$

となる．したがって，固有値は 0 か 1 に限る．M の階数が $n-k-1$ であることと正方行列の階数とトレースが等しいことから，固有値を適当に並べ換えて，

$$\Lambda = \begin{pmatrix} I_{n-k-1} & 0 \\ 0 & 0 \end{pmatrix}$$

とすることができる．

$$v = \frac{e'e}{\sigma^2}$$

と置くと，

$$v = \frac{e'e}{\sigma^2} = \frac{\dot{y}'M'M\dot{y}}{\sigma^2} = \frac{(Z\gamma + \varepsilon)'M'M(Z\gamma + \varepsilon)}{\sigma^2} = \frac{\varepsilon'M\varepsilon}{\sigma^2} = \frac{1}{\sigma^2}\ \varepsilon'G\Lambda G'\varepsilon$$

となる．

$$\frac{1}{\sigma}G'\varepsilon \sim N(0, I)$$

なので，結局，

$$v=\frac{e'e}{\sigma^2}=\frac{1}{\sigma^2}\ \varepsilon' G\Lambda G'\varepsilon \sim \chi^2(n-k-1)$$

となる．

なお，正則行列の性質として

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1}=\begin{pmatrix} E^{-1} & -E^{-1}BD^{-1} \\ -D^{-1}CE^{-1} & F^{-1} \end{pmatrix}$$

$$E^{-1}=(A-BD^{-1}C)^{-1},\ F^{-1}=(D-CA^{-1}B)^{-1}$$

が知られている（左右から掛けると単位行列になっている）．これを使うと，

$$(Z'Z)^{-1}=\left[(c_n \quad \dot{X})'(c_n \quad \dot{X})\right]^{-1}=\begin{pmatrix} n & c_n'\dot{X} \\ \dot{X}'c_n & \dot{X}'\dot{X} \end{pmatrix}^{-1}$$

$$=\begin{pmatrix} * & * \\ * & (\dot{X}'\dot{X}-\dfrac{\dot{X}'c_n c_n'\dot{X}}{n})^{-1} \end{pmatrix}=\begin{pmatrix} * & * \\ * & (\dot{X}'J\dot{X})^{-1} \end{pmatrix}=\begin{pmatrix} * & * \\ * & (X'X)^{-1} \end{pmatrix}$$

となり，

$$g\sim N\left(\gamma,\sigma^2(Z'Z)\right)$$

に適用すると，

$$b=\left(0 \quad I_k\right)g\sim N\left(\beta,\sigma^2(0 \quad I_k)(Z'Z)^{-1}(0 \quad I_k)'\right)=N\left(\beta,\sigma^2(X'X)^{-1}\right)$$

となる．

さらに，$\sigma^2(X'X)^{-1}$ は正定値行列なので，

$$\sigma^2(X'X)^{-1}=HH'$$

と分解できる．

次に，$d=H^{-1}b$ と置くと，

$$b=Hd,\ E\left(d\right)=H^{-1}\beta$$

なので，$H^{-1}\sigma^2(X'X)^{-1}H'^{-1}=I_k$ を使うと，

$$\begin{aligned} E\left(d-H^{-1}\beta\right)\left(d-H^{-1}\beta\right)' &= E\left[\left(H^{-1}b-H^{-1}\beta\right)\left(H^{-1}b-H^{-1}\beta\right)'\right] \\ &= E\left[H^{-1}\left(b-\beta\right)\left(b-\beta\right)' H'^{-1}\right] \\ &= H^{-1}\sigma^2(X'X)^{-1}H'^{-1}=I_k \end{aligned}$$

となる．つまり，

$$d=H^{-1}b\sim N\left(H^{-1}\beta,I_k\right)$$

なので，$\dfrac{X'X}{\sigma^2}=\left(HH'\right)^{-1}$ となることを利用すれば，

$$d'd=(H^{-1}b)'H^{-1}b=b'H'^{-1}H^{-1}b=b'(HH')^{-1}b$$
$$=\frac{b'X'Xb}{\sigma^2}$$

は，非心度

$$\hat{\lambda}=(H^{-1}\beta)'H^{-1}\beta=\frac{\beta'X'X\beta}{\sigma^2}$$

自由度 k の非心カイ二乗分布

$$u=\frac{b'X'Xb}{\sigma^2}\sim\chi^2\left(\hat{\lambda},k\right)$$

であることがわかる．

さて，カイ二乗分布

$$x\sim\chi^2(\omega,l),\quad y\sim\chi^2(m)$$

の比 $\dfrac{x}{y}$ の密度関数は，

$$f\left(\frac{x}{y}\right)=\sum_0^{\infty}\frac{e^{-\frac{\omega}{2}}\left(\frac{\omega}{2}\right)^j}{j!}\frac{1}{\mathrm{B}\left(\frac{l}{2}+j,\frac{m}{2}\right)}\left(\frac{x}{y}\right)^{\frac{l}{2}+j-1}\left(1+\frac{x}{y}\right)^{-\frac{l+m}{2}-j}$$

となることが知られている．

ところで，

$$\frac{R^2}{1-R^2}=\frac{b'X'Xb}{e'e}=\frac{\frac{b'X'Xb}{\sigma^2}}{\frac{e'e}{\sigma^2}}=\frac{u}{v}$$

は，

$$u\sim\chi(\hat{\lambda},k),\quad v\sim\chi(n-k-1)$$

の比となっている．したがって，

$$f(\frac{u}{v})=\sum_{j=0}^{\infty}\frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!}\frac{1}{B\left(\frac{k}{2}+j,\frac{n-k-1}{2}\right)}\left(\frac{u}{v}\right)^{\frac{k}{2}+j-1}\left(1+\frac{u}{v}\right)^{-\frac{n-1}{2}-j}$$

あるいは，

$$f(\frac{R^2}{1-R^2})=\sum_{j=0}^{\infty}\frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!}\frac{1}{B\left(\frac{k}{2}+j,\frac{n-k-1}{2}\right)}\left(\frac{R^2}{1-R^2}\right)^{\frac{k}{2}+j-1}\left(\frac{1}{1-R^2}\right)^{-\frac{n-1}{2}-j}$$

となる.

R^2 の密度関数は, $\dfrac{d\left(\frac{R^2}{1-R^2}\right)}{dR^2}=\left(1-R^2\right)^{-2}$ を利用すれば,

$$h(R^2\,|\,X)=\sum_{j=0}^{\infty}\frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!}\frac{1}{\mathrm{B}\left(\frac{k}{2}+j,\frac{n-k-1}{2}\right)}\left(\frac{R^2}{1-R^2}\right)^{\frac{k}{2}+j-1}\left(\frac{1}{1-R^2}\right)^{-\frac{n-1}{2}-j}\left(1-R^2\right)^{-2}$$

$$=\sum_{j=0}^{\infty}\frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!}\frac{1}{\mathrm{B}\left(\frac{k}{2}+j,\frac{n-k-1}{2}\right)}\left(R^2\right)^{\frac{k}{2}+j-1}\left(1-R^2\right)^{\frac{n-k-1}{2}-1}$$

となることがわかる. なお, ある X という条件のもとであることを明示するため, $h(R^2\,|\,X)$ とした.

