測定誤差と回帰係数の偏り

洲　浜　源　一

目　次

1．序

この報告は経済変数に含まれる測定誤差あるいは変数誤差が，回帰係数の最小２乗推定量に与える影響のうち特に「偏り」について検討する。大多数の経済変数はその測定時においてしばしば「四捨五入」等によって切り上げあるいは切り捨てがおこなわれる。その結果それぞれの「真の値」から乖離した変数となる。このためこれらの変数を用いた各種の経済モデルの推定結果には少なからず現実からの「ズレ」が発生する。例えば消費モデルの推定において，「四捨五入」等がおこなわれた可処分所得値をその説明変数として用いた消費モデルでは，推定された消費性向等がその「真の係数」から「ズレ」るおそれがある。

ところでこの種の測定誤差に基づく回帰係数の「偏り」については，これまで大標本を前提とした漸近的分析結果のみで，小標本を前提とし分析結果は殆ど知られていない。[1] そこでこの報告は，(1)単一の説明変数から構成される単純回帰モデルに限定し，説明変数に含まれる測定誤差が回帰係数の推定結果に与える「偏り」を小標本の下で定式化する。(2)測定誤差は互いに独立な正規確率

変数と仮定するが，その「分散」は変数測定時における「四捨五入」等に関連付けて分析する。(3)数値計算によって回帰係数の「偏り」係数を計算する。その結果はモンテ・カルロ実験によって確かめられる。(4)測定誤差が回帰係数を過小推定するという大標本に関する帰結は，小標本のもとでもほぼ成立する[2)]。

2．モ　デ　ル

次のような回帰モデルに限定する。

$$Y_i = \alpha + \beta X_i^* \qquad (\mathrm{i}=1,\ 2,\ \ldots,\ \mathrm{n}) \tag{2.1}$$

$$X_i = X_i^* + v_i \qquad (\mathrm{i}=1,\ 2,\ \ldots,\ \mathrm{n}) \tag{2.2}$$

ここで Y_i はモデルの被説明変数，X_i は実際に測定されるモデルの説明変数とする。上記（2.2）式は説明変数 X_i がその「真値」である X_i^* と測定誤差 v_i との二つの項の和で表されることを示す。このモデルはかなり単純化されている。まず（2.1）式は通常含まれる「方程式の誤差」を含んでいない。また被説明変数 Y_i には測定誤差が含まれていない。しかしこれらの誤差は回帰係数の推定にあたってその「偏り」の原因とはならない。ゆえにあとの帰結はこれらの誤差の存在に関係なく成立する。次に測定誤差を表す v_i は互いに独立な正規確率変数で平均 0，また以下の数式展開を簡明にするためにさしあたってその分散 σ^2 を 1 とする。さらに誤差 v_i は「真値」X_i^* との間でも独立と仮定する。したがって X_i は互いに独立に正規分布 $N(X_i^*,\ 1)$ で分布する。そこで（2.1），（2.2）式から

$$Y_i = \alpha + \beta(X_i - v_i) \qquad (\mathrm{i}=1,\ 2,\ \ldots,\ \mathrm{n}) \tag{2.3}$$

が得られる。これより回帰係数 β の最小 2 乗推定量 $\hat{\beta}$ は

$$\hat{\beta} = \beta \frac{M(x,\ x^*)}{M(x,\ x)} \tag{2.4}$$

となる。ここで $M(x,\ x) = (1/n)\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})$ である。$M(x,\ x^*)$ についても同様に定義される。周知のように，大標本の場合は（2.4）式の確率極限より

$$P\lim \hat{\beta} = \beta \left[\frac{p\lim M(x^*,\ x^*)}{p\lim M(x^*,\ x^*) + \sigma^2} \right] \tag{2.5}$$

となる。あきらかに右辺の括弧の内の係数は１より小である。ゆえにモデルの説明変数に含まれる測定誤差は，大標本に限定すれば，回帰係数の最小２乗推定量を過小に推定する。他方小標本の場合は（2.4）式の両辺の期待値をとり

$$E(\hat{\beta}) = \beta E \left[\frac{M(x,\ x^*)}{M(x,\ x)} \right] \tag{2.6}$$

となる。上式の右辺の期待値が推定量の「偏り」の大小を表す係数である。この係数が１であれば，測定誤差が存在しても回帰係数の最小２乗推定量には「偏り」が存在しない。しかしこれは必ずしも成立しない。また係数が１より小さいときは，回帰係数は過小に推定される。

