

Title	回帰モデルの誤差項の正規性検定(1)
Sub Title	Tests for normality of disturbances in regression model (1)
Author	蓑谷, 千鳳彦
Publisher	慶應義塾経済学会
Publication year	1990
Jtitle	三田学会雑誌 (Keio journal of economics). Vol.83, No.2 (1990. 7) ,p.362(148)- 400(186)
JaLC DOI	10.14991/001.19900701-0148
Abstract	
Notes	論説
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00234610-19900701-0148

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.

回帰モデルの誤差項の正規性検定*（1）

蓑谷千凰彦

はじめに

回帰モデルの誤差項が正規分布にしたがうという仮定は中心極限定理によって正当化され、パラメータ推定および仮説検定の基礎となっている。この誤差項の正規性の仮定に対しては計量経済学者のほとんど誰もが疑問を抱かず、Haavelmo [18] は聖典とされ、もし誤差項が正規分布しないとすれば、モデルの定式化が悪い、データ収集方法が悪い、データ数が足りないかであると考えられるようになった。1世紀前の Galton [16] の次のような正規分布への賛美歌は現在においてもなお計量経済学者の耳に快よく響く。

「私は“誤差の度数法則”によって表されている宇宙の秩序の驚くべき形式ほど感銘を与えるものを何も知らない。もし古代ギリシア人がこの法則を知っていたならば、それは擬人化され、神格化されたであろう。この法則は、非常な無秩序のなかで静かにそして全く表面に出ないように君臨している。群が大きいほど、また外見上無秩序が甚だしいほど、この法則の支配力は完全である。それは無理由の至高の法則である。無秩序な諸要素のなかから大標本を抽出し、それらを大きさの順に並べるならば、思いもよらぬ非常に美しい秩序の形式がかくされていたことが明らかになる。」

正規性の仮定は、最小2乗法や t 検定を支えている重要な補助仮説であるにもかかわらず、この補助仮説自体の検定および正規性の仮定がくずれた場合のパラメータ推定や仮説検定への影響、非正規性への対処については計量経済学は等閑視してきた。初級、上級を問わず、計量経済学のテキストを繙いても正規性検定についてはほとんど全く言及されていない。

正規分布の仮定を検定するということが等閑視されてきたのは、ひとつには確かに正規性の仮定が成立していることが多いからである。自己相関なしの仮定のようにくずれることが多い仮定であるならば正規性の仮定はもっと注目されたにちがいない。第2に正規性の仮定が成立しなくても、誤差項が期待値0、均一分散、自己相関なしの仮定を満たしているならば、線形回帰モデルの回帰係数の通常の最小2乗推定量（OLSE）は最良線形不偏推定量（BLUE）である。このようにもっともよく使われる最小2乗法は良い特性をもつ推定量を与えるから正規性の仮定がくずれた場合の

* この研究は慶應義塾学事振興資金の研究助成を受けている。記して感謝したい。

OLSEの特性は考察されなかった。第3に誤差項の正規性からの若干の乖離があっても、 t 検定の有意水準は余り大きく影響されないという妥当性の頑健 *robustness of validity* を t 検定は有しているからである。

しかし、最近、実証分析に携る人たちによって、回帰モデルにおける大きな残差は、正規分布からの「外れ値」*outlier* ではなく、むしろ正規分布より長い尾をもつ（両すそが厚いといわれることもある）分布からの誤差項の実現値ではないかと思われる例が多く指摘されるようになってきた。廣松他〔20〕は Geary の検定を用いることによって、誤差項の正規分布の仮定が成立していない関数の例として、輸出、労働時間、卸売物価、消費のインプリシットデフレーター、サービス輸出のインプリシットデフレーター、政府移転支出、直接税を挙げている。岩田〔24〕は $u_t = \Delta \log p_t$ (p_t は t 日の株価終値) が急尖の分布 *leptokurtic distribution* にしたがうことを示している。高木〔33〕は為替レートの日次変化率が正規分布より長い尾をもつ分布にしたがっていることに注意を促している。岩田、高木の例は回帰モデルの誤差項の分布ではないが、株価変化率、為替レート変化率が被説明変数とするモデル構築においては当然考慮すべき分布の型である。

本論文は、回帰モデルの誤差項の正規性の仮定も、自己相関なしの仮定と同様検定すべき仮説の1つであるという考えのもとに、さまざまな正規性の検定方法を検討し、スチューデント化残差とランキットの正規確率プロット、D'Agostino 検定および Shapiro-Wilk 検定が計量経済学の分野において有用な正規性検定の方法であることを、理論・応用両面にわたって展開する。

まず正規分布より長い尾をもつ分布の紹介（1節）、この長い尾をもつ分布のもとで、位置パラメータ（期待値 μ あるいは回帰モデルの β ）の OLSE の有効性は著しく損われること（2節）、仮説検定において効率の頑健性 *robustness of efficiency* はきわめて小さくなること（3節）を述べる。