密度関数が得られたので, 期待値を含めモーメントを一般的に導出する.

$$E(R^{2p}\,|\,X)=\int_0^1 R^{2p}\sum_{j=0}^{\infty}\frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!}\frac{1}{\mathrm{B}\left(\frac{k}{2}+j,\frac{n-k-1}{2}\right)}\left(R^2\right)^{\frac{k}{2}+j-1}\left(1-R^2\right)^{\frac{n-k-1}{2}-1}dr$$

まず, $\mathrm{B}(a,b)=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ なので,

$$=\int_0^1\sum_{j=0}^{\infty}\frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!}\frac{\Gamma\left(\frac{n-1}{2}+j\right)}{\Gamma\left(\frac{k}{2}+j\right)\Gamma\left(\frac{n-k-1}{2}\right)}\left(R^2\right)^{\frac{k}{2}+j+p-1}\left(1-R^2\right)^{\frac{n-k-1}{2}-1}dr$$

となり, 次に分子・分母に $\Gamma\left(\frac{k}{2}+j+p\right)\Gamma\left(\frac{n-1}{2}+j+p\right)$ を掛けて,

$$=\int_0^1 \sum_{j=0}^{\infty} \frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!} \frac{\Gamma\left(\frac{k}{2}+j+p\right)\Gamma\left(\frac{n-1}{2}+j\right)}{\Gamma\left(\frac{k}{2}+j\right)\Gamma\left(\frac{n-1}{2}+j+p\right)}$$

$$\times \frac{\Gamma\left(\frac{n-1}{2}+j+p\right)}{\Gamma\left(\frac{k}{2}+j+p\right)\Gamma\left(\frac{n-k-1}{2}\right)} \left(R^2\right)^{\frac{k}{2}+j+p-1} \left(1-R^2\right)^{\frac{n-k-1}{2}-1} dr$$

とする．

さらに，$\Gamma(a+1)=a\Gamma(a)$ を繰り返し使い，

$$=\sum_{j=0}^{\infty} \frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!} \frac{\left(\frac{k}{2}+j+p-1\right)\cdots\left(\frac{k}{2}+j\right)}{\left(\frac{n-1}{2}+j+p-1\right)\cdots\left(\frac{n-1}{2}+j\right)}$$

$$\times \int_0^1 \frac{1}{\mathrm{B}\left(\frac{k}{2}+j+p, \frac{n-k-1}{2}\right)} \left(R^2\right)^{\frac{k}{2}+j+p-1} \left(1-R^2\right)^{\frac{n-k-1}{2}-1} dr$$

を得る．最後に，ベータ関数の定義から $\int_0^1 \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)} dx = 1$ なので，

$$=\sum_{j=0}^{\infty} \frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!} \frac{\left(\frac{k}{2}+j+p-1\right)\cdots\left(\frac{k}{2}+j\right)}{\left(\frac{n-1}{2}+j+p-1\right)\cdots\left(\frac{n-1}{2}+j\right)}$$

を得る．したがって，

$$E\left(R^2 \mid X\right) = \sum_{j=0}^{\infty} \frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!} \frac{\frac{k}{2}+j}{\frac{n-1}{2}+j}$$

$$E\left((R^2)^2 \mid X\right) = \sum_{j=0}^{\infty} \frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!} \frac{\left(\frac{k}{2}+j+1\right)\left(\frac{k}{2}+j\right)}{\left(\frac{n-1}{2}+j+1\right)\left(\frac{n-1}{2}+j\right)}$$

が求まる．なお，

$$\rho^2 = \frac{\hat{\lambda}}{\hat{\lambda}+n},\quad \hat{\lambda} = \frac{n\rho^2}{1-\rho^2}$$

なので，

$$E\left(R^2 \mid X\right) = \sum_{j=0}^{\infty} \frac{e^{-\frac{n\rho^2}{2(1-\rho^2)}}\left(\frac{n\rho^2}{1-\rho^2}\right)^j}{j!} \frac{\frac{k}{2}+j}{\frac{n-1}{2}+j}$$

$$E\left((R^2)^2 \mid X\right) = \sum_{j=0}^{\infty} \frac{e^{-\frac{n\rho^2}{2(1-\rho^2)}}\left(\frac{n\rho^2}{1-\rho^2}\right)^j}{j!} \frac{\left(\frac{k}{2}+j+1\right)\left(\frac{k}{2}+j\right)}{\left(\frac{n-1}{2}+j+1\right)\left(\frac{n-1}{2}+j\right)}$$

とも書ける．

Cramer (1987, p. 259)は，解析的には求められないので，この式を用いて数値計算を行い, X が確率変数でない場合の R^2 の偏りと散らばりを報告している. それによると，独立変数が 2 個，真の R^2 が 0.667 の場合の平均，バイアス及び標準偏差は次のようになる．

標本数	平均	バイアス	標準偏差
5	0.840	+0.173	0.152
10	0.746	+0.079	0.128
20	0.705	+0.038	0.096
50	0.682	+0.015	0.062
100	0.674	+0.007	0.044
150	0.672	+0.005	0.036
200	0.670	+0.003	0.028

上方へのバイアスは標本数が 50 程度あれば，ほとんど消える一方，標準偏差は標本数が 200 になってもかなり残る．すなわち，標本数が少ないときに R^2 を使うことの問題点は，バイアスよりも，散らばりの大きさによる推定値の低い信頼性にあることがわかる．

３．当てはまりの指標としての R^2：独立変数が確率変数の場合

ここまでは，X が固定の場合を考えてきたけれども，ここでは，X が確率変数の場合を考える．まず，R^2 の漸近的性質を確認する．通常の仮定が全て満たされている場合を考えているので，

$$\frac{X'X}{n} \to_p E(x_i x_i') = Q,\ \frac{e'e}{n} \to_p E(e_i^2) = \sigma^2$$

$$b \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right) \Rightarrow b \to_p \beta$$

$$\rho^2 = \frac{\beta' Q\beta}{\beta' Q\beta + \sigma^2}$$

となる．したがって，

$$R^2 = \frac{b'X'Xb}{b'X'Xb + e'e} = \frac{b'\dfrac{X'X}{n}b}{b'\dfrac{X'X}{n}b + \dfrac{e'e}{n}} \to_p \frac{\beta' Q\beta}{\beta' Q\beta + \sigma^2} = \rho^2$$

となり，R^2 は真の値 ρ^2 の一致推定量になっている．

次に，$\lambda = n\dfrac{\beta' Q\beta}{\sigma^2}$ とすると，

$$\rho^2 = \frac{\lambda}{\lambda + n},\ \lambda = \frac{n\rho^2}{1-\rho^2}$$

という関係が導ける．ここでは X が確率変数なので，２節と異なり，

$$\hat{\lambda} = \frac{\beta' X'X\beta}{\sigma^2} = \frac{\beta'\left(\sum_{i=1}^n x_i x_i'\right)\beta}{\sigma^2} = \sum_{i=1}^n \left(\frac{\beta' x_i}{\sigma}\right)^2$$

が確率変数となる．

$$E\left(\frac{\beta' x_i}{\sigma}\right) = 0$$

なので，$\dfrac{\beta' x_i}{\sigma}$ の分散は，

$$\Omega = \frac{E\left(\beta' x_i x_i' \beta\right)}{\sigma^2} = \frac{\beta' Q\beta}{\sigma^2} = \frac{\lambda}{n} = \frac{\rho^2}{1-\rho^2}$$

となる．$\phi = \dfrac{\rho^2}{1-\rho^2}$ と置くと，

$$\varphi^{-\frac{1}{2}} \frac{\beta' x_i}{\sigma} \sim N(0,1)$$

なので,

$$\phi^{-1}\sum_{i=1}^{n}\left(\frac{\beta' x_i}{\sigma}\right)^2=\phi^{-1}\hat{\lambda}\sim\chi^2(n-1)$$

あるいは,

$$\hat{\lambda}\sim\phi\chi^2(n-1)$$

となる（x が平均値を引いて得た数値なので自由度は $n-1$ となる）.

