慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	計量経済学における頑健推定(1)
Sub Title	Robust estimation in econometrics (1)
Author	蓑谷, 千凰彦
Publisher	慶應義塾経済学会
Publication year	1989
Jtitle	三田学会雑誌 (Keio journal of economics). Vol.82, No.3 (1989. 10) ,p.413(1)- 455(43)
JaLC DOI	10.14991/001.19891001-0001
Abstract	
Notes	論説
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00234610-19891001-0001

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

「三田学会雑誌」82巻3号（1989年10月）

計量経済学における頑健推定（1）*

蓑谷千凰彦

はじめに

頑健推定 robust estimation[(1)] は1964年 Huber の先駆的論文 "Robust estimation of a location parameter"（*Annals of Mathematical Statistics*, 35）によって基礎が築かれて以来，最近20年間に著しい発展を遂げている統計学の新しいパラメータ推定法である。

残念ながら，計量経済学の分野においては頑健推定が試みられた例はきわめて少ない。わずかに Fair〔6〕，Welsch〔28〕の論文を散見するにすぎない。Koenker〔21〕の "Robust methods in econometrics" という頑健法に関するサーベイ論文があるが，計量経済学への頑健推定の応用を意図した内容ではなく，応用例も含まれていない。計量経済学の分野における頑健推定の理論と実際例はほとんど皆無であるから，本論文は計量経済学における頑健推定のサーベイ論文ではない。頑健推定も計量経済学の分析道具の箱の中へ入れるべき有用な方法であることを理論・応用両面から示すことが本論文の目的である。

なぜ頑健推定が必要であるかを考えてみよう。

古典的な線形回帰モデルを用いて実証分析を行なうとき，モデルにはさまざまな諸仮定が課されている。数学的に定式化されたモデル自体，1つの仮説の提示であるが，このモデルは，パラメータに関して線形であり，誤差項は平均0，均一分散，自己相関なしの正規分布に従い，説明変数とは統計的に独立であると仮定される。分析に用いられる統計的推測方法が，このような諸仮定のいずれかが成立しないとき，いいかえれば，いわば「理想モデル」からのわずかな乖離が生じたとき，その乖離に過剰に，あるいは敏感に反応し，「理想モデル」からの結論を大幅に変えるならば，統計分析の方法として好ましくない。

古典的正規線形回帰モデルの仮定の1つ，誤差項の正規性を考えてみよう。これまで正規分布のもとで最適と考えられてきた統計的方法のいくつかは，最近数年間の研究によって，正規分布からのわずかな乖離に対して，きわめて敏感であることが明らかになってきた。とすれば線形回帰モデルの誤差項の

(1) 正規性の仮定は適切なのか。

---

* この研究は慶應義塾学事振興資金による研究助成金を受けている。記して感謝したい。

注（1） robust estimation は『数学辞典』（第3版，岩波書店）ではロバスト推定と訳されているが，わが国では頑健推定という訳語が定着しつつあり，本論文もそれにしたがった。

(2) 正規性からの乖離に対して，通常の最小2乗推定量は，そして正規性の仮定の下で成立している t 検定や F 検定はどの程度頑健なのか。

(3) 正規性からの乖離に対していかに対処すべきか。

という問題が発生する。

本論文は上記(3)に焦点を絞りたい。論文全体の概要を述べておこう。

1節では線形回帰モデルにおける誤差項の正規性の仮定について簡単に検討し，正規分布より長い尾をもつ分布[(2)]の例をあげる。長い尾をもつ分布かどうかは正規性の検定でチェックできない場合が多い。本論文を通して私が採用する方法は，計量経済学ではこれまで使われたことのない順序化されたスチューデン化残差と正規分布の変位値 quantile の比較による正規確率プロットである。

2節で最小2乗法と t 検定の頑健性について従来主張されてきた内容を検討し，頑健推定の必要性を論ずる。3節では，正規分布より長い尾をもつ分布の例として Huber の最小好適分布 least favorable distribution (l. f. d.) を説明し，l. f. d. の正規確率プロットを与えた。

4節からは頑健推定を行なうための理論的準備である。影響関数（4節），影響関数の有限標本近似（5節），影響関数にもとづく影響力の尺度（6節），ジャックナイフ推定法（7節）の説明をする。前述の Huber の先駆的論文以来，頑健推定の分野におけるもっとも重要な理論的成果は Hampel〔12〕の影響関数である。影響関数の観点からみれば，「外れ値」に対する頑健推定は被説明変数の誤差の方向のみを考え，説明変数の高い作用点（説明変数の観測値の平均点から遠く離れた所に位置している点）のことは考慮されていないことがわかる。最近の研究はこの高い作用点にも頑健なパラメータ推定へと関心が向いている。この影響関数の議論は頑健推定との関連ばかりでなく，ジャックナイフ推定法，回帰診断の統計量である Cook の D, DFFITS などとも関連をもっている。したがって別々に論じられてきた回帰診断と頑健推定は統一的に用いられるべき方法であることを強調したい。

8節から10節は，最尤法を一般化したM推定量を説明する。そして影響関数と比例関係にあるM推定の ψ 関数のなかでもとくに Huber の ψ が計量経済学においても有用なM推定値を与える。したがって11節では Huber の ψ についてその特徴を検討し，総誤差感度 (g. e. s.)，局所方向感度 (l. s. s.)，漸近的分散などを計算した。以上11節までが本号の内容である。M推定量の求め方，計量経済モデルの頑健推定の実際例，応用上生ずる問題とその解決は本論文(2)で示す。

1 正規性の仮定

線形回帰モデルの誤差項に正規分布が仮定されるとき，援用されるのは中心極限定理 CLT である。計量経済分析の行動方程式の導出にあたって，経済主体の行動法則から出発し，次に集計概念

注（2） long tailed あるいは heavy tailed とよばれる分布のことであり，両すそが厚い，すそが広がっているともいわれる。

へと進むのは，個人あるいは企業の行動に含まれている確率的要素が，集計によって平均的な市場関係式を導くからである。

「個人の関係式においては確率要素の分布に関する仮定が非常に弱いにしても，すべての市場なりすべての社会なりについて誘導される平均のあるいは合計の関係式は，大数の法則によってほぼ正規分布するいくつかの合成確率変数（たとえば，個人の誤差項の合計）によって特性が与えられるのが普通である[3]」

大数の法則によってというよりは CLT によってと言うべきであろうが，Haavelmo のこの見解を，計量経済学者は受け入れてきた。線形回帰モデルの，とくに集計式の誤差項が無数の非系統的要因[4]の総和であることは認めるとしても，CLT 成立のための条件であるそれら個々の要因のその総和である誤差項への影響が微小であること，およびそれらの諸要因が独立であることをいかにしてわれわれは知ることができるのであろうか。

非系統的要因として処理されたすべての諸要因が何であるかがわかっていないにもかかわらず，それらの影響力の大きさ，それらの独立性に言及することも奇妙な話である。なぜなら誤差項の中に含められている多数の，影響力の小さい，系統的でない諸要因は，管理された実験においてそれぞれ互いに独立となるようにコントロールされた要因ではなく，われわれの想像能力のみがその存在と構造を仮定しうるにすぎないからである[5]。

CLT が成立するから誤差項の正規性は正当化されるというよりも，正規分布が成立している筈だから CLT がその論拠として援用されるという論理の逆転がここにはある。モデルがどのように定式化されようと，誤差項の分布はいつも正規分布が仮定される。あるいは誤差項は正規分布に従う筈であり，正規分布をしないとすればモデルが悪いからである，と暗黙のうちに仮定されている。19世紀末から20世紀初頭にかけての統計学の世界の雰囲気を表現した次の名言がある。「数学者は，正規分布は実験によって得られた結果であると信じ，実験者は正規分布を数学の定理であると信じ，誰もが正規分布を信じている」（ポアンカレ『確率の計算』(1912) に引用され有名になった Lippman の言葉)。この 19 世紀末の正規分布信仰の世界にわれわれ計量経済学者は依然として安住し，惰眠を貪っているのではないであろうか。

モデルは，その定式化，誤差項に関する諸仮定を含めて，いうまでもなく，つねに仮説的である。仮説的とはいえ，パラメータを推定し，検定にかけようとする自らのモデルを不適切な諸仮定に立脚しているモデルと考えている人はいない。したがって適切と考えているモデルのパラメータを通常の最小 2 乗法 OLS で推定する。このとき外れ値 outlier と判断される大きな残差があるとき，それは系統的要因が欠除しているか，あるいは説明変数のラグ構造が適切でないか，あるいは突発的事故や制度変更などに対処していない等々の意味でモデルの定式化が不適切である証拠とみなさ

注（3） Haavelmo〔10〕p. 69.

（4） システマティックに被説明変数に影響を与える要因以外の諸要因のことであり，回帰モデルで一括して誤差項として処理されている。

（5） 蓑谷〔24〕p. 99～100.

れる。そしてこれは，残差の大きさなど定式化を変えればどんな風にでも変わるさという自嘲の響きとなる。この自嘲には1片の真実と，自らのモデルに対する不信表明——どうせ自分のモデルは大したものじゃない——が混在している。定式化次第でどんなふうにでも変わる残差であるとすれば，OLS で十分ではないかという諦めと，ある意味では居直りでもある。そしてこの自嘲は，誤差項の分布が何であろうと，確率誤差項が平均0で均一分散であり，自己相関をもたず，線形回帰モデルの説明変数と統計的に独立であるならば，通常の最小2乗推定量 OLSE は最良線形不偏推定量 BLUE であり，さらに誤差項が正規分布するならば，OLSE は最小分散不偏推定量になるという OLS への強い信頼感とは全く別である。

ところで，大きな残差は正規分布からの外れ値であり，(無意味なダミー変数を説明変数に含めるという場合は除いて) 定式化を変えれば外れ値ではなくなるのであろうか。大きな残差はつねに定式化の不適切と考える立場は，決定係数の高さを競う当てはめゲームへと惰していかないであろうか。外れ値ではないかと疑いをもたれている大きな残差の値が，モデルの定式化のミス（重要な系統的要因が欠除している，あるいは無意味な要因が含まれているという定式化のミス）ではなく，誤差項が正規分布より長い尾をもつ（両すその厚い）分布からの実現値とすれば，OLS はむしろ危険である。

実際，他の分野の確率モデルと同様，計量経済モデルにおいても，誤差項の分布は正規分布より長い尾をもつ分布であることが多いということもわかってきた。

例として，貨幣賃金率変化率，EC から日本向け輸出および1次同次のコブ・ダグラス生産関数の定式化をあげよう。図1，図2，図3はそれぞれ次の(1)，(2)，(3)式の最小2乗残差のスチューデント化残差 r_i を順序化して，標準正規分布の変位値 quantile（この値は rankits とよばれることもある）に対してプロットした図である（以下，正規確率プロットとよぶことにする）。

$$WDOT=\underset{(-4.228)}{-4.8475}+\underset{(8.184)}{16.518}\left(\frac{1}{RU}\right)+\underset{(10.759)}{0.90106}\,CPIDOT \tag{1}$$

$$\bar{R}^2=0.932,\ s=1.69(\%),\ DW=1.355 \qquad 1965\sim1987\text{年度}$$

ここで

$WDOT$＝貨幣賃金率変化率（%）

RU＝完全失業率（%）

$CPIDOT$＝消費者物価変化率（%）

係数の下の（　）内の値は t 値

$$\log QXECJ\$=\underset{(1.996)}{4.9221}+\underset{(8.702)}{1.0532}\log GNPJ\$-\underset{(-3.029)}{1.2711}\log\left(\frac{PXEC\$}{WPIJ\$}\right)_{-1} \tag{2}$$

$$\bar{R}^2=0.923,\ s=0.113,\ DW=1.662 \qquad 1967\sim1987\text{年度}$$

ここで

$QXECJ\$$＝EC から日本向け輸出（1987年基準，億ドル）

$GNPJ\$$＝日本の実質 GNP（1987年基準，100億ドル）

$PXEC\$$＝EC の輸出価格指数（1987年＝100，ドル建て）

図 1　*WDOT* の正規確率プロット

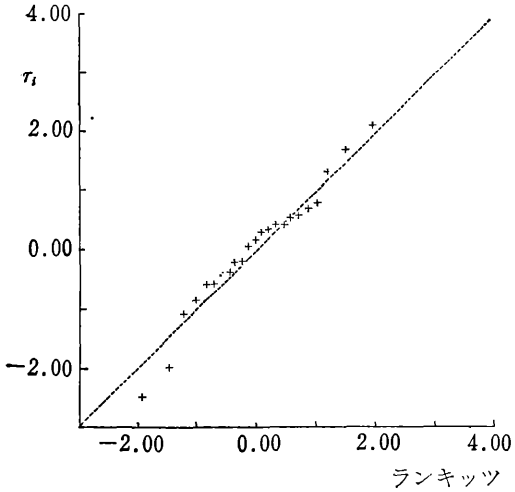


図 2　log *QXECJ*$ の正規確率プロット

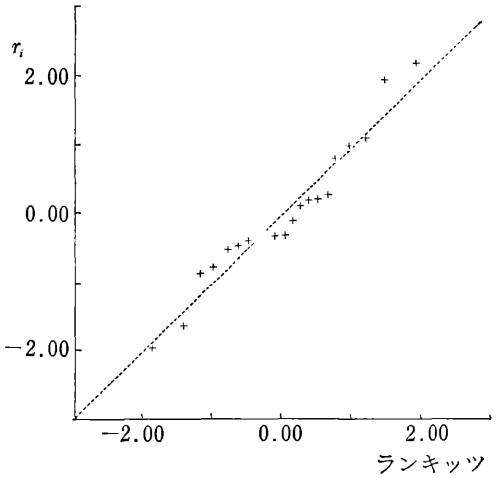
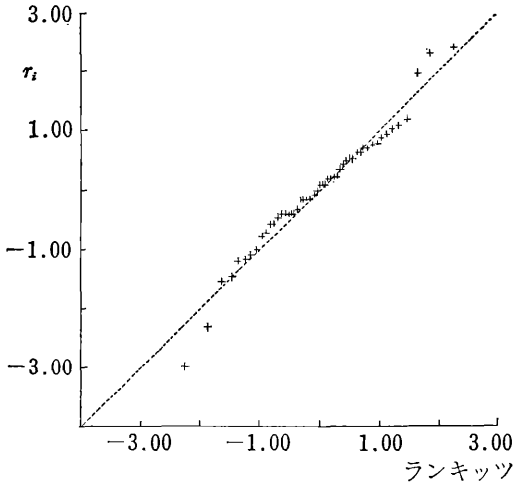


図 3　log *VL* の正規確率プロット



WPIJ$＝日本の卸売物価指数（1987年＝100，ドル建て）

$$\log VL = 5.2573 + 0.43774 \log KL \tag{3}$$
$$(11.768)\quad(9.161)$$

$\bar{R}^2 = 0.624$，$s = 0.200$　　1987年度製造業51産業

ここで *VL*，*KL* は，ともに資本金10億円以上の法人企業の値であり，

VL＝従業員1人当り粗付加価値額（千円）

KL＝労働装備率（千円）

である。

(1)〜(2)式の推定に用いたデータはすべて本論文の(2)に掲げる。これら3式の最小2乗残差を e_i とすれば，スチューデント化残差 r_i は

$$r_i = \frac{e_i}{s(1-h_{ii})^{\frac{1}{2}}} \tag{4}$$

によって与えられる。ここで s=誤差項の標準偏差の推定値，h_{ii}=ハット行列 H の対角要素である。図1～図3の r_i は(4)で与えられる r_i を小さい順に順序化した値とランキッツとの対応である。

図1，図2，図3いずれも(1)～(3)の線形回帰モデルの誤差項の分布は正規分布より長い尾をもつ分布に従っているということを示している（次節参照）。真の分布が正規分布より長い尾をもつ分布であれば，正規分布のもとでは外れ値と判断される極端な値が観測される可能性は高い。OLS はすべての残差に等ウエイトを与えるから，（絶対値で）極端に大きい残差にも小さい残差と同じウエイト1を与え，したがってこのような極値に対して OLS は無防備である。すなわち OLS は，正規分布の仮定のもとでは「外れ値」とみなされる値に対して無防備であり，この「外れ値」の OLSE への影響は大きく現れる。

このように真の分布が正規分布より長い尾をもつ分布であり，「外れ値」とみなされる値が，正規分布のもとで観測されるよりもこの分布のもとではより大きな確率で観測されるとすれば，極端な値にも同じウエイトを与える無防備な OLS ではなく，このような極端な値に対してはウエイトをダウンする頑健な推定法が望ましい。頑健推定は，後述するように，外れ値に対して無防備な OLS にくらべて「抵抗力」resistance があり，真の分布が正規分布のときにも推定量の有効性 efficiency を余り失わない推定法である。ここで「抵抗力」のある resistant とは，「推定値あるいは検定統計量の値が，標本データの小さな変化（標本データすべての小さな変化あるいは若干の大きな変化）に対して敏感に反応しない[(6)]」という性質である。

2　最小2乗法と t 検定の頑健性

誤差項の分布が正規分布をしていなくても，OLS によるパラメータ推定と t 検定は深刻な影響を受けないといわれている。この一般に受け入れられている見解の論拠となっているのは次の2点である。

(1) OLSE は誤差項の分布に関わりなく最良線形不偏性をもつ。したがって誤差項が正規分布から逸脱していようと，あるいはいかなる分布に従っていようと，OLS は線形不偏推定量のクラスのなかで最適な推定量を与えるから，OLSE の特性について思い煩うことはない。

ガウス・マルコフの定理で知られるこの OLSE の最良線形不偏性は確かに魅力的ではある。次の3点のみ OLS の問題点として指摘しておこう。

①誤差項が正規分布に従わないとき，OLSE は最尤推定量ではない。いいかえれば，誤差項が正規分布に従っていないとき，OLSE にのみ拘泥していれば，線形推定量の中からすべての最尤推定量が除外される。

②歪みがあり，不均一な分散をもつ誤差項の分布においては，Box-Cox 変換で知られる非線形

注（6） Huber〔19〕p. 7.