以下これらのことを数値計算等で明らかにする。

3．偏りの定式

本節では（2.6）式の右辺の期待値で表された「偏り」係数を評価する。この左辺は次式のように変形される。

$$E\left[\frac{M(x,\ x^*)}{M(x,\ x)} \right] = \sum_{i=1}^{n} X_i^* E \left(\frac{A_i X}{X' A X} \right) \tag{3.1}$$

ここで X は X_i を要素とするｎ次列ベクトルであり，X' はその転置ベクトルである。また A はｎ次正方行列で次のように定義する。（3.1）式右辺の A_i は行列 A の第ｉ行の要素から構成されるｎ次行ベクトルとする。次に行列 A を直

$$A=\begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots\cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots\cdots & -\frac{1}{n} \\ \cdot & \cdot & \cdot \quad \cdot & \cdot \\ \cdot & \cdot & \cdot \quad \cdot & \cdot \\ \cdot & \cdot & \cdot \quad \cdot & \cdot \\ -\frac{1}{n} & -\frac{1}{n} & \cdots\cdots & 1-\frac{1}{n} \end{pmatrix}$$

交行列 C で直交化し，あらたにn次列ベクトル Z を $Z=CX$ と定義すれば $X'AX=Z'DZ$ が成立する。ここで D は行列 A のn個の固有根を対角要素とする対角行列で，そのうち $n-1$ 個は1で残りは0である。そこで一般性を失うことなく対角行列 D の第n対角要素を0とする。このとき（3.1）式の右辺の期待値は

$$E\left(\frac{A_iX}{X'AX}\right)=\sum_{j=1}^{n} b_{ij}E\left(\frac{Z_j}{Z_1^2+Z_2^2+\ldots+Z_{n-1}^2}\right) \tag{3.2}$$

と表される。上式右辺の係数 b_{ij} は行列 $B=CD$ の第ij要素を表す。ここで測定誤差 v_i に関する仮定より，説明変数 X_i は互いに独立に正規分布 N（X_i^*，1）に従う。これより変数 Z_i も互いに独立に正規分布 $N(\mu_i,1)$ にしたがって分布する。ただし平均 μ_i は

$$\mu_i=\sum_{j=1}^{n} C_{ij}X_j^* \tag{3.3}$$

である。ここで C_{ij} は直交行列 C の第ij要素とする。この結果（3.2）式右辺の分母 $(Z_1^2+Z_2^2+\ldots+Z_{n-1}^2)$ は自由度 $(n-1)$ の非心 x^2 分布すなわち $x^2(n-1,\delta)$ にしたがって分布する。ただし δ はその非心度を表し

$$\delta=\sum_{i=1}^{n-1} \mu_i^2=\sum_{i=1}^{n} (X_i^*-\overline{X}^*)^2 \tag{3.4}$$

である。これは説明変数の「真値」X^*_i の変動を表している。以上より，

(3.2) 式の右辺の期待値に関して次の命題1が成立する。

命題1 「説明変数 X_i が互いに独立に正規分布 $N(X_i{}^*,\ 1)$ に従うとき,

$$E\left(\frac{Z_j}{Z_1^2+Z_2^2+\ldots+Z_{n-1}^2}\right)=\mu_j E\left(\frac{1}{x^2(n+1,\ \delta)}\right) \tag{3.5}$$

が成立する。」

この命題は Judge and Bock (1978) の定理1 (p. 321) より証明することができる。[3] また上式 (3.5) の右辺の期待値の評価については次の命題がある。

命題2 「上記期待値の存在を仮定すれば,

$$E\left(\frac{1}{x^2(n+1,\ \delta)}\right)=\int_0^\infty (1+2\mathrm{t})^{\frac{-(n+1)}{2}} e^{\frac{-t}{1+2t}\delta}\,dt \tag{3.6}$$

が成立する。」[4]

さて以上の (3.1), (3.2), 命題1の (3.5) 式および命題2の (3.6) 式を前節の (2.6) 式に代入すれば, 説明変数の測定誤差に基づく回帰係数の「偏り」は次のように評価できる。ただし以下では測定誤差 v_i の分布を正規分布 $N(0,\ \sigma^2)$ に戻す。

定理 「回帰モデル (2.1), (2.2) において誤差を含んだ説明変数 X_i が正規分布 N $(X_i^*,\ \sigma^2)$ にしたがって分布するとき, 回帰係数の最小2乗推定量の期待値は

$$E(\hat{\beta})=\beta\left[\lambda\int_0^\infty (1+2\mathrm{t})^{\frac{-(n+1)}{2}} e^{\frac{-t}{1+2t}\lambda}\,dt\right] \tag{3.7}$$