カイ二乗分布の密度関数は

$$f(y)=\frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)}y^{\frac{m}{2}-1}e^{-\frac{y}{2}}$$

なので,

$$\int_0^\infty e^{-\hat{\lambda}}\left(\frac{\hat{\lambda}}{2}\right)^j f(u)du=\int_0^\infty e^{-\phi u}\left(\frac{\phi u}{2}\right)^j\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}u^{\frac{n-1}{2}-1}e^{-\frac{u}{2}}du$$

$$=\frac{\phi^j}{2^j}\int_0^\infty\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}u^{\frac{n-1}{2}+j-1}e^{-\frac{u+\phi u}{2}}du$$

分子と分母に $\Gamma\left(\frac{n-1}{2}+j\right)$ を掛けて,

$$=\phi^j\frac{\Gamma\left(\frac{n-1}{2}+j\right)}{\Gamma\left(\frac{n-1}{2}\right)}\int_0^\infty\frac{1}{2^{\frac{n-1}{2}+j}\Gamma\left(\frac{n-1}{2}+j\right)}u^{\frac{n-1}{2}+j-1}e^{-\frac{u+\phi u}{2}}du$$

$v=(1+\phi)u$ と置くと,

$$=\frac{\phi^j}{(1+\phi)^{\frac{n-1}{2}+j-1}}\frac{\Gamma\left(\frac{n-1}{2}+j\right)}{\Gamma\left(\frac{n-1}{2}\right)}\int_0^\infty\frac{1}{2^{\frac{n-1}{2}+j}\Gamma\left(\frac{n-1}{2}+j\right)}v^{\frac{n-1}{2}+j-1}e^{-\frac{v}{2}}\frac{1}{1+\phi}dv$$

$$=\frac{\phi^j}{(1+\phi)^{\frac{n-1}{2}+j}}\frac{\Gamma\left(\frac{n-1}{2}+j\right)}{\Gamma\left(\frac{n-1}{2}\right)}\int_0^\infty\frac{1}{2^{\frac{n-2j-1}{2}}\Gamma\left(\frac{n-2j-1}{2}\right)}v^{\frac{n-2j-1}{2}-1}e^{-\frac{v}{2}}dv$$

積分部分はカイ二乗の累積分布関数になっているので,

$$= \frac{\phi^j}{(1+\phi)^{\frac{n-1}{2}+j}} \frac{\Gamma\left(\frac{n-1}{2}+j\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

が得られる．この結果を２節で得た条件付密度関数

$$h(R^2 \mid X) = \sum_{j=0}^{\infty} \frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!} \frac{1}{B\left(\frac{k}{2}+j, \frac{n-k-1}{2}\right)} \left(R^2\right)^{\frac{k}{2}+j-1} \left(1-R^2\right)^{\frac{n-k-1}{2}-1}$$

に代入すると，

$$h(R^2) = \sum_{j=0}^{\infty} \frac{\phi^j}{(1+\phi)^{\frac{n-1}{2}+j}} \frac{\Gamma\left(\frac{n-1}{2}+j\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{j!B\left(\frac{k}{2}+j, \frac{n-k-1}{2}\right)} \cdot \left(R^2\right)^{\frac{k}{2}+j-1} \left(1-R^2\right)^{\frac{n-k-1}{2}-1}$$

$\phi = \frac{\rho^2}{1-\rho^2}$ と $\mathrm{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ から，

$$= \sum_{j=0}^{\infty} \frac{\left(\frac{\rho^2}{1-\rho^2}\right)^j}{\left(\frac{1}{1-\rho^2}\right)^{\frac{n-1}{2}+j}} \frac{\Gamma\left(\frac{n-1}{2}+j\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\Gamma\left(\frac{n-1}{2}+j\right)}{j!\Gamma\left(\frac{k}{2}+j\right)\Gamma\left(\frac{n-k-1}{2}\right)} \left(R^2\right)^{\frac{k}{2}+j-1} \left(1-R^2\right)^{\frac{n-k-1}{2}-1}$$

$$= \frac{\left(1-R^2\right)^{\frac{n-k-1}{2}-1} (1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-k-1}{2}\right)} \sum_{j=0}^{\infty} \frac{\left(\rho^2\right)^j \left(R^2\right)^{\frac{k}{2}+j-1} \Gamma^2\left(\frac{n-1}{2}+j\right)}{j!\Gamma\left(\frac{k}{2}+j\right)}$$

さらに，もう一度 $\mathrm{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ を使い，

$$= \frac{\left(1-R^2\right)^{\frac{n-k-1}{2}-1} (1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \sum_{j=0}^{\infty} \frac{\left(\rho^2\right)^j \left(R^2\right)^{\frac{k}{2}+j-1} \Gamma\left(\frac{n-1}{2}+j\right)}{j!\mathrm{B}\left(\frac{n-k-1}{2}, \frac{k}{2}+j\right)}$$

が得られた．

こうして得られた R^2 の密度関数を使って，R^2 のモーメントを求める．

$$E(R^{2p})=\int_o^1 \frac{\left(1-R^2\right)^{\frac{n-k-1}{2}-1}(1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-k-1}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\rho^2\right)^j\left(R^2\right)^{\frac{k}{2}+j-1}\Gamma^2\left(\frac{n-1}{2}+j\right)}{j!\Gamma\left(\frac{k}{2}+j\right)}R^{2p}dR^2$$

$$=\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-k-1}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\rho^2\right)^j\Gamma^2\left(\frac{n-1}{2}+j\right)}{j!\Gamma\left(\frac{k}{2}+j\right)}\int_0^1\left(1-R^2\right)^{\frac{n-k-1}{2}-1}\left(R^2\right)^{\frac{k}{2}+j+p-1}dR^2$$

ここで分子と分母に$\Gamma\left(\frac{k}{2}+j+p\right)\Gamma\left(\frac{n-1}{2}+j+p\right)$を掛けて，

$$=\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\rho^2\right)^j\Gamma^2\left(\frac{n-1}{2}+j\right)\Gamma\left(\frac{k}{2}+j+p\right)}{j!\Gamma\left(\frac{k}{2}+j\right)\Gamma\left(\frac{n-1}{2}+j+p\right)}$$

$$\times\int_0^1\frac{\Gamma\left(\frac{n-1}{2}+j+p\right)}{\Gamma\left(\frac{n-k-1}{2}\right)\Gamma\left(\frac{k}{2}+j+p\right)}\left(1-R^2\right)^{\frac{n-k-1}{2}-1}\left(R^2\right)^{\frac{k}{2}+j+p-1}dR^2$$