変換が有効であることが認識されつつある。[7] したがって推定量を選ぶべきクラスを線形推定量のクラスにのみ限定すべきではない。線形推定量のクラスから離れれば，ガウス・マルコフの定理によって支えられている OLSE の優位性はくずれる。

③すでに1節で述べたように，真の分布が正規分布より長い尾をもつ分布であるならば，正規分布のもとでは「外れ値」とみなされる大きな残差も，この分布のもとでは外れ値ではない。この「外れ値」に対して OLS は無防備であり，OLSE はこの「外れ値」から大きな影響を受ける。

(2) 最小2乗法と正規分布の仮定にもとづいた回帰係数の t 検定は，誤差項が正規分布から若干逸脱しても，漸近的に正しい有意水準を与える。すなわち正規分布の仮定から導かれた t 検定による有意性は，正規分布の仮定自体が厳密に成立せず，少し乖離していたとしても，大きな影響を受けない (validity)。(これは妥当性の頑健 robustness of validity とよばれている)。したがって正規分布からのズレがあっても，通常の検定の手続きと結論に関して心配する必要はない。

t 検定の頑健性とはこの妥当性の頑健のことであった。しかし検定力の観点から考えてみよう(効率の頑健性 robustness of efficiency)。信頼区間や標準誤差の大きさは正規分布からのズレに対して全く頑健ではないことがわかってきた。[8] いいかえれば，正規分布からの逸脱があるとき，未知パラメータ0という帰無仮説を0でないという対立仮説に対して検定する場合に検定の有意水準（第I種の過誤確率）については漸近的に信頼できるが，検定力は小さくなるから（第II種の過誤確率は大きくなるから)，有意な結果（H_0 を棄却し，H_1 を採択する）は信頼できるが，有意でない結果（H_0 を棄却できない。H_1 を支持する証拠はない）は，OLSE の低い有効性にもとづいているのかも知れないということである。線形回帰モデルにおける実際の誤差項の分布は，正規分布より両すその厚い分布であることが多いとすれば，検定の観点からもやはり頑健推定の必要が生ずる。

3 長い尾をもつ分布と頑健推定

標準正規分布より長い尾をもつ分布の例として次の望ましさが最小であるという意味の最小好適分布 least favorable distribution（以下 l. f. d. と略す）を考えよう。

l. f. d. の確率密度関数は次式で与えられる。

$$f(x)=\begin{cases}\dfrac{1-\varepsilon}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} & |x|\leqq k\\[2ex] \dfrac{1-\varepsilon}{\sqrt{2\pi}}e^{\frac{k^2}{2}-k|x|} & |x|>k\end{cases} \tag{5}$$

$f(x)$ は確率密度関数であるから，パラメータ k と ε は次の関係にある（証明は数学注(1)を参照)。

注（7） たとえば Carroll and Ruppert〔2〕をみよ。
（8） Hampel et al. [13] p. 405.

表 1　正規分布の変位値と最小好適分布（l. f. d.）の変位値（$n=30$）

i	p_i	正規分布の変位値 (rankits)	最小好適分布（l. f. d.）の変位値				
			$\varepsilon=0.01$ $k=1.945$	$\varepsilon=0.05$ $k=1.399$	$\varepsilon=0.10$ $k=1.140$	$\varepsilon=0.15$ $k=0.98$	$\varepsilon=0$ $k=1.345$
1	0.0207	−2.0407	−2.1474	−2.5390	−2.9596	−3.3658	−2.6533
2	0.0531	−1.6159	−1.6592	−1.8644	−2.1318	−2.4027	−1.9516
3	0.0862	−1.3648	−1.3919	−1.5190	−1.7067	−1.9082	−1.5913
4	0.1193	−1.1786	−1.1981	−1.2859	−1.4216	−1.5766	−1.1786
5	0.1524	−1.0262	−1.0412	−1.1072	−1.2068	−1.3267	−1.0262
6	0.1855	−0.8944	−0.9064	−0.9582	−1.0340	−1.1264	−0.8944
7	0.2186	−0.7767	−0.7863	−0.8279	−0.8876	−0.9582	−0.7767
8	0.2517	−0.6688	−0.6767	−0.7104	−0.7581	−0.8135	−0.6688
9	0.2848	−0.5682	−0.5746	−0.6019	−0.6402	−0.6841	−0.5682
10	0.3179	−0.4731	−0.4782	−0.5001	−0.5306	−0.5653	−0.4731
11	0.3510	−0.3821	−0.3862	−0.4033	−0.4271	−0.4541	−0.3821
12	0.3841	−0.2942	−0.2973	−0.3102	−0.3281	−0.3482	−0.2942
13	0.4172	−0.2086	−0.2107	−0.2197	−0.2322	−0.2461	−0.2086
14	0.4503	−0.1245	−0.1258	−0.1311	−0.1384	−0.1467	−0.1245
15	0.4834	−0.0414	−0.0418	−0.0436	−0.0460	−0.0487	−0.0414
16	0.5166	0.0414	0.0418	0.0436	0.0460	0.0487	0.0414
17	0.5497	0.1245	0.1258	0.1311	0.1384	0.1467	0.1245
18	0.5828	0.2086	0.2107	0.2197	0.2322	0.2461	0.2086
19	0.6159	0.2942	0.2973	0.3102	0.3281	0.3482	0.2942
20	0.6490	0.3821	0.3862	0.4033	0.4271	0.4541	0.3821
21	0.6821	0.4731	0.4782	0.5001	0.5306	0.5653	0.4731
22	0.7152	0.5682	0.5746	0.6019	0.6402	0.6841	0.5682
23	0.7483	0.6688	0.6767	0.7104	0.7581	0.8135	0.6688
24	0.7814	0.7767	0.7863	0.8279	0.8876	0.9582	0.7767
25	0.8145	0.8944	0.9064	0.9582	1.0340	1.1264	0.8944
26	0.8476	1.0262	1.0412	1.1072	1.2068	1.3267	1.0262
27	0.8807	1.1786	1.1981	1.2859	1.4216	1.5766	1.1786
28	0.9138	1.3648	1.3919	1.5190	1.7067	1.9082	1.5913
29	0.9469	1.6159	1.6592	1.8644	2.1318	2.4027	1.9516
30	0.9793	2.0407	2.1474	2.5390	2.9596	3.3658	2.6533

$$\frac{2\phi(k)}{k}-2\Phi(-k)=\frac{\varepsilon}{1-\varepsilon} \tag{6}$$

ここでϕは標準正規分布の確率密度関数，Φはその累積分布関数である。

(5)式で $\varepsilon=0$, $k\to\infty$ のとき，l. f. d. は標準正規分布になる。次に l. f. d. の変位値 quantile を求めよう。累積分布関数を $F(x)$ とすれば，$F(x)=p$ に対する変位値は $F^{-1}(p)$ によって与えられるから，まず l. f. d. の $F(x)$ を求めると次のようになる（数学注(2)参照)。

$$F(x)=\begin{cases}(1-\varepsilon)\dfrac{\phi(k)}{k}e^{k^2+kx} & x<-k\\[2ex] \dfrac{\varepsilon}{2}+(1-\varepsilon)\Phi(x) & -k\leq x\leq k\\[2ex] 1-(1-\varepsilon)\dfrac{\phi(k)}{k}e^{k^2-kx} & x>k\end{cases} \tag{7}$$

図 4　標準正規分布と l. f. d.

（a）　確率密度関数　　（b）　変位値

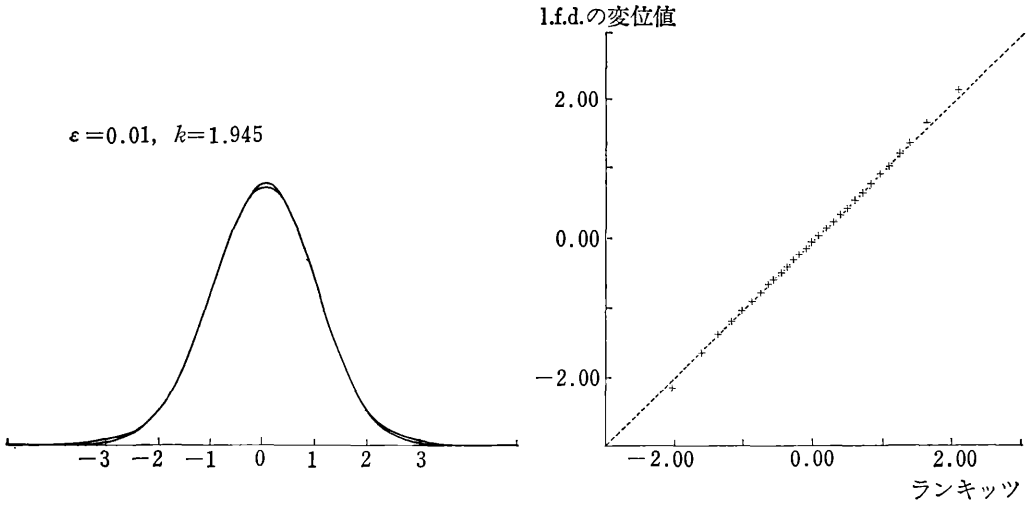
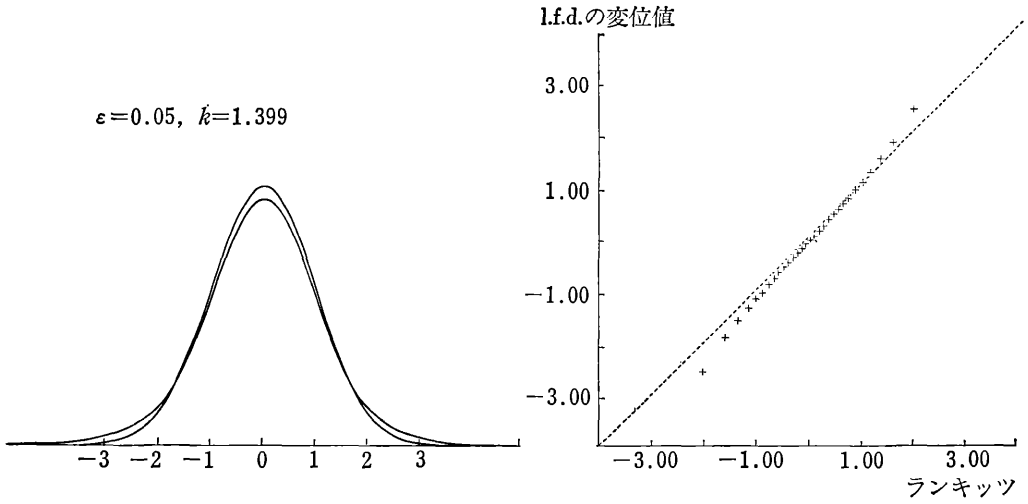


図 5　標準正規分布と l. f. d.

（a）　確率密度関数　　（b）　変位値



したがって $F(x)=p$ とすると，l. f. d. の変位値は次式によって与えられる。

$$F^{-1}(p)=\begin{cases}\dfrac{1}{k}\{\log p-\log(1-\varepsilon)-\log\phi(k)+\log k-k^2\} & x<-k\\ \Phi^{-1}\left(\dfrac{p-\frac{\varepsilon}{2}}{1-\varepsilon}\right) & -k\leq x\leq k\\ \dfrac{1}{k}\{-\log(1-p)+\log(1-\varepsilon)+\log\phi(k)-\log k+k^2\} & x>k\end{cases} \tag{8}$$

以上の結果を用いて，$n=30$ のとき l. f. d. の変位値と正規分布の変位値（ランキッツ）を計算したのが表 1 である。l. f. d. の変位値は(6)式によって関係づけられている 4 通りの（ε, k）の組合せ，

図 6　標準正規分布と l. f. d.

（a）　確率密度関数　　（b）　変位値

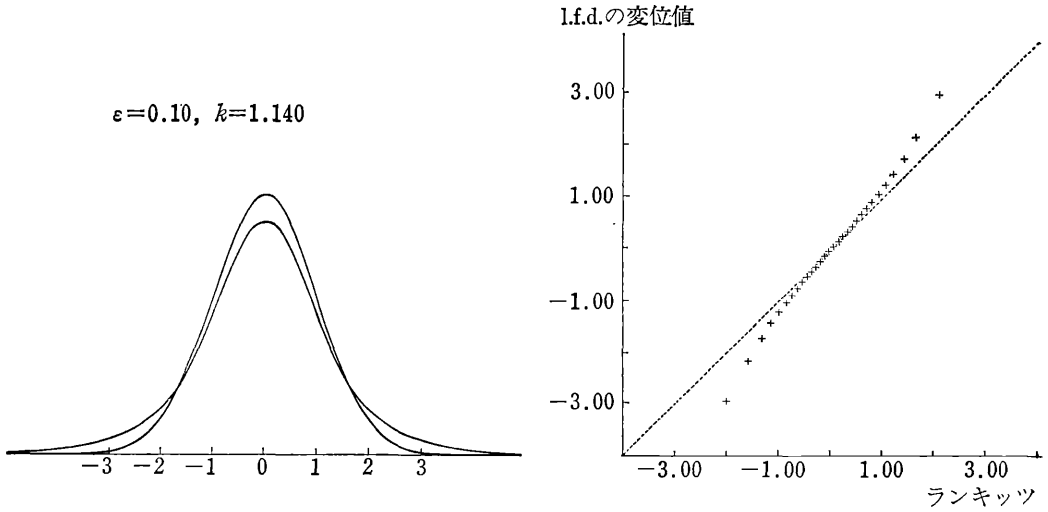
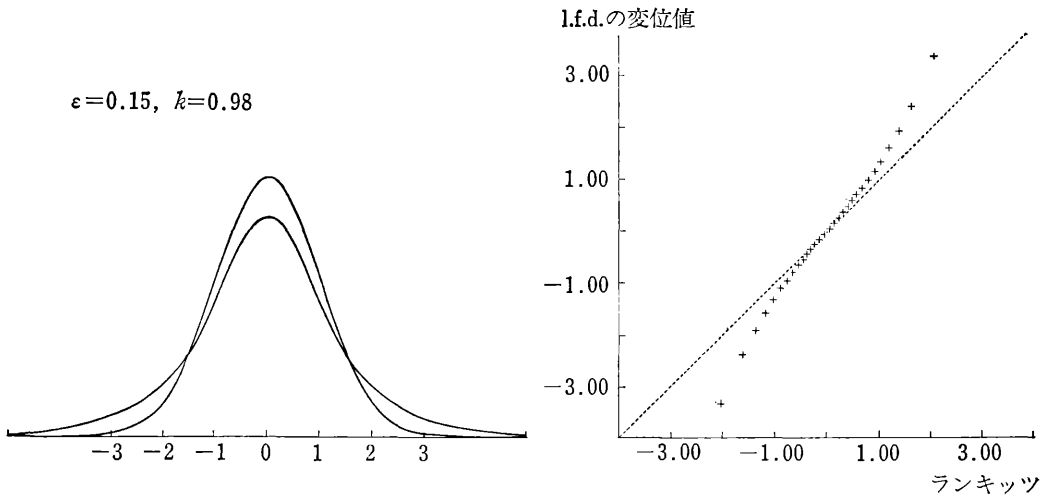


図 7　標準正規分布と l. f. d.