となる。ただし係数λは自由度 $(n+1)$ の非心 x^2 分布の非心度で

$$\lambda=\frac{\delta}{\sigma^2}\quad,\quad \delta=\sum_{i=1}^{n}\ (X_i^*-\overline{X})^2$$

で表される。」[5]

上記定理の (3.7) 式の右辺括弧の値が, 回帰係数 β の最小2乗推定量の「偏り」係数である。同括弧内の積分値によっても影響されるが, 回帰係数の「偏り」のだいたいの傾向は非心度 λ によってきまる。すなわち非心度が高まれば

「偏り」は減少する，逆に非心度が低下すれば「偏り」は増加する。しかし係数が1すなわち回帰係数の不偏推定量が成立するためには，説明変数の真値の変動（δ）と測定誤差の変動（σ^2）との微妙な関係が必要であることがわかる。

4．数 値 計 算

本節では若干の非心度について「偏り」係数を数値計算する。そのために測定誤差の分布とその分散の関係について下記の表1を設定する。本報告では(a)誤差が平均0を中心として対称的に発生する場合に限定する。その典型的な場合が小数点以下の四捨五入である。これはケース(2)においてa＝0.5の場合に当たる。(b)一様分布を同一分散を持った正規分布によって近似させる。これはかなり制約的な方法である。厳密には，切断（truncated）正規分布等によって近似されるべきである。また正規分布に従うケース(1)は典型的な場合として取り上げる。測定誤差は理論上は無限に分布するが，その大半は平均値すなわち「真値」に集中すると考えられる。したがって正規分布にしたがうケース(1)は，「偏り」係数の一般的傾向を知るために重要である。

表1　測定誤差 v_1 の分布

	誤差分布	範　囲	平　均	分　散
ケース(1)	正規分布	$(-\infty, +\infty)$	0	σ^2
ケース(2)	一様分布 正規近似	$(-a, a)$	0	$a^2/3$

表2はケース(1)の数値計算の結果を主な非心度について掲載している。数値は（3.7）式右辺の括弧の値すなわち「偏り」係数である。この値が1であれば不偏推定量を意味する。この表より非心度（$\lambda=\delta/\sigma^2$）の上昇は「偏り」係数を引き上げる効果がある。したがって真値の変動（δ）を一定とすれば非心度の上昇は測定誤差の分散（σ^2）の低下を意味するので，誤差分散の低下は「偏り」係数を引き上げ，最小2乗推定量を不偏定量に近づける。また標本数の増加自体は「偏り」係数を低下させる傾向がある。なお表2の二，三の数値に関するモンテ・カルロ実験結果は脚注6）において説明している。

表3，4はケース(2)の数値計算の結果を掲載している。これらの表は，測定誤差の分布範囲が「偏り」係数におよぼす影響を表している。表3は標本数15および表4は同25の結果である。これらの表から誤差範囲の拡大は誤差分散の増加を通して「偏り」係数を引き下げる。また小数点以下の「四捨五入」（a＝0.5）は回帰係数にそれほど大きな「偏り」をもたらさない。

表2 「偏り」係数 ケース(1)

λ（非心度）	n＝5	n＝15	n＝25	n＝35
50	0.960	0.801	0.694	0.604
60	0.967	0.829	0.729	0.647
70	0.971	0.850	0.758	0.682
80	0.975	0.679	0.781	0.771
90	0.978	0.880	0.801	0.724
100	0.980	0.891	0.817	0.755
200	0.990	0.943	0.900	0.861
300	0.993	0.961	0.931	0.903
400	0.995	0.971	0.948	0.926
500	0.996	0.976	0.958	0.940
1000	0.998	0.988	0.978	0.969
2000	0.999	0.994	0.989	0.984
3000	0.999	0.996	0.993	0.989
4000	0.999	0.997	0.995	0.992
5000	0.999	0.998	0.996	0.994

表3 「偏り」係数 ケース(2) n＝15

δ（真値変動）	a（誤差範囲）					
	0.5	1.0	1.5	2.0	2.5	3.0
100	0.990	0.961	0.916	0.859	0.794	0.724
200	0.995	0.980	0.957	0.925	0.887	0.845
300	0.997	0.987	0.971	0.949	0.392	0.891
400	0.998	0.990	0.978	0.961	0.941	0.916
500	0.998	0.992	0.982	0.969	0.952	0.932
600	0.998	0.993	0.985	0.974	0.960	0.943
700	0.999	0.994	0.987	0.978	0.965	0.951
800	0.999	0.995	0.989	0.980	0.970	0.957
900	0.999	0.996	0.990	0.982	0.973	0.961
1000	0.999	0.996	0.991	0.984	0.976	0.965