ここで， $\mathrm{B}(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ と $\int_0^1\frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)}dx=1$から，

$$=\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\rho^2\right)^j\Gamma^2\left(\frac{n-1}{2}+j\right)\Gamma\left(\frac{k}{2}+j+p\right)}{j!\Gamma\left(\frac{k}{2}+j\right)\Gamma\left(\frac{n-1}{2}+j+p\right)}$$

さらに， $\Gamma(a+1)=a\Gamma(a)$を繰り返し使い，

$$=\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\rho^2\right)^j\left(\frac{k}{2}+j+p-1\right)\cdots\left(\frac{k}{2}+j\right)\Gamma^2\left(\frac{n-1}{2}+j\right)\Gamma\left(\frac{k}{2}+j\right)}{j!\left(\frac{n-1}{2}+j+p-1\right)\cdots\left(\frac{n-1}{2}+j\right)\Gamma\left(\frac{k}{2}+j\right)\Gamma\left(\frac{n-1}{2}+j\right)}$$

$$=\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\rho^2\right)^j\left(\frac{k}{2}+j+p-1\right)\cdots\left(\frac{k}{2}+j\right)\Gamma\left(\frac{n-1}{2}+j\right)}{j!\left(\frac{n-1}{2}+j+p-1\right)\cdots\left(\frac{n-1}{2}+j\right)}$$

となる．こうして，Xが確率変数である場合のR^2のn次モーメントが求められた．したがって，

$$E\left(R^2\right)=\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\rho^2\right)^j\left(\frac{k}{2}+j\right)\Gamma\left(\frac{n-1}{2}+j\right)}{j!\left(\frac{n-1}{2}+j\right)}$$

$$E\left((R^2)^2\right)=\frac{(1-\rho^2)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\rho^2\right)^j\left(\frac{k}{2}+j+1\right)\left(\frac{k}{2}+j\right)\Gamma\left(\frac{n-1}{2}+j\right)}{j!\left(\frac{n-1}{2}+j+1\right)\left(\frac{n-1}{2}+j\right)}$$

となる．

４．当てはまりの指標としての $\overline{R}^2$

まず，２節で得た独立変数が固定された場合の期待値，

$$E(R^2\mid X)=\sum_{j=0}^{\infty}\frac{e^{-\frac{\hat{\lambda}}{2}}\left(\frac{\hat{\lambda}}{2}\right)^j}{j!}\frac{\frac{k}{2}+j}{\frac{n-1}{2}+j}$$

を，さらに以下のように展開し，

$$=e^{-\frac{\hat{\lambda}}{2}}\left[\frac{k}{2}\sum_{j=0}^{\infty}\frac{\left(\frac{\hat{\lambda}}{2}\right)^j}{j!\left(\frac{n-1}{2}+j\right)}+\sum_{j=1}^{\infty}\frac{\left(\frac{\hat{\lambda}}{2}\right)^j\cdot j}{j!\left(\frac{n-1}{2}+j\right)}\right]$$

$$=e^{-\frac{\hat{\lambda}}{2}}\left[\frac{k}{2}\sum_{j=0}^{\infty}\frac{\left(\frac{\hat{\lambda}}{2}\right)^j}{j!\left(\frac{n-1}{2}+j\right)}+\sum_{j=1}^{\infty}\frac{\left(\frac{\hat{\lambda}}{2}\right)^j}{\left(j-1\right)!\left(\frac{n-1}{2}+j\right)}\right]$$

$$=e^{-\frac{\hat{\lambda}}{2}}\left[\frac{k}{2}\sum_{j=0}^{\infty}\frac{\left(\frac{\hat{\lambda}}{2}\right)^j}{j!\left(\frac{n-1}{2}+j\right)}+\frac{\hat{\lambda}}{2}\sum_{j=0}^{\infty}\frac{\left(\frac{\hat{\lambda}}{2}\right)^j}{j!\left(\frac{n-1}{2}+j+1\right)}\right]$$

を得る．

ここからの展開は次の微分方程式の知識を利用する．

$$z\frac{d^2w}{dz^2}+\left(b-z\right)\frac{dw}{dz}-aw=0$$

は Kummer 方程式と呼ばれる微分方程式で，その一般解は，

$$M(a,b,z) = \sum_0^\infty \frac{a^j z^j}{b^j j!}$$

となる．ただし，$a_j = a(a+1)\cdots(a+j-1)$，$b_j = b(b+1)\cdots(b+j-1)$である．さらに，$b = a+1$ ならば，

$$M(a,z) = a\sum_0^\infty \frac{z^j}{(a+j)j!}$$

となる．ここで，$p(a,z) = \dfrac{M(a,z)}{a}$ と置き，

$$p(a,z) = \sum_0^\infty \frac{z^j}{(a+j)j!}$$

とすると，さらに，

$$p(a,z) = \int_0^1 e^{zt} t^{a-1} dt$$

と積分の形に変形できる．数学的帰納法で証明する．まず，$a = 1$ のとき，

$$p(1,z) = \sum_0^\infty \frac{z^j}{(1+j)j!} = 1 + \frac{z}{2!} + \frac{z^2}{3!} + \cdots = \frac{e^z}{z} - \frac{1}{z} = \int_0^1 e^{zt} t^{1-1} dt$$

となる．

$$p(k,z) = \sum_0^\infty \frac{z^j}{(k+j)j!} = \frac{1}{k} + \frac{z}{k+1} + \frac{z^2}{(k+2)2!} + \cdots = \int_0^1 e^{zt} t^{k-1} dt$$

が成り立つとすると，

$$\begin{aligned}
p(k+1,z) &= \sum_0^\infty \frac{z^j}{(k+1+j)j!} = \frac{1}{k+1} + \frac{z}{k+2} + \frac{z^2}{(k+3)2!} + \cdots \\
&= \frac{d\left(\dfrac{1}{k} + \dfrac{z}{k+1} + \dfrac{z^2}{(k+2)2!} + \dfrac{z^3}{(k+3)3!} + \cdots\right)}{dz} \\
&= \frac{d}{dz}\int_0^1 e^{zt} t^{k-1} dt = \int_0^1 \frac{d(e^{zt} t^{k-1})}{dz} dt = \int_0^1 e^{zt} t^{k+1-1} dt
\end{aligned}$$

となり，

$$p(a,z) = \int_0^1 e^{zt} t^{a-1} dt$$

であることが証明された．

次に，部分積分を行うと，

$$p(a,z)=\int_0^1 e^{zt}t^{a-1}dt=\left[\frac{e^{zt}t^a}{a}\right]_0^1-\int_0^1\frac{ze^{zt}t^a}{a}dt=\frac{e^z}{a}-\frac{z}{a}p(a,z+1)$$

なので,

$$p(a,z+1)=\frac{e^z-a\cdot p(a,z)}{z}$$

となる．

R^2の期待値をもう一度見ると,

$$\frac{\left(\frac{\hat{\lambda}}{2}\right)^j}{j!\left(\frac{n-1}{2}+j\right)}=p(\frac{n-1}{2},z),\ \frac{\left(\frac{\hat{\lambda}}{2}\right)^j}{j!\left(\frac{n-1}{2}+1+j\right)}=p(\frac{n-1}{2}+1,z)$$