（a）　確率密度関数　　（b）　変位値

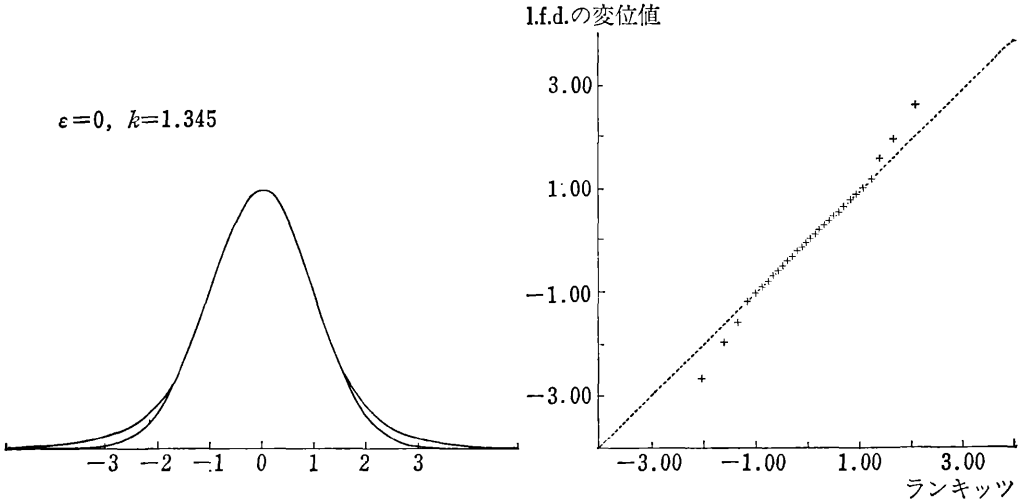


(0.01, 1.945), (0.05, 1.399), (0.10, 1.140), (0.15, 0.98) と，(6)式との関係ではなく，後述する Huber の ψ関数において，正規分布が正しいとき95%の漸近的有効性を与える (0, 1.345) のケース，合計5ケースについて計算されている。また図4から図8はこの l. f. d. の5ケースについて，標準正規分布との比較で描かれた確率密度関数および正規確率プロットである。

$\varepsilon=0$, $k\to\infty$ のとき l. f. d. は標準正規分布になることはすでに述べた。ε が小さく，k の大きい図4の $\varepsilon=0.01$, $k=1.945$ のケースにおいては，l. f. d. の確率密度関数は標準正規分布とほとんど同じであり，正規確率プロットにおいても，l. f. d. の変位値はほぼ 45° 線上にある。これに対して $\varepsilon=0.15$, $k=0.98$ と，ε は大きく，k は小さいケースの l. f. d. は正規分布よりかなり長い尾をもつ

図 8　標準正規分布と l. f. d.

（a）　確率密度関数　　（b）　変位値



分布であり，変位値も極値が45°線よりかなり乖離している。図7(b)は正規分布より長い尾をもつ分布の典型的なパターンである。図1～図3の正規確率プロットから，誤差項の分布は正規分布より長い尾をもつ分布であると判断した理由が今や明らかであろう。

表1からは，たとえば次のようなことがわかる。−2.04 以下あるいは 2.04 以上の値が観測される確率は，正規分布のとき約0.04と小さい。しかし同じ確率で $\varepsilon=0.10$, $k=1.140$ の l. f. d. の場合には絶対値で3をこえる値が観測される。いいかえれば l. f. d. からは $\varepsilon=0.10$, $k=1.140$ の場合は確率0.10で，$\varepsilon=0.15$, $k=0.98$ の場合は確率0.17で絶対値が2をこえる可能性があり，正規分布の場合より極端に小さいあるいは大きい値が得られる可能性は高い。

前節で OLSE は外れ値に対して無防備であると述べたが，推定量が観測値に対してどのように反応するかを調べるためには影響関数あるいる影響曲線という概念が有用である。

4　影　響　関　数[(9)]

推定量の影響関数 influence function あるいは影響曲線 influence curve という概念は Hampel〔11〕,〔12〕によって導入された。この影響関数は，推定量の個々の観測値に対する感度 sensitivity を測ること，および推定量の漸近的分散を計算するという2つの目的をもっている。

$z_1, z_2, \cdots\cdots, z_n$ は分布関数 F からの無作為標本であり，パラメータ θ の推定量を

$$T_n = T_n(z_1, z_2, \cdots\cdots, z_n)$$

注（9）　4～6節は主として Cook and Weisberg〔4〕, Hampel〔11〕,〔12〕, Hampel et al.〔13〕, Huber〔19〕, Chatterjee et al.〔3〕にもとづいている。

としよう。そして F_n を n 個の $z_1, z_2, \cdots\cdots, z_n$ にもとづく経験的分布関数とする。また

$$T_n = T_n(z_1, z_2, \cdots\cdots, z_n) = T(F_n)$$

と表わし

$$T(F_n) \xrightarrow[n\to\infty]{} T(F)$$

とする。たとえば

$$T_n = \frac{1}{n}\sum_{i=1}^{n} z_i = \bar{z}$$

とすれば

$$T(F) = \int z dF(z) = E(z) = \mu$$

$$T(F_n) = \int z dF_n(z) = \bar{z}$$

である。

このとき，きわめて大きな無作為標本に1個の観測点 z が追加される場合，T がどのような影響を受けるかは次の影響関数によって示される。

$$IC(z; F, T) = \lim_{\varepsilon\to 0} \frac{T[(1-\varepsilon)F + \varepsilon\delta_z] - T(F)}{\varepsilon} \tag{9}$$

ここで δ_z は z で1をとり，それ以外では0の値をとる分布関数である。この IC は有限標本における $T(F_n)$ への z の影響ではなく，無限標本における $T(F)$ $(\xleftarrow[n\to\infty]{} T(F_n))$ への z の影響を示している。

分布関数 $F_\varepsilon = (1-\varepsilon)F + \varepsilon\delta_z$ は，ε が十分小さく0に近いときほとんど F に等しいが，点 z で追加的なウエイト $\varepsilon\delta_z$ をもっている。したがって，$T[(1-\varepsilon)F + \varepsilon\delta_z] - T(F)$ は，点 z における推定量 T_n の確率収束 T のわずかな変化を示しており，IC は $\varepsilon=0$ で評価した $T(F_\varepsilon)$ の導関数，すなわち

$$\left.\frac{dT(F_\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0}$$

である。

たとえば，$T_n = \bar{z}$ とすれば

$$T(F) = \mu = \int z dF(z)$$

であるから

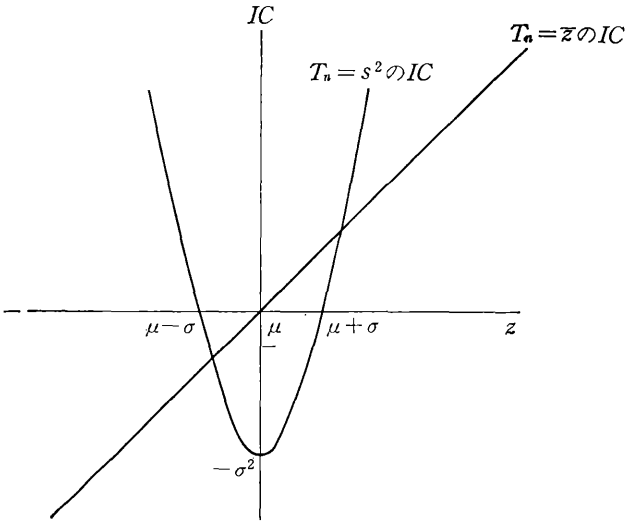
$$T[(1-\varepsilon)F + \varepsilon\delta_z] = \int (1-\varepsilon) z dF + \int \varepsilon z d\delta_z = (1-\varepsilon)\mu + \varepsilon z$$

したがって

$$IC(z; F, T) = \lim_{\varepsilon\to 0} \frac{(1-\varepsilon)\mu + \varepsilon z - \mu}{\varepsilon} = z - \mu \tag{10}$$

となり，F の z における小さな変化は推定量の大きな変化を，そして z が μ から遠ければ遠いほど

図 9　$\bar{z}$ と s^2 の IC



z の影響は大きく，しかも限界 bound のない変化をひきおこすことがわかる（図9）。

$T_n=s^2$ のときには

$$T(F)=\sigma^2=\int(z-\mu)^2\,dF(z)$$

であるから

$$T[(1-\varepsilon)F+\varepsilon\delta_z]=\int(1-\varepsilon)(z-\mu)^2\,dF+\int\varepsilon\,(z-\mu)^2\,d\delta z$$
$$=(1-\varepsilon)\,\sigma^2+\varepsilon\,(z-\mu)^2$$

したがって

$$IC(z;F,T)=\lim_{\varepsilon\to 0}\frac{(1-\varepsilon)\,\sigma^2+\varepsilon\,(z-\mu)^2-\sigma^2}{\varepsilon}=(z-\mu)^2-\sigma^2 \tag{11}$$

となり，F の z における小さな変化は，やはり推定量の限界のない，大きな変化をひきおこすことがわかる（図9）。

次に線形回帰モデル

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u} \tag{12}$$

における $\boldsymbol{\beta}$ の最小2乗推定量 $\hat{\boldsymbol{\beta}}$ の影響関数を求めよう。ここで $\mathbf{y}$ は $n\times 1$ の被説明変数ベクトル，$\mathbf{X}$ は $n\times k$ の説明変数行列，$\boldsymbol{\beta}$ は $k\times 1$ のパラメータベクトル，$\mathbf{u}$ は $n\times 1$ の誤差項ベクトルである。

分布関数 F のもとで $\mathbf{x}$ を $1\times k$ のベクトル，Y を 1×1 のスカラーとし，

$$E_F(\mathbf{x}'\mathbf{x})=\textstyle\sum_{\mathbf{xx}}(F),\quad E_F(\mathbf{x}'Y)=\sum_{\mathbf{x}y}(F)$$

とする。$T(F_n)=\hat{\boldsymbol{\beta}}(F_n)$ とすれば

$$T(F)=\boldsymbol{\beta}(F)=\textstyle\sum_{\mathbf{xx}}^{-1}(F)\sum_{\mathbf{x}y}(F)$$

で与えられるから，$\boldsymbol{\beta}$ の IC は次式となる。

$$IC[\mathbf{x},\ Y;\boldsymbol{\beta}(F)]=\sum_{\mathbf{xx}}^{-1}(F)\,\mathbf{x}'\{Y-\mathbf{x}\,\boldsymbol{\beta}(F)\} \tag{13}$$

(13)式は次のようにして得られる。

$$\mathbf{z}=(\mathbf{x}\ \ Y)$$

とおく。このとき $\boldsymbol{\beta}(F)$ の影響関数は次式で定義される。

$$IC[\mathbf{z};F,\boldsymbol{\beta}(F)]=\lim_{\varepsilon\to 0}\frac{\boldsymbol{\beta}\{(1-\varepsilon)F+\varepsilon\delta_{\mathbf{z}}\}-\boldsymbol{\beta}(F)}{\varepsilon}$$

上式の分子を評価しよう。

$$\boldsymbol{\beta}\{(1-\varepsilon)F+\varepsilon\delta_z\}=\boldsymbol{\beta}(F_\varepsilon)=\sum_{\mathbf{xx}}^{-1}(F_\varepsilon)\sum_{\mathbf{x}y}(F_\varepsilon)$$

$$\sum_{\mathbf{xx}}(F_\varepsilon)=\sum_{\mathbf{xx}}\{(1-\varepsilon)F+\varepsilon\delta_{\mathbf{z}}\}$$

であり,

$$E_F(\mathbf{x}'\mathbf{x})=\int\mathbf{x}'\mathbf{x}\,dF=\sum_{\mathbf{xx}}(F)$$

であることに注意すれば次の結果を得る。

$$\begin{aligned}E_{F_\varepsilon}(\mathbf{x}'\mathbf{x})&=\int\mathbf{x}'\mathbf{x}\,dF_\varepsilon=\int\mathbf{x}'\mathbf{x}\,[(1-\varepsilon)\,dF+\varepsilon\,d\delta_{\mathbf{z}}]\\&=(1-\varepsilon)\int\mathbf{x}'\mathbf{x}\,dF+\varepsilon\int\mathbf{x}'\mathbf{x}\,d\delta_{\mathbf{z}}\\&=(1-\varepsilon)\sum_{\mathbf{xx}}(F)+\varepsilon\,\mathbf{x}'\mathbf{x}\end{aligned}$$

同様に,

$$E_F(\mathbf{x}'Y)=\int\mathbf{x}'Y\,dF=\sum_{\mathbf{x}y}(F)$$

であるから,

$$\begin{aligned}E_{F_\varepsilon}(\mathbf{x}'Y)&=\int\mathbf{x}'Y\,[(1-\varepsilon)\,dF+\varepsilon\,d\delta_{\mathbf{z}}]\\&=(1-\varepsilon)\int\mathbf{x}'\,Y\,dF+\varepsilon\int\mathbf{x}'Y\,d\delta_{\mathbf{z}}\\&=(1-\varepsilon)\sum_{\mathbf{x}y}(F)+\varepsilon\,\mathbf{x}'Y\end{aligned}$$

を得る。結局次の結果が得られた。

$$\sum_{\mathbf{xx}}(F_\varepsilon)=E_{F_\varepsilon}(\mathbf{x}'\mathbf{x})=(1-\varepsilon)\sum_{\mathbf{xx}}(F)+\varepsilon\,\mathbf{x}'\mathbf{x}$$

$$\sum_{\mathbf{x}y}(F_\varepsilon)=E_{F_\varepsilon}(\mathbf{x}'Y)=(1-\varepsilon)\sum_{\mathbf{x}y}(F)+\varepsilon\,\mathbf{x}'Y$$

したがって

$$\sum_{\mathbf{xx}}^{-1}(F_\varepsilon)=\frac{1}{1-\varepsilon}\left\{\sum_{\mathbf{xx}}(F)+\frac{\varepsilon}{1-\varepsilon}\mathbf{x}'\mathbf{x}\right\}^{-1}$$

となる。さらに

$$\left\{\sum_{\mathbf{xx}}(F)+\frac{\varepsilon}{1-\varepsilon}\mathbf{x}'\mathbf{x}\right\}^{-1}$$

$$=\sum_{\mathbf{xx}}^{-1}(F)-\frac{\varepsilon}{1-\varepsilon}\sum_{\mathbf{xx}}^{-1}(F)\,\mathbf{x}'\left\{1+\frac{\varepsilon}{1-\varepsilon}\,\mathbf{x}\sum_{\mathbf{xx}}^{-1}(F)\,\mathbf{x}'\right\}^{-1}\mathbf{x}\sum_{\mathbf{xx}}^{-1}(F)$$

と表すことができるから次式を得る。

$$\sum_{\mathbf{xx}}^{-1}(F_\varepsilon)\sum_{\mathbf{x}y}(F_\varepsilon)$$

$$=\frac{1}{1-\varepsilon}\Big[\sum_{\mathbf{xx}}^{-1}(F)-\frac{\varepsilon}{1-\varepsilon}\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'\Big\{1+\frac{\varepsilon}{1-\varepsilon}\mathbf{x}\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'\Big\}^{-1}\mathbf{x}\sum_{\mathbf{xx}}^{-1}(F)$$

$$\cdot[(1-\varepsilon)\sum_{\mathbf{x}y}(F)+\varepsilon\,\mathbf{x}'Y]$$

$$=\boldsymbol{\beta}(F)+\frac{\varepsilon}{1-\varepsilon}\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'Y$$

$$-\frac{\varepsilon}{1-\varepsilon}\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'\Big\{1+\frac{\varepsilon}{1-\varepsilon}\mathbf{x}\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'\Big\}^{-1}\mathbf{x}\,\boldsymbol{\beta}(F)$$

$$-\Big(\frac{\varepsilon}{1-\varepsilon}\Big)^2\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'\Big\{1+\frac{\varepsilon}{1-\varepsilon}\mathbf{x}\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'\Big\}^{-1}\mathbf{x}\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'Y$$