次に正規確率プロット、いくつかの正規性検定統計量とその検定力を説明し、前述の3通りの方法が正規性を検定するにあたって今後、積極的に計量経済学の標準的方法として採用すべき方法であることを提唱したい。

1 正規分布より長い尾をもつ分布

正規分布は歪度 $\sqrt{\beta_1} = 0$ 、尖度 $\beta_2 = 3$ をもつ。

ここで

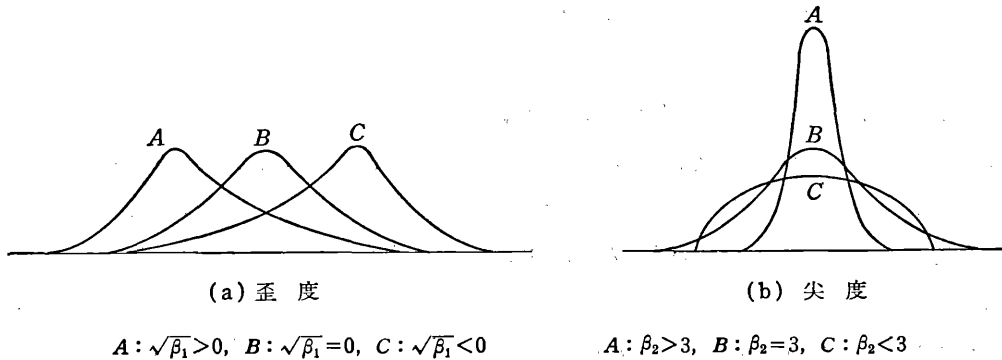
$$\sqrt{\beta_1} = \frac{\mu_3}{\sigma^3}$$

$$\beta_2 = \frac{\mu_4}{\sigma^4} \tag{1}$$

μ_j = 期待値のまわりの j 次モーメント

σ = 標準偏差 である。

図 1 歪度と尖度の異なる分布



正規分布からの乖離は、分布の非対称性が $\sqrt{\beta_1}$ 、すその長さが β_2 によって表される(図1参照)。分布の非対称性は歪度によって示され、正規分布は期待値を中心に対称的である。 $\sqrt{\beta_1} > 0$ の分布は正の歪み、 $\sqrt{\beta_1} < 0$ の分布は負の歪みをもつ分布といわれる。

分布のすその長さは β_2 によって表され、 $\beta_2 > 3$ の分布は長い(あるいは厚い)すそをもつ分布、 $\beta_2 < 3$ は薄いすそをもつ分布といわれている。 $\beta_2 > 3$ をもつ分布は急尖的分布 leptokurtic、 $\beta_2 < 3$ の分布は緩尖的分布 platykurtic、 $\beta_2 = 3$ の分布は中尖的分布 mesokurtic とよばれ、 β_2 が分布のとがりを表すといわれる。しかし $\beta_2 > 3$ の分布がつねに図1(b)のAのようにとがっているわけではなく、あるいは $\beta_2 < 3$ の分布がつねに図1(b)のCのようにとがりの小さい分布を表すわけではない(図2~図7参照)。 β_2 は期待値の近辺における分布のとがり具合というよりは、むしろ期待値から離れた両すそのところで確率密度がどれくらい早く0に近づいていくかを示している。この確率が相対的にゆっくり0に近づいていく分布が長い尾をもつ分布である。急尖的分布という誤解を与える名称ではなく、すそ長分布の方が良い。さらに注意すべきことは $\beta_2 > 3$ の分布が図2~図7のような分布とは限らず、図15のジョンソン無限分布 $SU(0, 2)$ の β_2 も 4.51 と3より大きい。分散が 0.324 と小さいから図15のような形状になる。

正規分布より長いすそをもつ分布にはロジスティック、ラプラス、コーシー、混合正規、最小好適 least favorable distribution (以下、l. f. d. と略す)、スラッシュ slash distribution 等々がある。これらの分布の確率密度関数 p. d. f. は以下の通りである。

ロジスティック分布 Logistic distribution

$$f(x; \theta, \lambda) = \frac{(1/\lambda)e^{-(x-\theta)/\lambda}}{[1+e^{-(x-\theta)/\lambda}]^2} \quad (2)$$