なので,

$$E(R^2\mid X)=e^{-\frac{\hat{\lambda}}{2}}\left[\frac{k}{2}\sum_{j=0}^{\infty}\frac{\left(\frac{\hat{\lambda}}{2}\right)^j}{j!\left(\frac{n-1}{2}+j\right)}+\frac{\hat{\lambda}}{2}\sum_{j=0}^{\infty}\frac{\left(\frac{\hat{\lambda}}{2}\right)^j}{j!\left(\frac{n-1}{2}+j+1\right)}\right]$$

$$=e^{-\frac{\hat{\lambda}}{2}}\left[\frac{k}{2}p\left(\frac{n-1}{2},\frac{\hat{\lambda}}{2}\right)+\frac{\hat{\lambda}}{2}p\left(\frac{n-1}{2}+1,\frac{\hat{\lambda}}{2}\right)\right]$$

となる．さらに，$p(a,z+1)=\dfrac{e^z-a\cdot p(a,z)}{z}$なので,

$$=e^{-\frac{\hat{\lambda}}{2}}\left[\frac{k}{2}p\left(\frac{n-1}{2},\frac{\hat{\lambda}}{2}\right)+\frac{\hat{\lambda}}{2}\frac{e^{\frac{\hat{\lambda}}{2}}-\frac{n-1}{2}p\left(\frac{n-1}{2},\frac{\hat{\lambda}}{2}\right)}{\frac{\hat{\lambda}}{2}}\right]$$

整理して,

$$=e^{-\frac{\hat{\lambda}}{2}}\left[\frac{k}{2}p\left(\frac{n-1}{2},\frac{\hat{\lambda}}{2}\right)+e^{\frac{\hat{\lambda}}{2}}-\frac{n-1}{2}p\left(\frac{n-1}{2},\frac{\hat{\lambda}}{2}\right)\right]$$

$$=1-e^{-\frac{\hat{\lambda}}{2}}\frac{n-k-1}{2}p\left(\frac{n-1}{2},\frac{\hat{\lambda}}{2}\right)$$

が得られる．

次に，変数追加による上方バイアスを「修正」することで真の R^2 のより良い推定

値として利用されることが多い，自由度修正済 R^2 は，

$$\overline{R}^2 = 1 - \frac{\dfrac{e'e}{n-k-1}}{\dfrac{\dot{y}'\dot{y} - n\overline{y}}{n-1}} = \frac{\dfrac{e'e}{n-k-1}}{\dfrac{y'y}{n-1}} = 1 - \frac{n-1}{n-k-1}\left(1-R^2\right) = \frac{n-1}{n-k-1}R^2 - \frac{k}{n-k-1}$$

$$= R^2 - \frac{k}{n-k-1}\left(1-R^2\right)$$

なので，必ず

$$\overline{R}^2 < R^2$$

となる．

$$E(\overline{R}^2 \mid X) = \frac{n-1}{n-k-1}E(R^2 \mid X) - \frac{k}{n-k-1}$$

にさきほど導出した，

$$E(R^2 \mid X) == 1 - e^{-\frac{\hat{\lambda}}{2}}\frac{n-k-1}{2}p\left(\frac{n-1}{2}, \frac{\hat{\lambda}}{2}\right)$$

を代入すると，

$$E(\overline{R}^2 \mid W) = \frac{n-1}{n-k-1}\left[1 - e^{-\frac{\hat{\lambda}}{2}}\frac{n-k-1}{2}p\left(\frac{n-1}{2}, \frac{\hat{\lambda}}{2}\right)\right] - \frac{k}{n-k-1}$$

$$= 1 - e^{-\frac{\hat{\lambda}}{2}}\frac{n-1}{2}p\left(\frac{n-1}{2}, \frac{\hat{\lambda}}{2}\right)$$

$$= 1 - e^{-\frac{\hat{\lambda}}{2}}\frac{n-1}{2}\sum_{j=0}^{\infty}\frac{\left(\dfrac{\hat{\lambda}}{2}\right)^j}{j!\left(\dfrac{n-1}{2}+j\right)}$$

となる．

確かに, R^2 と異なり, $\overline{R}^2$ の期待値は k に依存しない. 第 2 節で引用した Cramer (1987) の数値計算例すなわち独立変数が 2 個で真の R^2 が 0.667 の場合，$\overline{R}^2$ の期待値を計算すると以下のようになり，標本数が少ない場合でもバイアスはほとんどない．

標本数	R^2平均	$\bar{R}^2$平均
5	0.840	0.680
10	0.746	0.673
20	0.705	0.670
50	0.682	0.668
100	0.674	0.667
150	0.672	0.668
200	0.670	0.667

しかし，第 2 節でも指摘したように，R^2の問題点はバイアスより信頼性の低さ（散らばりの大きさ）だとすれば，むしろ$\bar{R}^2$は問題を悪化させている．なぜなら，

$$E(\bar{R}^2 \mid X) = \frac{n-1}{n-k-1} E(R^2 \mid X) - \frac{k}{n-k-1}$$

なので，標準偏差は,

$$SD(\bar{R}^2 \mid X) = \frac{n-1}{n-k-1} SD(R^2 \mid X)$$

となり，$\frac{n-1}{n-k-1} > 1$なので，必ず,

$$SD(\bar{R}^2 \mid X) > SD(R^2 \mid X)$$

となる．つまり，信頼性はさらに低下するのである．

なお，この Theil (1961)に帰せられることの多い$\bar{R}^2$は，バイアスの小さい推定のための A-Transformation として、Fisher (1924)により最初に導入された．ただし，バイアスは小さいけれども不偏推定量ではない．

５．モデル選択基準としての R^2 と $\bar{R}^2$

Theil (1961)以来, 実証分析における$\bar{R}^2$は, 母集団での当てはまりの指標ではなく, モデル選択の基準として使われてきた．R^2は定義式,

$$R^2 = 1 - \frac{e'e}{y'y} = 1 - \frac{e'e}{\dot{y}'\dot{y} - n\bar{y}}$$

を見ればわかるように，$y'y = \dot{y}'\dot{y} - n\bar{y}$が一定なので，$e'e$が小さいほど R^2が大きくなることがわかる．それゆえ，独立変数を追加すれば，$e'e$は同じままか小さくなるので，R^2の大小を基準にすると，必ず独立変数の多いモデルを選択することになる．そ

こで，このモデル選択基準としての「欠陥」を解消すべく提唱されたのが$\bar{R}^2$であった．独立変数追加による推定の際の自由度減少を考慮したとして，しばしば自由度修正済R^2と呼ばれる．

$$\bar{R}^2 = 1 - \frac{\dfrac{e'e}{n-k-1}}{\dfrac{\dot{y}'\dot{y} - n\bar{y}}{n-1}} = 1 - \frac{\dfrac{e'e}{n-k-1}}{\dfrac{y'y}{n-1}}$$