そして

$$\lim_{\varepsilon\to 0}\Big\{1+\frac{\varepsilon}{1-\varepsilon}\mathbf{x}\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'\Big\}^{-1}=1$$

に注意すれば

$$IC=\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'Y-\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'\mathbf{x}\,\boldsymbol{\beta}(F)$$

$$=\sum_{\mathbf{xx}}^{-1}(F)\mathbf{x}'[Y-\mathbf{x}\,\boldsymbol{\beta}(F)]$$

となる。

$Y-\mathbf{x}\boldsymbol{\beta}(F)$ に対応する標本概念は $Y-\mathbf{x}\hat{\boldsymbol{\beta}}=e$ である。このことから $\mathbf{z}=(\mathbf{x}\ Y)$ の $\hat{\boldsymbol{\beta}}$ への影響は，$\bar{z}$ や s^2 の場合と同じように，限界がないことがわかる。いいかえれば F の $\mathbf{z}$ における小さな変化の $\hat{\boldsymbol{\beta}}$ への影響は誤差項 $Y-\mathbf{x}\,\boldsymbol{\beta}(F)$（標本概念では残差 e）に比例的に現れ，しかも限界がない。

5　影響関数の有限標本近似

前節の議論は無限標本に適用された観測値の推定量への影響であった。しかし有限標本で分析せざるを得ないわれわれは，無限標本の世界で展開された $\mathbf{z}=(\mathbf{x}\ Y)$ を $(\mathbf{x}_i\ Y_i)$，$i=1, 2, \cdots\cdots, n$ で，推定値の漸近値 $T(F)$ を推定量 $T(F_n)$ でおきかえなければならない。われわれの関心は，具体的な $(\mathbf{x}_i\ Y_i)$ の値が，パラメータ推定量にどのような影響を与えるかにあるからである。

n 個の観測値 $(\mathbf{x}_i\ Y_i)$，$i=1, 2, \cdots\cdots, n$ にもとづいて IC を近似するとき次の3通りを考えることができる。

1　経験的影響関数あるいは曲線 empirical influence function　　*EIC*

2　標本影響関数あるいは曲線 sample influence function　　*SIC*

3　感度曲線 sensitivity curve　　*SC*

1　経験的影響関数 *EIC*

IC の $(\mathbf{x}\ Y)$ を $(\mathbf{x}_i\ Y_i)$，$\sum_{\mathbf{xx}}(F)$ を $\frac{1}{n}(\mathbf{X}'\mathbf{X})$，$\boldsymbol{\beta}(F)$ を $\hat{\boldsymbol{\beta}}(F_n)$ でおきかえることによ

って EIC が得られる。すなわち

$$EIC_i = n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'(Y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}) = n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'e_i \tag{14}$$

$$i=1,\ 2,\ \cdots\cdots,\ n$$

この EIC_i は，F_n に微小のゆらぎ perturbation があったときの説明変数の第 i 期のベクトル $\mathbf{x}_i$ における $\hat{\boldsymbol{\beta}}$ への効果を表している。

パラメータ推定において，第 i 期の観測値（$\mathbf{x}_i$ Y_i）を除いたときの EIC を $EIC(i)$ とすれば，IC の

$(\mathbf{x}\ Y)$ を $(\mathbf{x}_i\ Y_i)$ で

$\sum_{\mathbf{xx}}(F)$ を $\dfrac{1}{n-1}\mathbf{X}'(i)\mathbf{X}(i)$ で

$\boldsymbol{\beta}(F)$ を $\hat{\boldsymbol{\beta}}(i)(F_{n-1})$ で

おきかえて次の結果を得る。

$$EIC(i) = (n-1)[\mathbf{X}'(i)\mathbf{X}(i)]^{-1}\mathbf{x}_i'[Y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}(i)] \tag{15}$$

ここで

$\mathbf{X}(i) = \mathbf{X}(n\times k)$ から第 i 期のベクトル $\mathbf{x}_i$ を除いた $(n-1)\times k$ の説明変数行列

$\hat{\boldsymbol{\beta}}(i)$＝第 i 期の観測値を除いて推定される $\boldsymbol{\beta}$ の推定量

さらに

$$Y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}(i) = e_i(i) = \frac{e_i}{1-h_{ii}}$$

$$[\mathbf{X}'(i)\mathbf{X}(i)]^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}}{1-h_{ii}}$$

であるから

$$\begin{aligned}
&[\mathbf{X}'(i)\mathbf{X}(i)]^{-1}\mathbf{x}_i'e_i(i)\\
&= \left\{(\mathbf{X}'\mathbf{X})^{-1} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}}{1-h_{ii}}\right\}\mathbf{x}_i'\frac{e_i}{1-h_{ii}}\\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\frac{e_i}{1-h_{ii}} + \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'e_i}{(1-h_{ii})^2}\\
&= \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'e_i}{(1-h_{ii})^2}(1-h_{ii}+h_{ii})\\
&= \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'e_i}{(1-h_{ii})^2}
\end{aligned}$$

したがって $EIC(i)$ は次のように表すこともできる。

$$EIC(i) = (n-1)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\frac{e_i}{(1-h_{ii})^2} \tag{16}$$

2　標本影響関数 SIC

EIC は F_n あるいは F_{n-1} を無限標本から得ることができるという前提のもとで導かれた。n と

の関連を明示的に表す標本影響関数は，IC の定義において

z を $\mathbf{z}_i=(\mathbf{x}_i\ Y_i)$ でおきかえ

F を F_n でおきかえ

$\varepsilon=-\dfrac{1}{n-1}$ とおき

極限の演算 $\lim_{\varepsilon\to 0}$ を外すことによって得られる。このとき，$F_\varepsilon=(1-\varepsilon)F_n+\varepsilon\delta_{z_i}$ は，$\mathbf{z}_i$ が除外されたときの F_n の微小な変化を示し，標本の世界では F_{n-1} に対応していると考えることができる。すなわち

$$\begin{aligned}F_\varepsilon&=\left(1-\frac{1}{n-1}\right)F_n-\frac{1}{n-1}\delta_{z_i}\\&=\frac{n}{n-1}F_n-\frac{1}{n-1}\delta_{z_i}\\&\fallingdotseq F_{n-1}\end{aligned}$$

である。

したがって

$$\begin{aligned}SIC_i&=-(n-1)\left\{T\left(\frac{n}{n-1}F_n-\frac{1}{n-1}\delta_{z_i}\right)-T(F_n)\right\}\\&=-(n-1)\{T(F_{n-1})-T(F_n)\}\\&=(n-1)\{T(F_n)-T(F_{n-1})\}\end{aligned}\tag{17}$$

この式において

$$T(F_n)=\hat{\boldsymbol{\beta}}$$
$$T(F_{n-1})=\hat{\boldsymbol{\beta}}(i)$$

とおくことによって

$$\begin{aligned}SIC_i&=(n-1)[\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}(i)]\\&=(n-1)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\frac{e_i}{1-h_{ii}}\end{aligned}\tag{18}$$

と表すことができる。(18)式は(16)式とほとんど同じである。

この SIC_i はパラメータ推定に用いられるジャックナイフ法 jackknife method（次節参照）と関連をもっている。擬似ジャックナイフ値 jackknifed pseudovalue を

$$T^*_{ni}=nT_n(z_1,\ \cdots\cdots,\ z_n)-(n-1)\,T_{n-1}(z_1,\ \cdots\cdots,\ z_{i-1},\ z_{i+1},\ \cdots\cdots,\ z_n)\tag{19}$$

とする。たとえば $T_n=\bar{z}$ のとき $T_{n-1}=\dfrac{1}{n-1}\sum_{j\neq i}z_j$ であるから $T^*_{ni}=z_i$ である。

$IC(z;\ F,\ T)$ において，$z=z_i,\ F=F_n,\ \varepsilon=-\dfrac{1}{n-1}$ とおくと

$$\frac{T[(1-\varepsilon)F_n+\varepsilon\delta_{z_i}]-T(F_n)}{\varepsilon}$$

$$=\frac{T\left[\left(1+\frac{1}{n-1}\right)F_n-\frac{1}{n-1}\delta_{z_i}\right]-T(F_n)}{\left(-\frac{1}{n-1}\right)}$$

$$=(n-1)[T(F_n)-T(F_{n-1})]$$

$$=(n-1)[T_n(z_1,\cdots\cdots,z_n)-T_{n-1}(z_1,\cdots\cdots,z_{i-1},z_{i+1},\cdots\cdots,z_n)]$$

$$=nT_n(z_1,\cdots\cdots,z_n)-(n-1)T_{n-1}(z_1,\cdots\cdots,z_{i-1},z_{i+1},\cdots\cdots,z_n)-T_n(z_1,\cdots\cdots,z_n)$$

$$=T^*_{ni}-T_n(z_1,\cdots\cdots,z_n)$$

極限 $\varepsilon\to 0$ を除けば，この結果は擬似ジャックナイフ値 T^*_{ni} から T_n を引いた値である。いいかえれば $T^*_{ni}-T_n$ は SIC_i の有限標本近似である。前述の $T_n=\bar{z}$ の例でいえば

$$T^*_{ni}-T_n=z_i-\bar{z}$$

であり，これは(10)式で示される IC の標本対応にほかならない。

3 感度曲線 SC

IC への有限標本近似は Tukey の感度曲線によっても与えられる。

$n-1$ 個の標本 $(z_1, z_2, \cdots\cdots, z_{n-1})$ から出発すると，感度曲線は次式で示すことができる。

$$SC_i=n[T_n(z_1,\cdots\cdots,z_{n-1},z_n)-T_{n-1}(z_1,\cdots\cdots,z_{n-1})] \tag{20}$$

この SC_i は，IC において $\mathbf{z}$ に $\mathbf{z}_i=(\mathbf{x}_i\ Y_i)$, $F=F_{n-1}$, $\varepsilon=1/n$ とおき，極限を別にすれば

$$\frac{T[(1-\varepsilon)F_{n-1}+\varepsilon\delta_{z_i}]-T(F_{n-1})}{\varepsilon}$$

$$=\frac{T\left[\left(1-\frac{1}{n}\right)F_{n-1}+\frac{1}{n}\delta_{z_i}\right]-T(F_{n-1})}{\left(\frac{1}{n}\right)}$$

$$=n[T(F_n)-T(F_{n-1})]$$

に等しい。

$$\left(1-\frac{1}{n}\right)F_{n-1}+\frac{1}{n}\delta_{z_i}$$

は F_n の近似と考えることができるということを上式で用いた。$n\to\infty$ のとき SC_i が IC に収束することは明らかであろう。

$T(F_n)=\hat{\boldsymbol{\beta}}$, $T(F_{n-1})=\hat{\boldsymbol{\beta}}(i)$ とおけば

$$SC_i=n[\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}(i)]=n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\frac{e_i}{1-h_{ii}} \tag{21}$$

となる。

以上の結果をまとめると次の通りである。

$$EICi=n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'e_i$$

$$EIC(i)=(n-1)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\frac{e_i}{(1-h_{ii})^2}$$

$$SIC_i=(n-1)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\frac{e_i}{1-h_{ii}}$$

$$SC_i=n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\frac{e_i}{1-h_{ii}}$$

上式から，高い作用点 high leverage point（h_{ii} の大きい値）に一番敏感な尺度が $EIC(i)$，一番反応の小さいのが EIC_i であることがわかる。SIC_i と SC_i はほとんど同じであり，両者とも $\hat{\boldsymbol{\beta}}$ と $\hat{\boldsymbol{\beta}}(i)$ の距離 $\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}(i)=(\mathbf{X}'\mathbf{X})^{-1}\ \mathbf{x}_i'\ e_i/(1-h_{ii})$ に比例して（$\mathbf{x}_i\ Y_i$）のパラメータ推定量への影響力は変化する。

6　影響関数にもとづく影響力の尺度

$\hat{\boldsymbol{\beta}}$ の SIC_i は $k\times1$ のベクトルであるから，ベクトル SIC_i, $i=1$, ……, n を影響力の強さによって順序づけることはできない。そこで k 個のベクトルの要素を用いて，第 i 期の観測値（$\mathbf{x}_i\ Y_i$）の $\hat{\boldsymbol{\beta}}$ への影響力を示すスカラーの尺度を作ることを考えよう。影響力を表すために $k\times1$ ベクトルの規準化を行ない，この規準化された尺度にもとづいて，影響力の強さを順序化することを考える。

$$\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}(i)=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'\frac{e_i}{1-h_{ii}}$$

であるから，

$$(n-1)^{-1}SICi=\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}(i)$$

と表すことができる。この $\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}(i)$ を $k\times k$ の正値半定符号行列 $\mathbf{M}$ と，正の尺度要素 scale factor c を用いて規準化し，それを

$$D_i(\mathbf{M},\ c)=\frac{[\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}(i)]'\mathbf{M}[\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}(i)]}{c} \tag{22}$$

を定義すれば，スカラーの影響力を示す尺度となる。SIC_i を用いれば，

$$D_i(\mathbf{M},\ c)=\frac{(SIC_i)'\mathbf{M}(SIC_i)}{c(n-1)^2} \tag{23}$$

と表すことができる。

同様にして，SC_i, EIC_i, $EIC(i)$ を用いて $D_i(\mathbf{M},\ c)$ を表せば次のようになる。

$$D_i(\mathbf{M},\ c)=\frac{(SC_i)'\mathbf{M}(SC_i)}{cn^2} \tag{24}$$

$$D_i(\mathbf{M},\ c)=\frac{(EIC_i)'\mathbf{M}(EIC_i)}{cn^2(1-h_{ii})^2} \tag{25}$$

$$D_i(\mathbf{M},\ c)=\frac{(1-h_{ii})^2(EIC(i)'\mathbf{M}(EIC(i))}{c(n-1)^2} \tag{26}$$

回帰診断で用いられる Cook の D は(22)式において $\mathbf{M}=\mathbf{X}'\mathbf{X}$, $c=ks^2$ の場合であり，その他，DFFITS は$\mathbf{M}=\mathbf{X}'\mathbf{X}$, $c=s^2(i)$, Welsch の距離は $\mathbf{M}=\mathbf{X}'\mathbf{X}$, $c=(1-h_{ii})s^2(i)/(n-1)$, Atkinson の修正された Cook の D は $\mathbf{M}=\mathbf{X}'\mathbf{X}$, $c=\left(\frac{k}{n-k}\right)s^2(i)$ のことに他ならない。

7　ジャックナイフ推定量

5節で SIC_i と擬似ジャックナイフ値との関連を述べたが，ここでジャックナイフ推定量について説明しよう。

パラメータ推定において不偏推定量がつねに得られるならば，偏り補正の問題は生じない。しかし最尤推定量が偏りをもつ場合があるように，偏りのある推定量を得ることはしばしばある。しかもその偏りを正確に評価できず，したがって補正できないという場合がある。偏りのある推定量の偏りを0にする，あるいはより小さくする有力な方法がジャックナイフ法である。

“ジャックナイフ”という名称は，ボーイスカウトがもっているジャックナイフがどのような用途にも使用できる頼りになる道具であるように，どのような状況にも適用できる方法という意味から付けられた名称である。ジャックナイフ推定量の概念は Quenouille によって 1949年の論文 “Approximate tests of correlation in time series”, *Journal of the Royal Statistical Society, B,* において導入された。名称は1958年 Tukey の命名による（“Bias and confidence in not-quite large samples”, *Annals of Mathematical Statistics,* 29, 1958)。

ジャックナイフ法の目的は推定量の偏りを小さくし，パラメータの近似的な信頼区間を与えることにある。

未知パラメータ θ の推定量 $T_n(X_1, \cdots\cdots, X_n)$ の期待値および分散が次のように展開できると仮定しよう。

$$E(T_n)=\theta+\frac{a_1(\theta)}{n}+\frac{a_2(\theta)}{n^2}+\cdots\cdots$$

$$Var(T_n)=\frac{\sigma_1^2(\theta)}{n}+\frac{\sigma_2^2(\theta)}{n^2}+\cdots\cdots$$