表 4　「偏り」係数　ケース(2)　n＝25

δ (真値変動)	a (誤差範囲)					
	0.5	1.0	1.5	2.0	2.5	3.0
100	0.982	0.931	0.857	0.770	0.686	0.634
200	0.991	0.965	0.923	0.871	0.811	0.750
300	0.994	0.976	0.948	0.910	0.866	0.817
400	0.995	0.982	0.960	0.931	0.896	0.857
500	0.996	0.986	0.968	0.944	0.915	0.883
600	0.997	0.988	0.973	0.953	0.929	0.900
700	0.997	0.990	0.977	0.960	0.938	0.913
800	0.998	0.991	0.980	0.965	0.946	0.923
900	0.998	0.992	0.982	0.968	0.951	0.931
1000	0.998	0.993	0.984	0.971	0.956	0.938

5．結　　　語

この報告は経済変数に含まれる測定誤差が，回帰係数の最小 2 乗推定量に与える「偏り」について検討した。この報告は，(1)一つの説明変数から構成されるごく単純な回帰モデルに限定し，測定誤差が回帰係数の推定結果に与える「偏り」を特に小標本の下で定式化した。(2)測定誤差を正規分布あるいは正規近似の一様分布にしたがう確率変数とした。また(3)誤差分布の「分散」を測定時における「四捨五入」等に関連付けて分析した。その結果，（a）測定誤差の範囲の拡大は回帰係数の「偏り」係数を 0 方向に引き下げる。また（b）測定値の「真値」の変動の増加は回帰係数の「偏り」係数を 1 方向に引き上げる方向に働く。この結果（c）大標本の場合と同じように，小標本の場合も測定誤差は最小 2 乗推定量を過小推定する。

注

1）Greene（2000, p. 376）

2）この報告では測定誤差を含む回帰モデルの推定方法については触れない。各種の推定方法については Greene（2000, p. 376），洲浜（1979, p. 85）等を参照されたい。

3）この命題 1 と次の命題 2 は，本文（3.5）式の左辺の分母および分子の結合積率母関数を経由して直接証明することも可能である。

4) 一般に統計量 q ($q>0$) の積率母関数の存在を仮定して，それを $\phi(t)=E(e^{tq})$ とすれば，

$$E\left(\frac{1}{q}\right)=\int_0^{\infty}\phi(-t)\,dt$$

という関係が容易に成立する（洲浜 1988, p. 118)。そこで非心カイ 2 乗分布にしたがう統計量 $x^2(n+1,\ \delta)$ の積率母関数（ただし $\phi(-t)$）を，上記の被積分関数に代入すれば本文の (3.6) 式をえる。

5) 各係数間に次の関係がある。

$$\sum_i X_i^* \sum_j b_{ij}\mu_j = \sum_i X_i^* \sum_j b_j C_{ji} \sum_l C_{jl} X_l^*$$

$$= \sum_{i,l}\left(\sum_j b_j C_{ji} C_{il}\right) X_i^* X_l^*$$

$$= \sum_{i,l} a_{jl} X_i^* X_l^* = X^{*}AX^{*} = \delta$$

ここ d_j は対角行列 D の第 j 要素，a_{ij} は行列 A の第 i，j 要素とする。

6) 下表の二つの仮説例について定理 (3.7) 式のモンテ・カルロ実験（試行回数 1,000 回）をおこなった。結果は表の最後に掲載している。

「偏り」係数に関するモンテ・カルロ実験

	仮設例A	仮設例B
回帰モデル	$Y_i=\alpha+\beta X_i^*$, $X_i=X_i^*+v_i$ $X_i^*=1(1)15$, $n=15$ $\alpha=10$, $\beta=1$	
v_i δ λ	N(0, 4) 280 70	N(0, 0.25) 280 1120
「偏り」係数 モンテ・カルロ値 理論値	 0.850 0.850	 0.990 0.989

参　考　文　献

Greene, William H. (2000), *Econometric Analysis Forth Edition*, Prentice-Hall, Inc, New Jersey.

Judge, G. C. and M. E. Bock (1978), *The Statistical Implications of Pre-Test and Stein-rule*

Estimators in Econometrics, North-Holland, Amsterdam.
洲浜源一（1979）『観察不可能な変数を含む経済モデルの推定』（大阪府立大学経済研究叢書第50冊）大阪府立大学経済学部。
同　上（1988）「Cochrane-Orcutt 推定量の exact な分散とその効率－母平均推定モデルの場合－」『大阪府立大学経済研究』（大阪府立大学経済学会），33巻2号，111－125.