ここでは，前節までと異なり，定数項は独立変数$x_{i0} = 1$の係数と考えて，$\dot{k} = k+1$と置き換え，

$$\bar{R}^2 = 1 - \frac{\dfrac{e'e}{n-\dot{k}}}{\dfrac{\dot{y}'\dot{y} - n\bar{y}}{n-1}} = 1 - \frac{\dfrac{e'e}{n-\dot{k}}}{\dfrac{y'y}{n-1}}$$

の形を用いる．

定義式からわかるように，$\bar{R}^2$の大小は$\dfrac{e'e}{n-\dot{k}}$の大小で決まる．対数は順序を維持するので，$\bar{R}^2$を基準に用いるということは，$\log\left(\dfrac{e'e}{n-\dot{k}}\right)$が小さいモデルを選択するといっても良い．これを展開し，xが小さい場合の近似式$\log(1+x) \cong x$を用いると，

$$\begin{aligned}
\log\left(\frac{e'e}{n-\dot{k}}\right) &= \log\left(\frac{e'e}{n}\frac{n}{n-\dot{k}}\right) = \log(\hat{\sigma}^2) + \log\left(\frac{n}{n-\dot{k}}\right) \\
&= \log(\hat{\sigma}^2) + \log\left(1 + \frac{\dot{k}}{n-\dot{k}}\right) \\
&\cong \log(\hat{\sigma}^2) + \frac{\dot{k}}{n-\dot{k}} \\
&\cong \log(\hat{\sigma}^2) + \frac{\dot{k}}{n}
\end{aligned}$$

となる．これは変数を追加することで小さくなる（か同じままの）$\log(\hat{\sigma}^2)$に$\dfrac{\dot{k}}{n}$のペナルティーを加える形になっている．したがって，独立変数を増やしても，R^2と異なり$\bar{R}^2$が大きくなるとは限らない．

さて，$\gamma \;= g$での対数尤度は，

$$l_n(g, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}e'e$$

なので，

$$\frac{\partial l_n}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\left(\hat{\sigma}^2\right)^2} e'e = 0$$

から$\hat{\sigma}^2 = \dfrac{e'e}{n}$，したがって，

$$\hat{l}_n = -\frac{n}{2}\log\left(2\pi\hat{\sigma}^2\right) - \frac{1}{2\hat{\sigma}^2} e'e = -\frac{n}{2}\log\left(\hat{\sigma}^2\right) - \frac{n}{2}\log(2\pi) - \frac{n}{2}$$

となる．

変数$\dot{k}_1$個のモデルと，新たにk_p個の変数を加えて$\dot{k}_2 = \dot{k}_1 + k_p$個としたモデルの対数尤度比（の 2 倍）は，

$$\begin{aligned} LR_n &= 2\left(\hat{l}_n^2 - \hat{l}_n^1\right) = n\left(\log\hat{\sigma}_1^2 - \log\hat{\sigma}_2^2\right) = n \cdot \log\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = n \cdot \log\left(1 + \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} - 1\right) \\ &\cong n\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} - 1\right) = n\left(\frac{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}{\hat{\sigma}_2^2}\right) \end{aligned}$$

で，$1 - R^2 = \dfrac{e'e}{y'y} = \dfrac{n\hat{\sigma}^2}{y'y}$を使うと，

$$\begin{aligned} &= 2\left(\hat{l}_n^2 - \hat{l}_n^1\right) = n\left(\log\hat{\sigma}_1^2 - \log\hat{\sigma}_2^2\right) = n \cdot \log\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = n \cdot \log\left(1 + \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} - 1\right) \\ &\cong n\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} - 1\right) = n\left(\frac{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}{\hat{\sigma}_2^2}\right) \\ &= n\frac{1 - R_1^2 - (1 - R_2^2)}{1 - R_2^2} = \frac{n\left(R_2^2 - R_1^2\right)}{1 - R_2^2} \end{aligned}$$

となる．これは，$\dot{k}_2 = \dot{k}_1 + k_p$個のモデルに$k_p$の変数の係数がゼロという制約を加えた場合の検定となっており，（制約が正しい場合）漸近的には，

$$W_n = \frac{n\left(R_2^2 - R_1^2\right)}{1 - R_2^2} \to_d \quad \chi^2\left(\dot{k}_2 - \dot{k}_1\right)$$

と，自由度$\dot{k}_2 - \dot{k}_1$のカイ二乗分布に従う．

以上の準備をもとに，$\overline{R}^2$に基づくモデル選択の漸近的性質を検討する．真のモデルがM_1とすると，$\overline{R}^2$に基づくモデル選択が真のモデルを選択する確率は，

$$
\begin{aligned}
P\left(\hat{M}=M_1 \mid M_1\right) &= P\left(\overline{R}_1^2 > \overline{R}_2^2 \mid M_1\right) \\
&\cong P\left(\log\left(\hat{\sigma}_1^2\right) + \frac{\dot{k}_1}{n} < \log\left(\hat{\sigma}_2^2\right) + \frac{\dot{k}_2}{n} \mid M_1 \right) \\
&= P\left(n\left[\log\left(\hat{\sigma}_1^2\right) - \log\left(\hat{\sigma}_2^2\right)\right] < \dot{k}_2 - \dot{k}_1 \mid M_1\right) = P\left(LR_n < \dot{k}_2 - \dot{k}_1 \mid M_1\right) \\
&\to P\left(\chi^2\left(\dot{k}_2 - \dot{k}_1\right) < \dot{k}_2 - \dot{k}_1\right) < 1
\end{aligned}
$$

となり，標本数が無限にあっても，真のモデルを選択する確率は 100%とならず，モデル選択基準として一致性を持たない．直感的にいうと，ペナルティーが軽すぎて，変数の多いモデルを誤って選択する傾向がある．

では，他のモデル選択基準はどうであろうか．とくに時系列分析で用いられることの多いモデル選択基準に情報量規準 AIC がある．AIC は，

$$
AIC = -2(\hat{l}_n - \dot{k})
$$

の大小でモデルを選択する．ところで，

$$
\hat{l}_n = -\frac{n}{2}\log\left(2\pi\hat{\sigma}^2\right) - \frac{1}{2\hat{\sigma}^2}e'e = -\frac{n}{2}\log\left(\hat{\sigma}^2\right) - \frac{n}{2}\log\left(2\pi\right) - \frac{n}{2}
$$

を代入すると，

$$
\begin{aligned}
AIC = -2(\hat{l}_n - \dot{k}) &= -2\left[-\frac{n}{2}\log\left(\hat{\sigma}^2\right) - \frac{n}{2}\log\left(2\pi\right) - \frac{n}{2} - \dot{k} \right] \\
&= n\log\left(\hat{\sigma}^2\right) + n\log\left(2\pi\right) + \frac{n}{2} + 2\dot{k} \\
&= n\left[\log\left(\hat{\sigma}^2\right) + 2\frac{\dot{k}}{n} + \log\left(2\pi\right) + \frac{1}{2} \right]
\end{aligned}
$$