たとえば $\hat{\sigma}^2=\sum(X_i-\bar{X})^2/n$ は

$$E(\hat{\sigma}^2)=\left(\frac{n-1}{n}\right)\sigma^2=\sigma^2+\left(-\frac{\sigma^2}{n}\right)$$

となるから，偏りは $1/n$ の次数をもつ推定量である。

X_i を除いて得られる θ の推定量を

$$T_{n-1,(i)}=T_{n-1}(X_1, \cdots\cdots, X_{i-1}, X_{i+1}, \cdots\cdots, X_n)$$

とすれば，θ のジャックナイフ推定量は

$$J(T_n)=nT_n-(n-1)\,\bar{T}_{n-1,(i)} \tag{27}$$

によって与えられる。ここで

$$\bar{T}_{n-1,(i)}=\frac{1}{n}\sum_{i=1}^{n}T_{n-1,(i)} \tag{28}$$

である。この $J(T_n)$ は

$$J(T_n)=T_n+(n-1)(T_n-\bar{T}_{n-1,(i)}) \tag{29}$$

と表すこともできるから，ジャックナイフ推定量は全標本データから計算される推定値 T_n を，T_n と $\bar{T}_{n-1}$ の差に $(n-1)$ をかけた項だけ修正した推定値になっている。また i 番目の擬似ジャックナイフ値といわれる値を

$$T^*_{ni}=nT_n-(n-1)\,T_{n-1,(i)} \tag{30}$$

で表せば，

$$J(T_n)=T^*_n \tag{31}$$

である。ここで

$$T^*_n=\frac{1}{n}\sum_{i=1}^{n} T^*_{ni}$$

この $J(T_n)$ が T_n の偏りを補正するのは

$$I=(n-1)(T_n-\bar{T}_{n-1,(i)})$$

とすれば

$$E(T_{n-1,(i)})=\theta+\frac{a_1(\theta)}{n-1}+\cdots\cdots$$

であるから

$$\begin{aligned}E(I)&=(n-1)[E(T_n)-E(\bar{T}_{n-1,(i)})]\\&=(n-1)\left\{\theta+\frac{a_1(\theta)}{n}+\cdots\cdots-\left[\theta+\frac{a_1(\theta)}{n-1}+\cdots\cdots\right]\right\}\\&=(n-1)\left[\left(\frac{1}{n}-\frac{1}{n-1}\right)a_1(\theta)+\cdots\cdots\right]\\&=-\frac{a_1(\theta)}{n}+O(n^{-2})\end{aligned}$$

ここで上式の計算において

$$\begin{aligned}E(\bar{T}_{n-1,(i)})&=\frac{1}{n}\sum_{i=1}^{n}E(T_{n-1,(i)})\\&=\frac{1}{n}\sum_{i=1}^{n}\left[\theta+\frac{a_1(\theta)}{n-1}+\cdots\cdots\right]\\&=\theta+\frac{a_1(\theta)}{n-1}+\cdots\cdots\end{aligned}$$

を用いた。

したがって

$$\begin{aligned}E[J(T_n)]&=E(T_n)+E(I)\\&=\theta+\frac{a_1(\theta)}{n}+\cdots\cdots-\frac{a_1(\theta)}{n}+O(n^{-2})\\&=\theta+O(n^{-2})\end{aligned}$$

となり，偏り $\dfrac{a_1(\theta)}{n}$ は消え，$O(n^{-2})$ の次数の偏りのみが残るからである。

例 1

$X_1, X_2, \cdots\cdots, X_n$ は期待値 θ をもつ分布からの無作為標本であるとしよう。このとき

$$T_n=\bar{X}$$

とすれば,

$$T_{n-1,(i)}=\frac{1}{n-1}\sum_{j\neq i}X_j=\frac{1}{n-1}\left(\sum_{j=1}^{n}X_j-X_i\right)$$

$$=\frac{1}{n-1}(n\bar{X}-X_i)=\frac{n}{n-1}\bar{X}-\frac{1}{n-1}X_i$$

したがって

$$\bar{T}_{n-1,(i)}=\frac{n}{n-1}\bar{X}-\frac{\bar{X}}{n-1}=\bar{X}$$

ゆえに

$$J(T_n)=T_n+(n-1)(T_n-\bar{T}_{n-1,(i)})=\bar{X}+(n-1)(\bar{X}-\bar{X})=\bar{X}$$

例 2

分散 σ^2 の推定量を

$$T_n=\frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X})^2$$

とすると

$$T_{n-1,(i)}=\frac{1}{n-1}\sum_{j\neq i}(X_j-\bar{X}_{(i)})^2$$

そして

$$X_j-\bar{X}_{(i)}=X_j-\bar{X}+\bar{X}-\bar{X}_{(i)}$$

$$\bar{X}-\bar{X}_{(i)}=\frac{1}{n}\sum_{j}X_j-\frac{1}{n-1}\sum_{j\neq i}X_j$$

$$=\frac{1}{n}\sum_{j}X_j-\frac{1}{n-1}(\sum_{j}X_j-X_i)$$

$$=\frac{-1}{n(n-1)}\sum_{j}X_j+\frac{1}{n-1}X_i$$

$$=\frac{1}{n-1}(X_i-\bar{X})$$

$$X_i-\bar{X}_{(i)}=X_i-\frac{1}{n-1}\sum_{j\neq i}X_j=X_i-\frac{1}{n-1}(\sum_{j}X_j-X_i)=\frac{n}{n-1}(X_i-\bar{X})$$

ゆえに

$$\sum_{j\neq i}(\bar{X}_j-\bar{X}_{(i)})^2$$

$$=\sum_{j=1}^{n}(X_j-\bar{X}+\bar{X}-\bar{X}_{(i)})^2-(X_i-\bar{X}_{(i)})^2$$

$$=\sum_{j=1}^{n}(X_j-\bar{X})^2+2(\bar{X}-\bar{X}_{(i)})\sum_{j=1}^{n}(X_j-\bar{X})+\frac{n}{(n-1)^2}(X_i-\bar{X})^2-(X_i-\bar{X}_{(i)})^2$$

$$=\sum_{j=1}^{n}(X_j-\bar{X})^2+\frac{n}{(n-1)^2}(X_i-\bar{X})^2-\frac{n^2}{(n-1)^2}(X_i-\bar{X})^2$$

$$=\sum_{j=1}^{n}(X_j-\bar{X})^2-\frac{n}{n-1}(X_i-\bar{X})^2$$

したがって

$$T^*_{ni}=nT_n-(n-1)\,T_{n-1,(i)}$$

$$=\sum_{j=1}^{n}(X_j-\bar{X})^2-\sum_{j=1}^{n}(X_j-\bar{X})^2+\frac{n}{n-1}(X_i-\bar{X})^2$$

$$=\frac{n}{n-1}(X_i-\bar{X})^2$$

結局 σ^2 のジャックナイフ推定量として

$$J(T_n)=T^*_n=\frac{1}{n}\sum_{i=1}^{n}T^*_{ni}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$$

が得られる。σ^2 の推定量 T_n は $-\frac{\sigma^2}{n}$ の偏りをもち，$O(n^{-2})$ の偏りはないから，ジャックナイフ化することによって，次数 $n-1$ の偏りは除去され，$J(T_n)$ は σ^2 の不偏推定量となる。

次にジャックナイフ推定量 $J(T_n)$ を用いて，未知パラメータ θ に関する検定問題を考えよう。T_n, $T_{n-1,i}$ がスカラーの場合を考える。

$$T^*_{ni}=nT_n-(n-1)\,T_{n-1,(i)}$$

$$J(T_n)=T^*_n=\frac{1}{n}\sum_{i=1}^{n}T^*_{ni}=nT_n-(n-1)\,\bar{T}_{n-1,(i)}$$

であった。

Tukey は，T^*_{ni} は近似的に独立で同一の分布に従う確率変数としてあつかうことができるので

$$t=\frac{T^*_n-\theta}{\sqrt{\dfrac{\sum_{i=1}^{n}(T^*_{ni}-T^*_n)^2}{n(n-1)}}} \tag{32}$$

は近似的に自由度 $n-1$ の t 分布に従う。したがって大標本においては標準正規分布で近似できると述べた。この主張が正しいことは Miller〔23〕によって証明された。したがって θ のジャックナイフ推定量 T^*_n を用いて，t 分布あるいは正規分布による θ に関する区間推定あるいは仮説検定をすることができる。

例 3

線形回帰モデルの回帰係数 β_j のジャックナイフ推定量を求めよう。

$$T_n=\hat{\beta}_j$$

$$T_{n-1,(i)}=\hat{\beta}_j(i)$$

とすれば

$$T^*_{ni}=n\hat{\beta}_j-(n-1)\,\hat{\beta}_j(i)$$

$$=\hat{\beta}_j+(n-1)[\hat{\beta}_j-\hat{\beta}_j(i)]$$

であるから，β_j のジャックナイフ推定量は

$$T^*_n=\hat{\beta}_j+(n-1)[\hat{\beta}_j-\overline{\hat{\beta}_j}(i)] \tag{33}$$

によって与えられる。ここで

$$\overline{\hat{\beta}}_j(i)=\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_j(i)$$

である。β_j の OLSE $\hat{\beta}_j$ と T^*_{ni} の相違は

$$(n-1)[\hat{\beta}_j-\hat{\beta}_j(i)]$$

であり，これは感度曲線の値であった。

具体例によって，ジャックナイフ推定値がどの程度 OLSE と異なってくるかをみてみよう。古典的な線形回帰モデルの仮定のもとで，定式化が正しければ OLSE は最良線形不偏推定量を与えるから，ジャックナイフ推定法を適用すべき理由はない。しかし決定係数が余り大きくなく，現在の定式化に含まれていない別の系統的要因があり，定式化の誤まり misspecification の恐れがあるのではないかと思われる場合がある。しかし理論的にどのような系統的要因をモデルに含めるべきかを明らかにすることはできず，説明力を高めようとすれば，ad hoc な説明変数を加えることになるという場合が実証分析でしばしば生ずる。系統的要因が抜けているという意味の定式化の誤まりがあれば，$\boldsymbol{\beta}$ の OLSE $\hat{\boldsymbol{\beta}}$ はもはや不偏性をもたない。これまでジャックナイフ法が計量経済学で使われたことはないが，前述の状況のことを考えれば，計量経済学の分野においても決して無意味ではない。

例を次のイギリスの賃金率関数にとろう。

表 2　イギリスのデータ

	RUUK1 (−1) (%)	CPIDUK1 (%)	WDOTUK1 (%)
1965	+1.8000000D+00	+4.8000000D+00	+1.0050000D+01
1966	+1.5000000D+00	+3.9000000D+00	+5.4800000D+00
1967	+1.5000000D+00	+2.5000000D+00	+4.5500000D+00
1968	+2.3000000D+00	+4.7000000D+00	+6.8300000D+00
1969	+2.5000000D+00	+5.4000000D+00	+8.3300000D+00
1970	+2.4000000D+00	+6.4000000D+00	+1.5210000D+01
1971	+2.6000000D+00	+9.4000000D+00	+1.1800000D+01
1972	+3.5000000D+00	+7.1000000D+00	+1.4030000D+01
1973	+3.8000000D+00	+9.1000000D+00	+1.4490000D+01
1974	+2.7000000D+00	+1.6000000D+01	+2.1800000D+01
1975	+2.6000000D+00	+2.4200000D+01	+2.6540000D+01
1976	+4.1000000D+00	+1.6500000D+01	+1.2680000D+01
1977	+5.6000000D+00	+1.5800000D+01	+8.2300000D+00
1978	+5.7000000D+00	+8.3000000D+00	+1.4870000D+01
1979	+5.7000000D+00	+1.3400000D+01	+1.7700000D+01
1980	+5.3000000D+00	+1.8000000D+01	+1.7510000D+01
1981	+6.8000000D+00	+1.1900000D+01	+1.0100000D+01
1982	+9.4000000D+00	+8.6100000D+00	+8.9900000D+00
1983	+1.0900000D+01	+4.5900000D+00	+7.2500000D+00
1984	+1.1700000D+01	+4.9800000D+00	+7.0700000D+00
1985	+1.1700000D+01	+6.0800000D+00	+7.9800000D+00
1986	+1.1800000D+01	+3.4000000D+00	+6.9200000D+00
1987	+1.1800000D+01	+4.1700000D+00	+6.7300000D+00

$$WDOTUK1=\beta_0+\beta_1(RUUK1-RUUK1_{-1})_{-1}$$
$$+\beta_2\ CPIDUK1+\varepsilon \qquad (34)$$

ここで

$WDOTUK1$＝貨幣賃金率変化率（%）

$RUUK1$＝失業率（%）

$CPIDUK1$＝消費者物価変化率（%）

データは表2に示されている。1965～1987年のデータを用いた OLS による推定結果は次の通りである。

	COEFFICIENT	SD	T-VALUE	有意確率	VIF
定数項	+5.3885050D+00	+1.1937709D+00	+4.514	0.00021	
B（1）	−2.2114942D+00	+7.2375551D−01	−3.056	0.00624	1.000
B（2）	+7.7423137D−01	+1.0813726D−01	+7.160	0.00000	1.000

$\bar{R}^2=0.728$, $s=2.90$（%）, $DW=1.686$

表 3　回帰係数の変化とジャックナイフ推定値

***** 回帰係数の変化　BETAj-BETAj (i)（DFBETAS）*****

	定数項	B（1）	B（2）
1	−0.0548(−0.045)	0.0278(0.038)	0.0027(0.024)
2	−0.5182(−0.443)	0.1805(0.254)	0.0295(0.278)
3	−0.4322(−0.362)	0.0814(0.113)	0.0289(0.267)
4	−0.0414(−0.034)	−0.0112(−0.015)	0.0028(0.026)
5	−0.0818(−0.067)	0.0113(0.015)	0.0044(0.040)
6	0.4546(0.400)	−0.1594(−0.231)	−0.0189(−0.184)
7	−0.0200(−0.016)	0.0058(0.008)	−0.0002(−0.002)
8	0.3070(0.275)	0.1672(0.247)	−0.0150(−0.149)
9	0.1317(0.110)	−0.0193(−0.027)	−0.0000(−0.000)
10	−0.0099(−0.008)	−0.1985(−0.270)	0.0202(0.183)
11	−0.4744(−0.397)	−0.1076(−0.148)	0.0738(0.681)
12	0.2127(0.177)	−0.1846(−0.253)	−0.0280(−0.256)
13	0.5205(0.497)	−0.5086(−0.801)	−0.0698(−0.736)
14	0.2128(0.180)	−0.0666(−0.093)	−0.0039(−0.036)
15	−0.0009(−0.001)	−0.0529(−0.072)	0.0125(0.114)
16	0.1619(0.136)	0.1666(0.231)	−0.0413(−0.382)
17	0.0281(0.023)	−0.0928(−0.126)	−0.0054(−0.049)
18	−0.0313(−0.026)	0.5584(0.778)	−0.0021(−0.020)
19	0.1372(0.113)	0.1287(0.175)	−0.0118(−0.107)
20	−0.0375(−0.031)	−0.0106(−0.014)	0.0025(0.023)
21	−0.2096(−0.174)	0.0583(0.080)	0.0096(0.088)
22	−0.1204(−0.099)	0.0191(0.026)	0.0077(0.070)
23	−0.2415(−0.199)	0.0535(0.073)	0.0143(0.130)

*** Jackknife 推定値 ***

	COEFFICIENT	SD	T-VALUE
定数項	+5.2857346D+00	+1.2122947D+00	+4.360
B（1）	−2.1671249D+00	+8.8367414D−01	−2.452
B（2）	+7.8607237D−01	+1.2461690D−01	+6.308

推定されたパラメータの符号条件，統計的有意性，多重共線性，1階の自己相関いずれも全く問題ないが，説明力は $\bar{R}^2=0.728$ と余り大きくない。系統的要因が欠除しているのではないか，したがってOLSE の不偏性は成立していないのではないかという疑いがもたれる。そこでジャックナイフ法を用いてみよう。

$$\hat{\beta}_j-\hat{\beta}_j(i),\quad j=0,\ 1,\ 2,\quad i=1,\ \cdots\cdots,\ 23$$