となる．つまり，AIC は，

$$
\log\left(\hat{\sigma}^2\right) + 2\frac{\dot{k}}{n}
$$

の大小によってモデル選択を行うことになる．$\overline{R}^2$ との違いは，ペナルティーの重さが 2 倍になっているところである．したがって，$\overline{R}^2$ 同様，真のモデルを選択する確率は，

$$
\begin{aligned}
P\left(\hat{M}=M_1 \mid M_1\right) &= P\left(AIC_1 < AIC_2 \mid M_1\right) = P\left(LR_n < 2\left(\dot{k}_2 - \dot{k}_1\right)\right) \\
&\to P\left(\chi^2\left(\dot{k}_2 - \dot{k}_1\right) < 2\left(\dot{k}_2 - \dot{k}_1\right)\right) < 1
\end{aligned}
$$

となり，モデル選択基準として一致性を持たない．$\overline{R}^2$ よりも改善したものの，いまだペナルティーが軽く，変数の多いモデルを選択する可能性が残る．

一方，AIC に類似した選択基準である BIC は，

$$BIC = -2\hat{l}_n + \log(n)\cdot \dot{k}$$

の大小でモデルを選択する．AIC と同様の代入を行い整理すると，BIC は,

$$\log\left(\hat{\sigma}^2\right) + \log(n)\cdot \frac{\dot{k}}{n}$$

の大小によってモデル選択を行うことになる．BIC は，$\overline{R}^2$ や AIC と異なり，標本数が多くなると，変数追加のペナルティーが大きくなる．実際，真のモデルを選択する確率は,

$$P\left(\hat{M} = M_1 \mid M_1\right) = P\left(BIC_1 < BIC_2 \mid M_1\right) = P\left(\frac{LR_n}{\log(n)} < \left(\dot{k}_2 - \dot{k}_1\right)\right)$$
$$\rightarrow P\left(0 < 2\left(\dot{k}_2 - \dot{k}_1\right)\right) = 1$$

となり, モデル選択基準として一致性を持つ. つまり, 標本数が多くなればなるほど, 選択されるモデルが真のモデルに近づく.

以上，三つのモデル選択基準は，変数を増やしたときのペナルティーの重さをどうするかに違いがあることがわかった．モデル選択が一致性を持つには，標本数増にともないペナルティーを大きくすることがカギである.

６．異なる標本間の比較

実証分析でしばしば忘れられているかに見えるのが，前節のモデル選択の議論は同じ従属変数を使った場合のみ妥当するということである．つまり，同じ y に対して，多くの説明変数の候補 $\{x_1, x_2, \ldots\}$ からどれをモデルに取り込むべきかという議論にしか使えない．ここで同じ y を使うという条件は変数変換をしないことも含む．たとえば, 従属変数に y を使ったモデルと $\log y$ を使ったモデルの選択には前節の議論は妥当しない.

さらに，同じ従属変数（と説明変数）を用いても，異なる標本間のモデル比較に R^2 を用いること[3]を正当化することは困難である．第 3 節冒頭で述べたように，X が確率変数かつ通常の仮定が全て満たされている場合,

$$\frac{X'X}{n} \rightarrow_p E(x_i x_i') = Q,\ \frac{e'e}{n} \rightarrow_p E(e_i^2) = \sigma^2$$

$$b \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right) \Rightarrow b \rightarrow_p \beta$$

[3] 当節の議論は R^2 と $\overline{R}^2$ 両者に共通なので，R^2 に代表させて検討する.

$$\rho^2 = \frac{\beta' Q \beta}{\beta' Q \beta + \sigma^2}$$

で，

$$R^2 = \frac{\hat{y}'\hat{y}}{y'y} = \frac{b'X'Xb}{b'X'Xb + e'e} = \frac{b'\dfrac{X'X}{n}b}{b'\dfrac{X'X}{n}b + \dfrac{e'e}{n}} \to_p \frac{\beta' Q \beta}{\beta' Q \beta + \sigma^2} = \rho^2$$

となる.

ここで, 議論の見通しをよくするために, Xが一つの変数xからなる場合を考える[4]. すなわち,

$$\frac{\sum x_i^2}{n} \to_p E(x_i^2) = q,\ \frac{\sum e_i^2}{n} = \hat{\sigma}^2 \to_p E(e_i^2) = \sigma^2$$

$$b \sim N\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right) \Rightarrow b \to_p \beta$$

$$\rho^2 = \frac{\beta^2 q}{\beta^2 q + \sigma^2}$$

で，

$$R^2 = \frac{b^2 \sum x_i^2}{b^2 \sum x_i^2 + \sum e_i^2} = \frac{b^2 \dfrac{\sum x_i^2}{n}}{b^2 \dfrac{\sum x_i^2}{n} + \hat{\sigma}^2} \to_p \frac{\beta^2 q}{\beta^2 q + \sigma^2} = \rho^2$$

となる場合である.

これを見ると，仮にモデルの誤差項 e が全く同じであっても，説明変数 x の散らばりが大きい場合，R^2 が大きくなる．例えば，景気変動が大きかった時期と安定していた時期を別々に，その時点の経済成長率を説明変数，失業率を従属変数として回帰分析した場合，誤差項 e のみならず係数 b が全く同じでも，景気変動つまり説明変数である経済成長率の変動が大きかった時期のモデルの R^2 が大きくなる．さらに，標本間の係数 b が異なる場合も同様に，係数 b（の絶対値）が大きい標本の R^2 が大きくなる．説明変数の散らばりやモデルの係数の大小で左右される指標を「当てはまりのよさ」に用いることは望ましいとはいえない.

さて，それでは異なる標本間でモデル比較をする場合，何を使うべきか．まず，考

[4]事前に平均ゼロに基準化してあると仮定する．したがって，定数項はない.

えられるのは，標本分散$\hat{\sigma}^2$（あるいは標準偏差$\hat{\sigma}$）であり，会計やファイナンスの実証研究でしばしば登場するプライシング・エラーとはモデルの標準偏差のことである．しかし，リターンあるいは価格水準の異なる標本間では単純な比較は好ましくない．たとえば，ある標本の平均リターンが 10%，他の標本は 5%の場合，前者を使ったモデルの標準偏差が 1.5%で，後者のそれが 1%だった場合，後者のモデルの方が当てはまりがよいとは必ずしもいえないだろう．こうした標本間の相違をどう数量化するか，有力な解決策が Gu (2007)により提示されている．

なお，ここでの議論はリターンあるいはレベル（価格）データのいずれを使うべきかという，value relevance研究に関してしばしば取り上げられる，いわゆるスケールの問題とは別である[5]．ここで取り上げた論点は仮にスケールの問題がなくても存在する，より本質的なものと見なすこともできる．

７．トレンドの存在

会計情報に限らず経済時系列データにはトレンド[6]が存在するのが普通であり，クロスセクションに比べ時系列データを用いた回帰分析でR^2が大きくなることは良く知られている[7]．しかし，このことは後者の方が前者より「当てはまりがよい」がいいから生じているのではない. 説明変数と従属変数に含まれるトレンドがR^2をかさ上げするのである．しかも，トレンドの存在を考慮せずに回帰した場合つまりモデル選択そのものが間違っている場合だけではなく，トレンドを考慮に入れた正しいモデルの場合でも，トレンドを考慮に入れる方法によってR^2が異なるのである．

真のモデルが

$$y_{NT} = X\beta + \varepsilon$$

で，説明変数Xがトレンド項tとその他X_{NT}からなるとすると，

$$X = (X_{NT}, t), \beta = \begin{pmatrix} \beta_{NT} \\ \beta_T \end{pmatrix}$$

$$t = \begin{pmatrix} 1 \\ \vdots \\ T \end{pmatrix}$$

5 スケールの問題については Gu (2005)参照.