と DFBETAS は表3に示されている。この表を用いてジャックナイフ推定値

$$\hat{\beta}_j+(n-1)[\hat{\beta}_j-\overline{\hat{\beta}_j(i)}],\quad j=0,\ 1,\ 2$$

と，標準偏差

$$\left\{\frac{\sum_{i=1}^{n}(T^*_{ni}-T^*_n)^2}{n(n-1)}\right\}^{\frac{1}{2}} \tag{35}$$

が計算された。ジャックナイフ推定値は表3の下の方の COEFFICIENT，標準偏差は SD に示されている。

労働市場の需給ギャップを示す失業率の変化に対する係数も，消費者物価変化率に対する係数も，OLS とジャックナイフ法の間にそれほど大きな違いはない。そしてジャックナイフ推定値も有意である。しかし前述したように，系統的要因が足りないかも知れないということを考えると，ジャックナイフ推定値を採用した方が良い。

ジャックナイフ法によって次数 $1/n^2$，さらに $1/n^3$，$1/n^4$…… の偏りを除去するための一般化ジャックナイフ法については Miller〔22〕に詳しい。またこの論文は1974年までのジャックナイフ法に関する参考文献を網羅している。

ジャックナイフ推定量 T^*_n は，擬似ジャックナイフ値 T^*_{ni} の単純平均であり，一般的には T^*_{ni} は独立ではなく，長い尾をもつ分布に従っている。単純平均は(10)式で示されているように，外れ値に対して頑健でないから，頑健なジャックナイフ推定量として削除平均 trimmed mean と同じ概念，削除ジャックナイフ trimmed jackknife を Hinkley and Wang〔14〕は提唱している。

計量経済学のパラメータ推定にジャックナイフ法が使われた例は皆無であるが，不偏推定量の世界にのみ安住していることができないことが多い以上，今後，真剣に検討されるべき方法であろう。

8　M 推 定 量

線形回帰モデルの誤差項の分布は正規分布より長い尾をもつ分布に従うことが多いこと，したがって正規分布のもとでは「外れ値」とみなされる極端な値も，より長い尾をもつ分布のもとでは決して「外れ値」ではないこと，しかし，通常の最小2乗法 OLS はこの「外れ値」に対して敏感に反応すること，したがって「外れ値」に対して頑健なパラメータ推定が望ましいということをこれまでに述べた。

誤差項の分布が正規分布より長い尾をもつ分布であるとき，頑健推定 robust estimation は適切な方法である。良い頑健推定法とは長い尾をもつ分布に対して有効であり，もし誤差項の分布が長い尾をもつ分布ではなく正規分布であったとしても，推定量の有効性の損失は少なく，かつ外れ値に対して抵抗力をもつようなパラメータ推定法でなければならない。

回帰の頑健法はM推定量，L推定量およびR推定量の3つに大別できるが，このなかで計量経済学の回帰モデルに対しても容易に適用可能であり，望ましい特性を有しているのはM推定量である。

$X_1, X_2, \cdots\cdots, X_n$ は確率密度関数 $f(x;\theta)$ からの無作為標本であるとしよう。θ は未知パラメータである。対数尤度関数 $L(\theta)$ を次のように表わそう。

$$\log L(\theta)=\sum_{i=1}^{n}\log f(x_i;\theta)=-\sum_{i=1}^{n}\rho(x_i;\theta)$$

ここで

$$\rho(x)=-\log f(x)$$

である。このとき θ の最尤推定量は

$$\frac{d\log L(\theta)}{d\theta}=\sum_{i=1}^{n}\frac{f'(x_i;\theta)}{f(x_i;\theta)}=-\sum_{i=1}^{n}\rho'(x_i;\theta)$$

$$=-\sum_{i=1}^{n}\psi(x_i;\theta)=0$$

の解である。ここで

$$f'=\frac{df}{d\theta},\quad \rho'=\frac{d\rho}{d\theta}=-\frac{f'}{f},\quad \psi(x)=\rho'(x)$$

である。

たとえば f は $\sigma=1$ で未知パラメータ μ をもつ正規分布のとき

$$f(x;\mu)=(2\pi)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(x-\mu)^2\right\}$$

であるから

$$\log f(x;\mu)=-\frac{1}{2}\log 2\pi-\frac{1}{2}(x-\mu)^2$$

$$\rho(x;\mu)=\frac{1}{2}\log 2\pi+\frac{1}{2}(x-\mu)^2$$

$$\psi(x;\mu)=x-\mu$$

である。あるいはラプラス分布（2重指数分布）のとき

$$f(x;\mu)=\frac{1}{2}\exp\{-|x-\mu|\}$$

であるから

$$\log f(x;\mu)=-\log 2-|x-\mu|$$

$$\rho(x;\mu)=\log 2+|x-\mu|$$

$$\psi(x;\mu)=\begin{cases}1 & x>\mu\\ -1 & x<\mu\end{cases}$$

である。

正規分布，ラプラス分布いずれの例においても

$$\sum_{i=1}^{n} \psi(x_i;\ \theta)=0 \tag{36}$$

の解は

$$\max \log L(\theta)$$

あるいは

$$\min \sum_{i=1}^{n} \rho(x_i;\ \theta) \tag{37}$$

の解を与える。この(37)式あるいは(36)式によって与えられる推定量をM推定量 M-estimator とよぶ。前述の例では，μ のM推定量は正規分布のとき $\bar{X}$，ラプラス分布のときメジアンである。M推定量という名称は Huber の命名であり，一般化最尤推定量 generalized maximum likelihood estimator のmにもとづいている。

$$\rho(x;\ \theta)=-\log f(x;\ \theta)$$

のとき，すなわち

$$\psi(x;\ \theta)=-\frac{f'(x;\ \theta)}{f(x;\ \theta)}$$

のときM推定量は最尤推定量であることは前述した通りであるが，一般に $\rho(x)=-\log f(x)$ である必要はない。

M推定量は，(36)式から明らかなように，ψ に任意の定数 $c>0$ をかけても変わらない。通常，M推定量は ψ 関数によって区別される。

9　回帰係数のM推定量

線形回帰モデル

$$Y_i=\mathbf{x}_i\boldsymbol{\beta}+u_i,\quad i=1,\ \cdots\cdots,\ n \tag{38}$$

を考えよう。$\mathbf{x}_i$ は $1\times k$ のベクトルである。$\boldsymbol{\beta}$ のM推定量を $\hat{\boldsymbol{\beta}}_{\mathrm{M}}$ とすれば，$\hat{\boldsymbol{\beta}}_{\mathrm{M}}$ は

$$\sum_{i=1}^{n} \rho(Y_i-\mathbf{x}_i\hat{\boldsymbol{\beta}}_{\mathrm{M}})=\min_{\beta} \sum_{i=1}^{n} \rho(Y_i-\mathbf{x}_i\boldsymbol{\beta})$$

の解である。OLS は

$$\rho(u)=\frac{u^2}{2}$$

というM推定量の特別の場合である。すなわち

$$\sum_{i=1}^{n} \rho(Y_i-\mathbf{x}_i\boldsymbol{\beta})=\frac{1}{2}\sum_{i=1}^{n}(Y_i-\mathbf{x}_i\boldsymbol{\beta})^2$$

が OLS の最小にすべき損失関数である。

$\rho(u)$ が微分可能な，0のまわりで対称的な凸関数のとき，$\boldsymbol{\beta}$ のM推定量 $\hat{\boldsymbol{\beta}}_{\mathrm{M}}$ は必要条件

$$\sum \rho'(Y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}_M)\mathbf{x}_i' = \sum \psi(e_i)\mathbf{x}_i' = 0 \tag{39}$$

を満たす。ここで $\psi=\rho'$, $e_i = Y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}_M$ である。OLS は $\rho(u)=u^2/2$ であるから $\rho'(u)=u$, すなわち

$$\psi(e_i)=\rho'(e_i)=e_i$$

したがって OLSE は

$$\sum_{i=1}^{n} e_i \mathbf{x}_i' = 0$$

の解である。

いま $\psi(u)=uw(u)$, すなわちウエイト関数 $w(u)$ を

$$w(u)=\frac{\psi(u)}{u} \tag{40}$$

によって定義すれば, (39)式は

$$\sum_{i=1}^{n} e_i w(e_i)\mathbf{x}_i' = \sum_{i=1}^{n} w_i e_i \mathbf{x}_i' = 0 \tag{41}$$

と表すこともできる。$w_i = w(e_i)$ であり, ウエイト w_i は残差 e_i に依存する。

(41)式は

$$\min \sum_{i=1}^{n} w_i e_i^2$$

すなわち加重最小2乗推定量 weighted least-squares estimator WLSE の解を与える。なぜならば, $\tilde{\boldsymbol{\beta}}$ を $\boldsymbol{\beta}$ の WLSE とし, $\tilde{e}_i$ をその残差

$$\tilde{e}_i = Y_i - \mathbf{x}_i \tilde{\boldsymbol{\beta}}$$

とすれば, $\tilde{\boldsymbol{\beta}}$ は

$$\sum_{i=1}^{n} w_i \tilde{e}_i^2 = \sum_{i=1}^{n} w_i (Y_i - \mathbf{x}_i \tilde{\boldsymbol{\beta}})^2$$

を最小にするから

$$\frac{d\sum w_i \tilde{e}_i^2}{d\tilde{\boldsymbol{\beta}}} = 0$$

すなわち

$$\sum_{i=1}^{n} w_i \tilde{e}_i \mathbf{x}_i' = 0$$

の解でなければならないからである。

(41)式は

$$\sum (Y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}_M) w_i \mathbf{x}_i' = 0 \tag{42}$$

すなわち

$$\sum_{i=1}^{n} \mathbf{x}_i' w_i \mathbf{x}_i \hat{\boldsymbol{\beta}}_M = \sum_{i=1}^{n} \mathbf{x}_i' w_i Y_i \tag{43}$$

と表すこともできる。行列で表せば, (43)式は次のような加重最小2乗法の正規方程式を与える。

$$\mathbf{X}'\mathbf{W}\mathbf{X}\boldsymbol{\beta}=\mathbf{X}'\mathbf{W}\mathbf{y}$$

ここで $\mathbf{X}$ は $n\times k$, $\mathbf{y}$ は $n\times 1$, $\boldsymbol{\beta}$ は $k\times 1$, $\mathbf{W}$ は

$$\mathbf{W}=\begin{bmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{bmatrix}$$

によって与えられる $n\times n$ の対角行列である。

(42)式（あるいは(41)式でも同じ）の解は，weight の W をとり，W 推定値とよばれることもある。

結局 M 推定量は，ウエイトが残差に依存する加重最小2乗回帰の解として求めることができることがわかった。M 推定量の求め方について詳しくは後で説明する。

OLS のとき

$$\psi(e_i)=e_i$$

であるから

$$w_i=\frac{\psi(e_i)}{e_i}=1,\ i=1,\ \cdots\cdots,\ n$$

である。すなわち加重回帰という観点からみれば，OLS はすべての残差に等ウエイト1を与える。残差が絶対値できわめて大きい値をとる外れ値と思われる場合であっても，OLS においては他の残差と同じウエイトが与えられる。

ところで頑健推定とは，外れ値 outlier のパラメータ推定値への影響を小さくしようとする方法であるから，加重回帰で考えれば，絶対値の大きな残差に対してはウエイトを小さくすればよい。ところが残差 e_i の大きさは被説明変数の単位に依存しているから，e_i の水準によって外れ値かどうかを判断することはできない。そこで e_i と同じ単位をもつ誤差項の標準偏差 σ で割り，標準化された残差 e_i/σ を考えると，σ は定数であるから，(39)式は

$$\sum_{i=1}^{n}\psi\left(\frac{e_i}{\sigma}\right)\mathbf{x}_i'=0 \tag{44}$$

と同等である。$e_i^*=e_i/\sigma$ とおけば，上式は

$$\sum_{i=1}^{n}\psi(e_i^*)\,\mathbf{x}_i'=0 \tag{45}$$

に等しく，(45)式は

$$w_i=\frac{\psi(e_i^*)}{e_i^*} \tag{46}$$

とすれば

$$\sum_{i=1}^{n}w_i\,e_i^*\,\mathbf{x}_i'=0 \tag{47}$$

と同等である。

頑健推定の例として Huber の ψ をとりあげよう。Huber の ψ 関数は次式で与えられる。

図10　OLS の ψ 関数

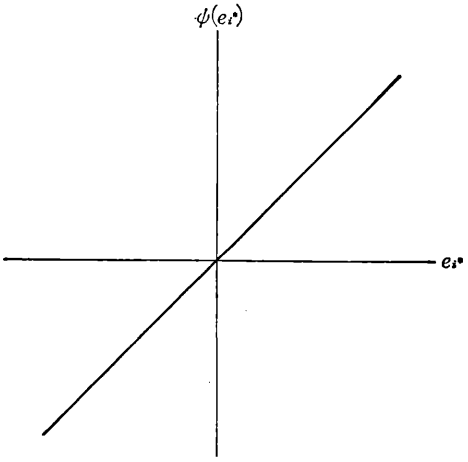
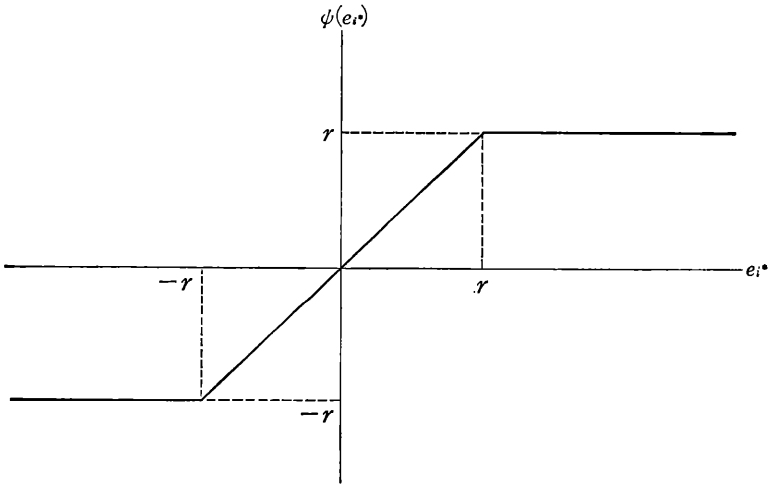


図11　Huber の ψ 関数



$$\psi(e_i^*)=\begin{cases} e_i^* & |e_i^*|\leq r \\ r & e_i^*>r \\ -r & e_i^*<-r \end{cases} \tag{48}$$

損失関数 ρ で示せば次の関数である。

$$\rho(e_i^*)=\begin{cases} \dfrac{e_i^{*2}}{2} & |e_i^*|\leq r \\ r|e_i^*|-\dfrac{r^2}{2} & |e_i^*|>r \end{cases} \tag{49}$$

Huber の $\psi(e_i^*)$ と，OLS の $\psi(e_i^*)=e_i^*$ をグラフで示したのが図10，図11である。図からわかるように，OLS の ψ が残差 e_i^* に対して限界がないのに対して，Huber の ψ は e_i^* が絶対値で r

をこえる場合には，その影響を r で止めてしまうことがわかる。ウエイトで示せば Huber の ψ は

$$w_i=\begin{cases} 1 & |e_i^*|\leq r \\ \dfrac{r}{|e_i^*|} & |e_i^*|>r \end{cases} \tag{50}$$

というウエイトを与えるから，$|e_i^*|>r$ のとき，ウエイトを $r/|e_i^*|<1$ と1より小さくさせる。

10 M推定量の影響関数

影響関数

M推定量の影響関数は次式で与えられる（Huber [19]，p. 45）。

$$IC(x\,;\,F,\,T)=\frac{\psi(x,\,T(F))}{-\int(\partial/\partial\theta)[\psi(x\,;\,T(F))]\,dF(x)} \tag{51}$$

この結果で重要な点は影響関数は ψ 関数に比例するという点である。すなわち

影響関数＝(定数)×ψ 関数

という関係である。したがってM推定量においては ψ が影響関数とよばれることもある。

とくに位置問題において ψ が $\psi(x\,;\,\theta)=\psi(x-\theta)$ の形をしているとき IC は次式のように書くことができる。

$$IC(x\,;\,F,\,T)=\frac{\psi[x-T(F)]}{\int\psi'[x-T(F)]\,dF(x)} \tag{52}$$

そして F が x の真のモデル分布，$T(F)$ がその分布の位置パラメータのとき

$$IC(x\,;\,F,\,T)=\frac{\psi[x-T(F)]}{\int\psi'[x-T(F)]\,dF(x)}=\frac{\psi}{E(\psi')} \tag{53}$$

となる（Hampel et al. [13]，p. 103）。

線形回帰モデルのパラメータ $\boldsymbol{\beta}$ のM推定量 $\hat{\boldsymbol{\beta}}_{\mathrm{M}}(F)$ の影響関数は次式で与えられる（Hampel et al. [13]，p. 313）。

$$IC[(\mathbf{x},\,Y)\,;\,F,\,\hat{\boldsymbol{\beta}}_{\mathrm{M}}(F)]=\psi[Y-\mathbf{x}\,\hat{\boldsymbol{\beta}}_{\mathrm{M}}(F)]\,\mathbf{B}^{-1}\mathbf{x}' \tag{54}$$

ここで

$$\mathbf{B}=\int\psi'[Y-\mathbf{x}\,\hat{\boldsymbol{\beta}}_{\mathrm{M}}(F)]\,\mathbf{x}'\mathbf{x}\,dF(\mathbf{x},\,Y)$$

(54)式は次の特徴を有している。

$\boldsymbol{\beta}$ の OLSE $\hat{\boldsymbol{\beta}}(F)$ の無限標本で評価された影響関数 (13) 式と比較することによって次のことがわかる。

① $\hat{\boldsymbol{\beta}}(F)$ は残差 $Y-\mathbf{x}\hat{\boldsymbol{\beta}}(F)$ に対して限界をもたないが，$\hat{\boldsymbol{\beta}}_{\mathrm{M}}(F)$ は ψ で与えられる限界を有する。

② OLS のとき $\psi(u)=u$ であるから $\psi'(u)=1$ となり，$\mathbf{B}=\sum_{\mathrm{xx}}(F)$ となるが，M推定にお

いては $\psi'(u)=1$ ではない。

しかし(54)式から

③ もし $\mathbf{x}$ に限界がなければ，$\mathbf{B}^{-1}\mathbf{x}'$ は限界をもたず，したがって ψ が大きな残差に限界を画しても，$\hat{\boldsymbol{\beta}}_{\mathrm{M}}$ の IC は限界をもたない。しかし，有限標本の観測値の世界で $\mathbf{x}$ が無限に大きくなっていく状況を経済学で想定する必要はない。それゆえ，$\mathbf{B}^{-1}\mathbf{x}'$ によって $\hat{\boldsymbol{\beta}}_{\mathrm{M}}$ の IC の限界がなくなってしまう可能性は深刻ではない。大きな残差 e_i^* に対して $\psi(e_i^*)$ がどのように限界を画し，残差のウエイトを小さくするかがやはり重要である。

例として確率変数 u の真のモデルが標準正規分布に従っているとき，Huber の ψ 関数をもつM推定量の影響関数を求めよう。Huber の ψ は

$$\psi(u)=\begin{cases} u & |u|\leq r \\ r & u>r \\ -r & u<-r \end{cases}$$

したがって

$$\psi'(u)=\begin{cases} 1 & |u|\leq r \\ 0 & |u|>r \end{cases} \tag{55}$$

であるから

$$E(\psi')=\int_{-r}^{r}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du=1-2\Phi(-r) \tag{56}$$

が得られる（数学注(1)参照)。

したがって Huber の ψ をもつM推定量の影響関数は次式で与えられる。

$$IC=\begin{cases} \dfrac{u}{1-2\Phi(-r)} & |u|\leq r \\ \dfrac{r}{1-2\Phi(-r)} & u>r \\ \dfrac{-r}{1-2\Phi(-r)} & u<-r \end{cases} \tag{57}$$

M推定量の不偏性と漸近的特性

ψ が奇関数

$$\psi(-u)=-\psi(u)$$

で，分布 F が中心 T のまわりで対称ならば，

$$\psi\left(\frac{X_i-T_n}{\sigma}\right)=0$$

の解であるM推定量 T_n は T の不偏推定量である（Goodall〔8〕，p. 364)。

たとえば前述の Huber の ψ は奇関数であり，X の分布が正規分布であるならば，正規分布は μ を中心に左右対称であるから，Huber の ψ を用いる μ のM推定量は不偏性をもつ。

真の分布が F であるときのパラメータ $T(F)$ のM推定量 T_n は

$$\sqrt{n}[T_n - T(F)] \xrightarrow{d} N[0,\ V(T,\ F)] \tag{58}$$

$$V(T,\ F) = \int IC(x\ ;\ T,\ F)^2\, dF(x) \tag{59}$$

と，漸近的に正規分布をし，その分散 $V(T,\ F)$ は影響関数を用いて(59)式のように表すことができる（Huber〔19〕p. 45，Hampel et al.〔13〕p. 85）。

IC が(53)式のように表すことができる場合には

$$V(T,\ F) = \frac{E(\psi^2)}{[E(\psi')]^2} \tag{60}$$

となる。

線形回帰モデルのパラメータ β の OLSE $\hat{\beta}$ の場合

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

となるが，M推定量 $\hat{\boldsymbol{\beta}}_{\mathrm{M}}$ の場合，σ^2 に代わって漸近的分散が(60)式によって与えられるから，$\hat{\boldsymbol{\beta}}_{\mathrm{M}}$ の漸近的分散・共分散行列は

$$V(\hat{\boldsymbol{\beta}}_{\mathrm{M}}) = \frac{E(\psi^2)}{[E(\psi')]^2}(\mathbf{X}'\mathbf{X})^{-1} \tag{61}$$

となる。

モデルの真の分布が標準正規分布のとき，Huber の ψ によるM推定量の漸近的分散を(60)式で求めよう。

$$E(\psi^2) = \int_{-r}^{r} u^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$
$$+ \int_{-\infty}^{-r} r^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du + \int_{r}^{\infty} r^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

そして

$$\int_{-r}^{r} \frac{u^2}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 2\int_{0}^{r} \frac{u^2}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

$$\int_{0}^{r} u^2 e^{-\frac{u^2}{2}} du = -ue^{-\frac{u^2}{2}}\Big|_0^r + \int_0^r e^{-\frac{u^2}{2}} du = -re^{-\frac{r^2}{2}} + \int_0^r e^{-\frac{u^2}{2}} du$$

ゆえに

$$\int_{-r}^{r} \frac{u^2}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = -\frac{2r}{\sqrt{2\pi}} e^{-\frac{r^2}{2}} + \int_0^r \frac{2}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$
$$= -2r\phi(r) + 1 - 2\Phi(-r)$$

したがって

$$E(\psi^2) = -2r\phi(r) + 1 - 2\Phi(-r) + 2r^2\Phi(-r) \tag{62}$$

この結果と(56)式を用いて，真の分布 F が標準正規分布のときの Huber の ψ のM推定量の漸近的分散は次式となる。

$$V(T,\ F) = \frac{E(\psi^2)}{[E(\psi')]^2} = \frac{-2r\phi(r) + 1 - 2\Phi(-r) + 2r^2\Phi(-r)}{[1 - 2\Phi(-r)]^2} \tag{63}$$

標準正規分布の分散は1であるから，真の確率分布が正規分布のときの，Huber の ψ によるM

推定量 T_n の漸近的有効性 $AE(T_n)$ は次式で与えられる。

$$AE(T_n)=\frac{[1-2\Phi(-r)]^2}{-2r\phi(r)+1-2\Phi(-r)+2r^2\Phi(-r)} \tag{64}$$

同様にして，線形回帰モデルのパラメータ β_j のM推定量 $\hat{\beta}_{Mj}$ の漸近的有効性 $AE(\hat{\beta}_{Mj})$ は，誤差項の真の分布が期待値0，分散 σ^2 の正規分布に従うとき

$$AE(\hat{\beta}_{Mj})=\frac{\sigma^2 m^{jj}}{\dfrac{E(\psi^2)}{[E(\psi')]^2}m^{jj}}=\frac{\sigma^2[E(\psi')]^2}{E(\psi^2)} \tag{65}$$

$$j=1,\ 2,\ \cdots\cdots,\ k$$

によって与えられる。ここで m^{jj} は $(\mathbf{X}'\mathbf{X})^{-1}$ の $(j,\ j)$ 要素，$j=1,\ \cdots\cdots,\ k$ である。

11 Huber の ψ 関数

M推定量を与えるいくつかの代表的な ψ 関数は次節で説明するが，ここではM推定量のもっとも代表的な Huber の ψ 関数についてこれまでに得られた結果をまとめながら，その特徴を考察しておこう。

Huber の ψ 関数の背後には次の損失関数 ρ がある。

$$\rho(u)=\begin{cases}\dfrac{u^2}{2} & |u|\leq r\\[2ex] r|u|-\dfrac{u^2}{2} & |u|>r\end{cases} \tag{49}$$

影響関数 ψ，ウエイト関数 w は次の通りであった。

$$\psi(u)=\begin{cases}u & |u|\leq r\\ r & u>r\\ -r & u<-r\end{cases} \tag{48}$$

$$w(u)=\begin{cases}1 & |u|\leq r\\[2ex] \dfrac{r}{|u|} & |u|>r\end{cases} \tag{50}$$

図12(a)～(c)は Huber の ψ によるM推定量の損失関数 ρ，影響関数 ψ およびウエイト関数 w である。図12(a)には OLS の損失関数 $\rho(u)=u^2/2$ も示されている。また図12は調整定数 tuning constant r の値は1.345で描かれている。$r=1.345$は，u の真の分布が標準正規分布のとき，Huber の ψ によるM推定量の漸近的有効性が0.95となる値である。

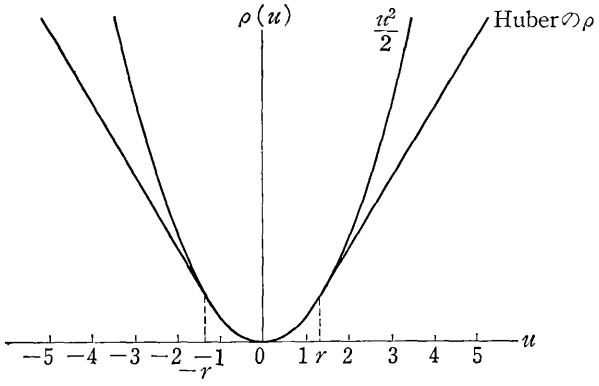
OLS は

$$\rho(u)=\frac{u^2}{2},\ \ \psi(u)=u,\ \ w(u)=1$$

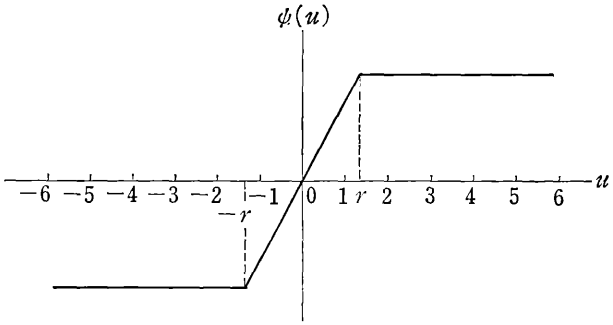
であったから，図12を参照しつつ OLS と比較すれば Huber のM推定量は次の特徴をもつことがわかる。Huber の ψ にもとづくM推定量（あるいは推定法）を $ME(H)$ と表すことにする。

図12　Huber の ρ, ψ, w

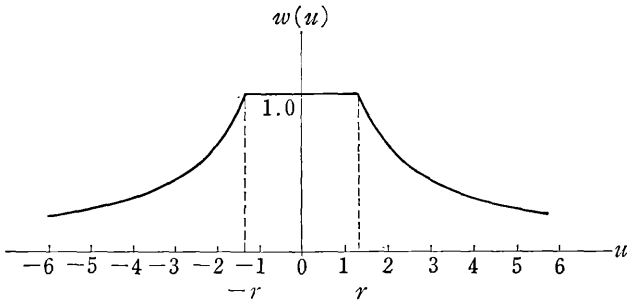
(a) 損失関数ρ



(b) 影響関数ψ



(c) ウエイト関数w



①OLS も $ME(H)$ も，損失関数でみるとuの正，負の値に対称的であり，損失は限界なく大きくなり得る。しかし $ME(H)$ は OLS と比較して，$|u|>r$ のuに対して損失関数の値は OLS のように u^2 に比例して大きくはならない。

②同じことをψでみれば，$|u|>r$ のuに対して，uのパラメータ推定値への影響は OLS は限界がないのに対して $ME(H)$ は限界をもつ。しかし $ME(H)$ のψ関数の値はあるuをこえると次

第に小さくなる，あるいは0になるという性質はもっていない。

このことをψ関数を評価する基準である総誤差感度 gross error sensitivity（以下 g.e.s. と略す），局所方向感度 local shift sensitivity（以下 l.s.s. と略す）および排除点 rejection point を用いて説明すれば次のような特徴を $ME(H)$ はもつ。

③まず g.e.s. とは，不適当な観測値 contaminated observation が，漸近的に，推定量 T_n に与える最大の効果のことであり，次式で定義される。

$$\text{g.e.s.}=\sup_u |IC(u;\ F,\ T)|$$

$\sup_u |IC(u;\ F,\ T)|$ とは任意のuに対して，$|IC(u;\ F,\ T)|$ が超えることのない最小の値である。推定量は g.e.s. が有界であるときにのみ効率の頑健性 robustness of efficiency をもち，抵抗力をもつ（resistant である）。もし $\bar{X}$ や $\hat{\boldsymbol{\beta}}$ の g.e.s. のように，g.e.s. が有界でなければ外れ値に対してその推定量は無防備であることを意味する。したがって g.e.s. が小さいほど頑健性は高い。

Huber のψに対する IC は

$$IC=\begin{cases} \dfrac{u}{1-2\Phi(-r)} & |u|\leq r \\ \dfrac{r}{1-2\Phi(-r)} & u>r \\ \dfrac{-r}{1-2\Phi(-r)} & u<-r \end{cases} \tag{57}$$

であったから，$ME(H)$ の g.e.s. は

$$\text{Huber の}\psi\text{の g.e.s.}=\frac{r}{1-2\Phi(-r)}$$

によって与えられ，有界である。これに対して OLS の g.e.s. は∞である。

④局所方向感度 l.s.s. とは次式で定義される。

$$\text{l.s.s.}=\sup_{u_1\neq u_2}\frac{|IC(u_2)-IC(u_1)|}{|u_2-u_1|}$$

IC 曲線が微分可能ならば，平均値の定理によって

$$IC(u_2)-IC(u_1)=IC'(u^*)(u_2-u_1)$$

であるから，l.s.s. は IC 曲線の最大の勾配である。すなわち

$$\text{l.s.s.}=\sup_u |IC'(u)|$$

l.s.s. が有界であれば高い抵抗力をもつが，しかし OLS によって得られる $\bar{X}$ や $\hat{\boldsymbol{\beta}}$ の l.s.s. は1で有限の大きさをもつのに対して，メジアンのように l.s.s. が∞という頑健推定量もあり，この l.s.s. 有界という基準を満たすことをψに要求することはきわめて厳しい条件を課すことになる。Huber のψは(57)式の IC 曲線から容易に分るように次の l.s.s. をもち，有界である。

$$\text{Huber の}\psi\text{の l.s.s.}=\frac{1}{1-2\Phi(-r)}$$

⑤ある値 r^* に対して，$|u|>r^*$ のとき $\psi(u)=0$ となるならば，この r^* は排除点 rejection

point とよばれる。絶対値で r^* をこえる u のパラメータ推定への影響は排除されるからである。排除点をもつψ関数はきわめて大きな外れ値に対して防備されている。Huber のψは図 12 (b), (c) からわかるように排除点は∞である。排除点が∞ということをウエイト関数でいえば，$|u|>r$ の u に対してウエイトは 1 より小さくなるが，0 になることはないという意味である。

⑥加重回帰という観点でみれば，OLS はすべての u に等ウエイトを与え，$ME(H)$ は絶対値で r をこえる u に対してはウエイトを 1 より小さくする。$|u|$ が大きくなればなるほどその u に対するウエイトは小さくなる。しかしウエイトが 0 になることはない。