6 ここでは，確定的トレンドのみ扱う.

7 前節同様，当節の議論もR^2と$\overline{R}^2$両者に共通なので，R^2に代表させて検討する.

と置けば，

$$y_{NT}=X\beta+\varepsilon=X_{NT}\beta_{NT}+t\beta_T+\varepsilon$$

となる．

ここで，トレンドを含んだ説明変数に基づく対称行列

$$M_X=I-X(X'X)^{-1}X'$$

を定義すると，

$$M_X\left(X_{NT},t\right)=M_XX=\left(I-X(X'X)^{-1}X'\right)X=0$$

を得る．すなわち，

$$M_XX_{NT}=0,\ M_Xt=0$$

である．M_Xを用いて最小自乗法による誤差項を表せば

$$e=y_{NT}-Xb=y_{NT}-X(X'X)^{-1}Xy_{NT}=M_Xy_{NT}=M_X(X\beta+\varepsilon)=M_X\varepsilon$$

となる．さらに，

$$M_T=I-t(t't)^{-1}t'$$

と定義すると，

$$M_Tt=0$$

で，

$$M_TM_X=\left(I-t(t't)^{-1}t'\right)M_X=M_X-t(t't)^{-1}t'M_X=M_X$$

を得る．

最小自乗法に基づく推計モデルは

$$y_{NT}=Xb+e=X_{NT}b_{NT}+tb_T+e$$

なので，

$$M_Te=M_TM_X\varepsilon=M_X\varepsilon=e$$

に注意し，両辺にM_Tを掛けて，

$$\begin{aligned}M_Ty_{NT}&=M_TX_{NT}b_{NT}+M_Ttb_T+M_Te\\&=M_TX_{NT}b_{NT}+e\end{aligned}$$

を得る．さらにX_{NT}を掛けると，

$$(X_{NT},t)'e=X'e=X'\left(y_{NT}-Xb\right)=X'\left(y_{NT}-X(X'X)^{-1}X'y_{NT}\right)=0$$

なので，

$$X'_{NT}M_T y_{NT} = X'_{NT}M_T X_{NT}b_{NT} + X'_{NT}e$$
$$= X'_{NT}M_T X_{NT}b_{NT}$$

となる．したがって，

$$b_{NT} = (X'_{NT}M_T X_{NT})^{-1}X'_{NT}M_T y_{NT}$$

を得る．

$$M'_T M_T = M_T^2 = \left(I - t(t't)^{-1}t'\right)\left(I - t(t't)^{-1}t'\right) = M_T$$

なので，新たに説明変数と従属変数

$$X_{DT} = M_T X_{NT}, \; y_{DT} = M_T y_{NT}$$

を定義すると，

$$b_{NT} = (X'_{DT}X_{DT})^{-1}X'_{DT}y_{DT}$$

となり，これは

$$y_{DT} = X_{DT}b + e$$

の最小自乗推計値となっている．

ところで，M_Tによる変換で変数はどうなっているのだろうか．まず，従属変数のトレンドを除去してみる[8]．トレンドの係数は

$$y_{NT} = b_{YT}t + e_{DT}$$

より，

$$b_{YT} = (t't)^{-1}t'y$$

なので，

$$e_{DT} = y_{NT} - (t't)^{-1}t'yt$$

を得る．さて，

$$y_{DT} = M_T y_{NT} = \left(I - t(t't)^{-1}t'\right)y_{NT} = y_{NT} - t(t't)^{-1}t'y_{NT}$$

で，

$$(t't)^{-1}t'y_{NT}$$

がスカラーであることに注意すると，

$$y_{DT} = y - t(t't)^{-1}t'y_{NT} = y - (t't)^{-1}t'y_{NT}t$$

となる．すなわち，M_Tによる変換は変数からトレンドを除去する操作なのである．ここでの操作は説明変数についても同じことである．つまり，トレンド項を入れたモデ

[8] 事前に平均ゼロに基準化してあると仮定する．したがって，定数項はない．

ルの係数推計値と誤差項は，あらかじめ説明・従属変数のトレンドを除去してから回帰したモデルと全く同じである[9].

にもかかわらず,

$$R^2 = \frac{\hat{y}'\hat{y}}{y'y} = 1 - \frac{e'e}{y'y}$$

なので，誤差項も係数も両モデルは同じ，つまり本質的に同じモデルなのに，

$$y'_{NT} y_{NT} > y'_{DT} y_{DT}$$

なため，トレンド除去モデルの R^2 が常に（大幅に）小さくなる．だからといって，トレンド除去モデルよりトレンドを含むモデルの方が「当てはまりがよい」とはいえないだろう．

８．おわりに

以上，母集団における真のあてはまりの推定値及びモデル選択基準としての R^2 と $\overline{R}^2$ について解説した．結論を一言でいえば，R^2 あるいは $\overline{R}^2$ の大小で，モデルの優劣を議論することは危険であるということである．今後，実証分析を進める際に，このノートが多少とも参考になれば幸いである．なお，ここで触れなかった密度関数の詳細については竹内 (1991)を，モデル選択を含む推測統計の基礎については細谷(2002)を参照してほしい．

[9] ここでの議論はトレンドにとどまらず，説明変数を二群に分け，段階的に回帰させる場合一般に適用できる．いわゆる Frisch-Waugh 定理 (1933)である．

参考文献

竹内彰通（1991）『現代数理統計学』創文社.

細谷雄三（2002）『統計的証拠とその解釈（増補版）』牧野書店.

Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*, Third Edition. Hoboken, U.S.A.: Wiley.

Cramer, J. S. 1987. Mean and Variance of R^2 in Small and Moderate Samples. *Journal of Econometrics* 35: 253-266.

Fisher, R. A. 1924. The Influence of Rainfall on the Yield of Wheat at Rothamsted. *Philosophical Transactions of the Royal Society of London* Series B 213: 89-142.

Frisch, R., and F. Waugh. 1933. Partial Time Regressions as Compared with Individual Trends. *Econometrica* 1 (4): 387-401.

Goldberger, A. S. 1991. *A Course in Econometrics*. Cambridge, U.S.A.: Harvard University Press.

Goldberger, A. S. 1998. *Introductory Econometrics*. Cambridge, U.S.A.: Harvard University Press.

Gu, Z. 2005. Scale Factor, R^2, and the Choice of Levels versus Returns Models. *Journal of Accounting, Auditing and Finance* 20 (1): 71-91.

Gu, Z. 2007. Across-Sample Incomparability of R^2s and Additional Evidence on Value Relevance Changes over Time. *Journal of Business Finance and Accounting* (forthcoming).

Hansen, B. E. 2007. Econometrics. Lecture Notes, Department of Economics, University of Wisconsin.

McGuirk, A. M., and P. Driscoll. 1995. The Hot Air in R^2 and Consistent Measures of Explained Variation. *American Journal of Agricultural Economics* 77 (2): 319-328.

Theil, H. 1961 (1958). *Economic Forecasting and Policy*, Second Edition. Amsterdam, Netherlands: North-Holland.