⑦Huber のψによるM推定量は最小好適分布 least favorable distribution ((5)式，以下 l. f. d. と略す）の最尤推定量である。

8 節でM推定量のψが確率密度関数 f と

$$\psi=-\frac{f'}{f}$$

の関係にあるならば，M推定量は最尤推定量であることを述べた。l. f. d. の f は(5)式で与えられるから

$$f'(u)=\begin{cases}\dfrac{1-\varepsilon}{\sqrt{2\pi}}(-u)\,e^{-\frac{u^2}{2}}=-uf(u) & |u|\leq r\\[2ex] \dfrac{1-\varepsilon}{\sqrt{2\pi}}(-r)\,e^{\frac{r^2}{2}-r|u|}=-rf(u) & |u|>r\end{cases}$$

となり，したがって(48)式で与えられる Huber のψが

$$\psi=-\frac{f'}{f}$$

の関係にあることは明らかである。

⑧u の真の分布が標準正規分布であるとき，OLS は位置パラメータの最小分散不偏推定量（有効推定量）を与える。もし，このときψ関数による頑健推定を行なえばM推定量の有効性は低下する。外れ値に対して頑健性をもち，真の分布が正規分布のときにも推定量の有効性が余り低下しない頑健推定を行ないたい。そのためには，Huber のψを用いるM推定量の有効性が，u の真の分布が標準正規分布の場合にくらべてどの程度変化するかを調べなくてはならない。

Huber のψを用いるM推定量の漸近的分散は(63)式，有効性は(64)式で与えられている。この式より $r=1.345$ のとき $ME(H)$ の漸近的有効性は0.95となることがわかる。

以上の結果を用いて，真の分布が標準正規分布であるとき Huber のψによるM推定量の

$A=E(\psi^2)$

$B=E(\psi')$

漸近的分散 $V=\dfrac{A}{B^2}$

表 4 Huber の M 推定量

r	A	B	V	e	G^*	L^*
1.000	0.5160585510	0.6826893091	1.10727	0.90312	1.46480	1.46480
1.050	0.5473536220	0.7062817812	1.09727	0.91136	1.48666	1.41587
1.100	0.5777049610	0.7286678553	1.08805	0.91908	1.50960	1.37237
1.150	0.6070179466	0.7498562336	1.07956	0.92631	1.53363	1.33359
1.200	0.6352146880	0.7698608637	1.07176	0.93305	1.55872	1.29894
1.250	0.6622331429	0.7887006998	1.06460	0.93932	1.58489	1.26791
1.300	0.6880261946	0.8063992262	1.05805	0.94514	1.61210	1.24008
1.350	0.7125604964	0.8229842186	1.05206	0.95052	1.64037	1.21509
1.400	0.7358156628	0.8384869099	1.04659	0.95548	1.66967	1.19262
1.450	0.7577830763	0.8529417515	1.04161	0.96005	1.70000	1.17241
1.500	0.7784649411	0.8663858175	1.03709	0.96424	1.73133	1.15422
1.550	0.7978727915	0.8788586855	1.03299	0.96806	1.76365	1.13784
1.600	0.8160270162	0.8904014826	1.02928	0.97155	1.79694	1.12309
1.650	0.8329548526	0.9010571241	1.02593	0.97473	1.83118	1.10981
1.700	0.8486907225	0.9108690023	1.02291	0.97760	1.86635	1.09785
1.750	0.8632736207	0.9198815823	1.02020	0.98020	1.90242	1.08710
1.800	0.8767476003	0.9281392097	1.01777	0.98254	1.93936	1.07742
1.850	0.8891600712	0.9356862307	1.01559	0.98465	1.97716	1.06873
1.900	0.9005609912	0.9425666332	1.01365	0.98653	2.01577	1.06093
1.950	0.9110025880	0.9488235712	1.01193	0.98821	2.05518	1.05394
2.000	0.9205376846	0.9544994831	1.01039	0.98971	2.09534	1.04767
2.050	0.9292208110	0.9596352577	1.00904	0.99105	2.13623	1.04206
2.100	0.9371049596	0.9642709494	1.00784	0.99222	2.17781	1.03705
2.150	0.9442443091	0.9684445858	1.00678	0.99327	2.22005	1.03258
2.200	0.9506911189	0.9721928835	1.00585	0.99418	2.26293	1.02860
2.250	0.9564960763	0.9755508900	1.00504	0.99499	2.30639	1.02506
2.300	0.9617086386	0.9785517454	1.00433	0.99569	2.35041	1.02192
2.350	0.9663772332	0.9812265635	1.00371	0.99630	2.39496	1.01913
2.400	0.9705468926	0.9836049080	1.00317	0.99684	2.44000	1.01667
2.450	0.9742604995	0.9857144356	1.00270	0.99730	2.48551	1.01449
2.500	0.9775594224	0.9875807762	1.00230	0.99770	2.53144	1.01258

r=Huber's tuning constant　　$G^*=r/B$ gross error sensitivity
$A=E(\mathrm{PSI}\ 2)$: $B=E(\mathrm{PSI}')$　　$L^*=1/B$ local shift sensitivity
$V=A/B$ 2 漸近的分散 : $e=1/V$ efficiency

漸近的有効性 $e=\dfrac{1}{V}$

総誤差感度 $G^*=\dfrac{r}{B}$

局所方向感度 $L^*=\dfrac{1}{B}$

を，調整定数 r の関数として計算したのが表4である。表4は r を 1.0 (0.05) 2.5 と動かしている。

頑健性という観点からいえば G^* の値は小さいほどよく，L^* は有界であればよい。真の分布が標準正規分布であるときの漸近的有効性 e は 0.95 あるいは 0.99 あればよいと思われるが，G^* と e

との間はトレード・オフの関係にあることがわかる。調整定数 r を大きくしていけば，Huber の ϕ を用いるM推定量は OLSE に近づいていくが，G^* を大きくし，外れ値に対する頑健性は悪くなる。r の値は通常，漸近的有効性 0.95を与える 1.345が採用されるが，外れ値に対する備えを強くしようとすれば $r=1.0$，ゆるくしてもよければ $r=2.0$ としてもよい。

Huber の ϕ を用いるM推定量は計量経済学においても注目すべき推定量である。正規分布より長い尾をもつ分布からの「外れ値」に対して頑健であり，排除点をもたないから，定式化のミスによる大きな「外れ値」をパラメータ推定から排除して間違ったモデルに都合のよいパラメータ推定をする危険はなく，漸近的有効性と調整定数との関係が明瞭であるという意味において注目すべきである。

数学注

(1) (6)式の証明

$f(x)$ は確率密度関数であるから

$$\int_{-\infty}^{\infty} f(x)\,dx=1$$

を満たさなければならない。すなわち

$$1=\int_{-\infty}^{-k} f(x)\,dx+\int_{-k}^{k} f(x)\,dx+\int_{k}^{\infty} f(x)\,dx$$

$$=\int_{-\infty}^{\infty}\frac{1-\varepsilon}{\sqrt{2\pi}}e^{\frac{k^2}{2}+kx}\,dx+\int_{-k}^{k}\frac{1-\varepsilon}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx+\int_{k}^{\infty}\frac{1-\varepsilon}{\sqrt{2\pi}}e^{\frac{k^2}{2}-kx}\,dx$$

を満たさなければならない。

上式の右辺第2項を累積分布関数を用いて表すと次のようになる。

$$1=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx$$

$$=\int_{-\infty}^{-k}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx+\int_{-k}^{k}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx+\int_{k}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx$$

$$=2\Phi(-k)+\int_{-k}^{k}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx$$

したがって

$$\int_{-k}^{k}\frac{1-\varepsilon}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx=(1-\varepsilon)[1-2\Phi(-k)]$$

次に右辺第1項と第3項を ϕ で表そう。

$$\int_{-\infty}^{-k}e^{kx}\,dx=\frac{e^{kx}}{k}\bigg|_{-\infty}^{-k}=\frac{e^{-k^2}}{k}$$

$$\int_{k}^{\infty}e^{-kx}\,dx=-\frac{e^{-kx}}{k}\bigg|_{k}^{\infty}=\frac{e^{-k^2}}{k}$$

であるから

$$\int_{-\infty}^{-k}\frac{1-\varepsilon}{\sqrt{2\pi}}e^{\frac{k^2}{2}+kx}\,dx=\left(\frac{1-\varepsilon}{k}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{k^2}{2}}=\left(\frac{1-\varepsilon}{k}\right)\phi(k)$$

同様に

$$\int_k^\infty \frac{1-\varepsilon}{\sqrt{2\pi}} e^{\frac{k^2}{2}-kx} dx = \left(\frac{1-\varepsilon}{k}\right) \phi(k)$$

したがって

$$1 = \frac{2(1-\varepsilon)}{k} \phi(k) - (1-\varepsilon)[1-2\Phi(-k)]$$

この式から(6)式が得られる。

(2) (7)式の証明

$x < -k$ のとき

$$F_1(x) = \frac{1-\varepsilon}{\sqrt{2\pi}} \int_{-\infty}^x e^{\frac{k^2}{2}+ku} du = \left(\frac{1-\varepsilon}{\sqrt{2\pi}} e^{\frac{k^2}{2}}\right) \frac{e^{ku}}{k} \Big|_{-\infty}^x$$

$$= \left(\frac{1-\varepsilon}{\sqrt{2\pi}} e^{\frac{k^2}{2}}\right)\left(\frac{e^{kx}}{k}\right) = \frac{1-\varepsilon}{\sqrt{2\pi}\, k} e^{\frac{k^2}{2}+kx} = \frac{1-\varepsilon}{k} \frac{1}{\sqrt{2\pi}} e^{-\frac{k^2}{2}} e^{k^2+kx}$$

$$= \frac{1-\varepsilon}{k} \phi(k) e^{k^2+kx}$$

$-k \leq x \leq k$ のとき

$$F_2(x) = F_1(-k) + \int_{-k}^x \frac{1-\varepsilon}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

$$= \frac{1-\varepsilon}{\sqrt{2\pi}\, k} e^{-\frac{k^2}{2}} + (1-\varepsilon)[\Phi(x) - \Phi(-k)]$$

$$= (1-\varepsilon)\left\{\frac{\phi(k)}{k} + \Phi(x) - \Phi(-k)\right\}$$

ところが(6)式より

$$\Phi(-k) = \frac{\phi(k)}{k} - \frac{\varepsilon}{2(1-\varepsilon)}$$

であるから，この式を上式右辺の $\Phi(-k)$ へ代入すれば

$$F_2(x) = \frac{\varepsilon}{2} + (1-\varepsilon)\Phi(x)$$

が得られる。

$x > k$ のとき

$$F_3(x) = F_2(k) + \frac{1-\varepsilon}{\sqrt{2\pi}} \int_k^x e^{\frac{k^2}{2}-ku} du$$

$$= (1-\varepsilon)\left\{\frac{\phi(k)}{k} + \Phi(k) - \Phi(-k)\right\} + \frac{1-\varepsilon}{\sqrt{2\pi}} e^{\frac{k^2}{2}} \int_k^x e^{-ku} du$$

そして

$$\Phi(k) - \Phi(-k) = 1 - 2\Phi(-k) = 1 - \frac{2\phi(k)}{k} + \frac{\varepsilon}{1-\varepsilon}$$

$$\frac{1-\varepsilon}{\sqrt{2\pi}} e^{\frac{k^2}{2}} \int_k^x e^{-ku} du = \frac{1-\varepsilon}{\sqrt{2\pi}\, k} e^{-\frac{k^2}{2}} (1 - e^{k^2-kx}) = (1-\varepsilon) \frac{\phi(k)}{k} (1 - e^{k^2-kx})$$

であるから

$$F_3(x)=(1-\varepsilon)\left\{1-\frac{\phi(k)}{k}+\frac{\varepsilon}{1-\varepsilon}\right\}+(1-\varepsilon)\frac{\phi(k)}{k}(1-e^{k^2-kx})$$

$$=1-(1-\varepsilon)\frac{\phi(k)}{k}(1-e^{k^2-kx})$$

参 考 文 献

〔1〕 Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W.: *Robust Estimates of Location: Survey and Advances,* Princeton University Press, 1972.

〔2〕 Carroll, R.J., and Ruppert, D.: *Transformation and Weighting in Regression,* Chapman and Hall, 1988.

〔3〕 Chatterjee, S., and Hadi, A.S.: *Sensitivity Analysis in Linear Regression,* John Wiley & Sons, 1988.

〔4〕 Cook, R.D., and Weisberg, S.: *Residuals and Influence in Regression,* Chapman and Hall, 1982.

〔5〕 Dudewicz, E.J., and Mishra, S.Y.: *Modern Mathematical Statistics,* John Wiley & Sons, 1988.

〔6〕 Fair, R.C., On the robust estimation of econometric models, *Annals of Economic and Social Measurement,* 3/4, 1974, 667-677.

〔7〕 Fox, J.: *Linear Statistical Models and Related Methods,* John Wiley & Sons, 1984.

〔8〕 Goodall, C.: M-estimators of location: an outline of the theory, in Hoaglin, D.C. et al. ed. *Understanding Robust and Exploratory Data Analysis,* John Wiley & Sons, 1983.

〔9〕 Green, P.J.: Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives, *Journal of the Royal Statistical Society,* B, 46, No. 2, 149～192.

〔10〕 Haavelmo, T.: *The Probability Approach in Econometrics,* Cowles Commission Paper, New Series No. 4, John Wiley & Sons, 1944. (山田勇訳編『計量経済学の確率的接近法』, 岩波書店, 1953).

〔11〕 Hampel, F.R.: A general qualitative defenition of robustness, *Annals of Mathematical Statistics,* 42, 1971, 1887-1896.

〔12〕 Hampel, F.R.: The influence curve and its role in robust estimation, *Journal of the American Statistical Association,* Vol. 69, No. 346, June 1974, 383-393.

〔13〕 Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A.: *Robust Statistics,* John Wiley & Sons, 1986.

〔14〕 Hinkley, D., and Wang, H.L.: A trimmed jackknife, *Journal of the Royal Statistical Society,* B, Vol. 42, No. 3, 1980, 347-356.

〔15〕 Hoaglin, D.C.: Using quantiles to study shape, in Hoaglin, D.C., Mosteller, F., and Tukey, J.W. ed., *Exploring Data Tables, Trends and Shapes,* John Wiley & Sons, 1985.

〔16〕 Hogg, R.V.: An introduction to robust procedures, *Communications in Statistics,* A 6, 1977, 789-794.

〔17〕 Holland, P.W.: Robust regression using iteratively reweighted least-squares, *Communications in Statistics,* A 6, 1977, 813-827.

〔18〕 Huber, P.J.: Robust regression: asymptotics, conjectures and Monte Carlo, *Annals of Statistics,* Vol. 1, No. 5, 1973, 799-821.

〔19〕 Huber, P.J.: *Robust Statistics,* John Wiley & Sons, 1981.

〔20〕 Iglewicz, B.: Robust scale estimators and confidence interval for location, in Hoaglin, D.C.

et al. ed. *Understanding Robust and Exploratory Data Analysis,* John Wiley & Sons, 1983.

〔21〕 Koenker, R.: Robust methods in econometrics, *Econometric Reviews,* 1(2), 213-255.

〔22〕 Li, G.: Robust regression, in Hoaglin, D.C. et al. ed. *Exploring Data Tables, Trends, and Shapes,* John Wiley & Sons, 1985.

〔23〕 Miller, R.G.: The jackknife-a review, *Biometrika,* Vol. 61(1), 1974, 1-15.

〔24〕 蓑谷千凰彦："経済学における統計的認識" (2),「三田学会雑誌」69巻5号, 1976, 89-106.

〔25〕 Mosteller, F., and Tukey, J.W.: *Data Analysis and Regression,* Addison-Wesley Publishing Company, 1977.

〔26〕 Myers, R.H.: *Classical and Modern Regression with Applications,* Duxbury Press, 1986.

〔27〕 Weisberg, S.: *Applied Linear Regression* 2nd ed., John Wiley & Sons, 1985.

〔28〕 Welsch, R.E., Regression sensitivity analysis and bounded-influence estimation, in Kmenta, J., and Ramsey, J.B. ed., *Evaluation of Econometric Models,* 1980, Academic Press.

〔29〕 Wetherill, G.B.: *Regression Analysis with Applications,* Chapman and Hall, 1986.

(経済学部教授)