$$\sqrt{\beta_1} = 0, \beta_2 = 4.2$$

ラプラス分布 Laplace distribution

$$f(x; \theta, \lambda) = \frac{1}{2\lambda} e^{-|x-\theta|/\lambda} \quad (3)$$

$$\sqrt{\beta_1} = 0, \beta_2 = 6.0$$

図 2 ロジスティック分布 ($\theta=0, \lambda=1$), $\sqrt{\beta_1}=0, \beta_2=4.2$

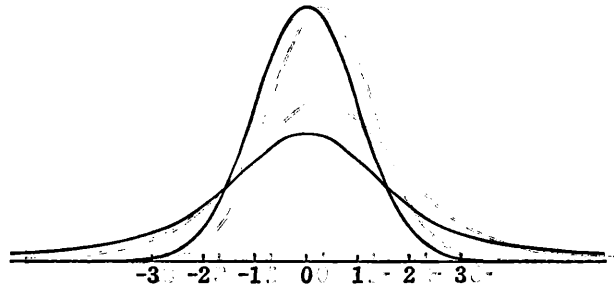


図 3 ラプラス分布 ($\theta=0, \lambda=1$), $\sqrt{\beta_1}=0, \beta_2=6.0$

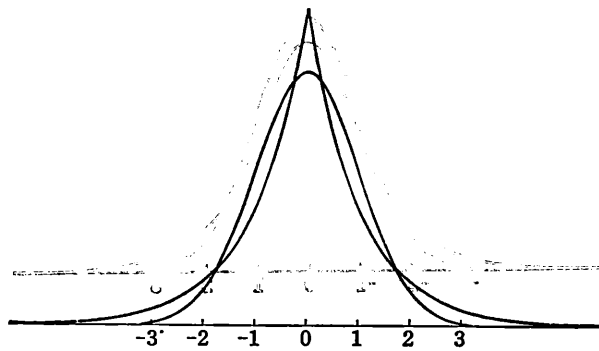


図 4 コーシー分布 ($\theta=0, \lambda=1$)

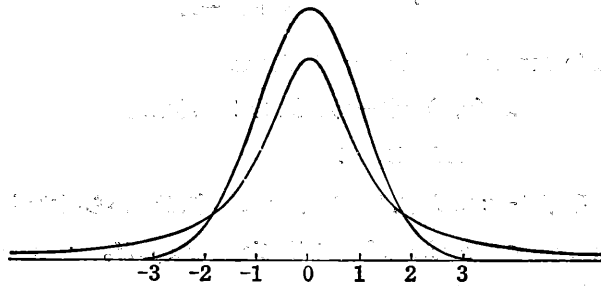


図 5 尺度混合正規分布 ($\alpha=0.15, k=3$), $\sqrt{\beta_1}=0, \beta_2=8.06$

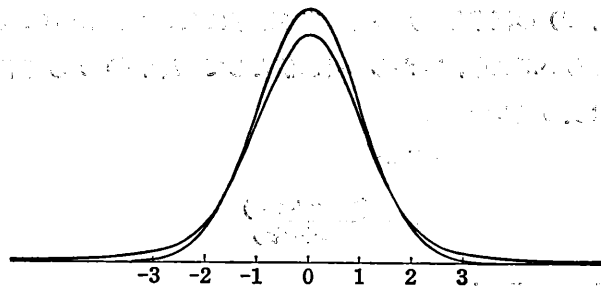


図 6 スラッシュ分布 ($\theta=0, \lambda=1$)

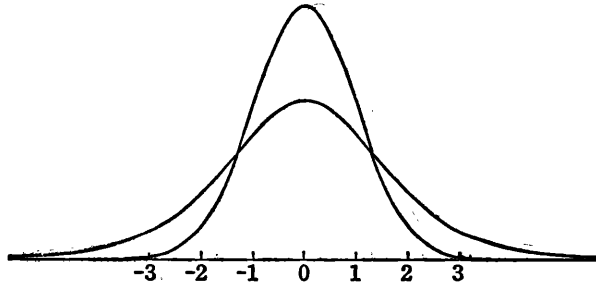
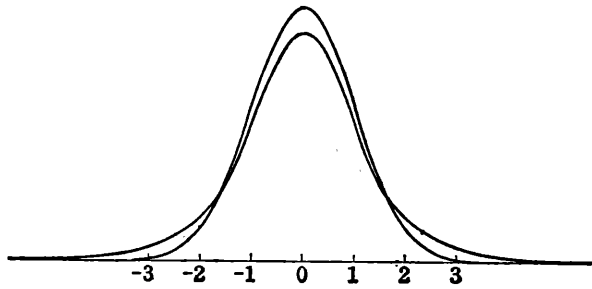


図 7 最小好適分布 ($\epsilon=0.1, k=1.14$), $\sqrt{\beta_1}=0, \beta_2=5.09$



コーシー分布 Cauchy distribution

$$f(x; \theta, \lambda) = \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x - \theta)^2} \quad (4)$$

尺度混合正規分布 Scale contaminated normal distribution

$$CN(\alpha, k) = (1 - \alpha)N(0, 1) + \alpha N(0, k^2) \quad (5)$$

$$0 \leq \alpha \leq 1, k > 1$$

この分布においては、確率 $1 - \alpha$ で $N(0, 1)$ 、確率 α で $N(0, k^2)$ から標本が発生する。この $CN(\alpha, k)$ を Tukey は混合分布 contaminated distribution とよんだ。 α は混合率 contaminated fraction, $N(0, k^2)$ から発生する標本観測値を混合値 contaminants という。たとえば $n=30, \alpha=0.15, k=3$ のとき $N(0, k^2)$ から期待される観測値の数は $30 \times 0.15 = 4.5$ 個。そして分布 $N(0, k^2)$ で $\mu \pm 3$ (3 は $N(0, k^2)$ の標準偏差) の外に出る確率は 0.32 であるから、 $4.5 \times 0.32 = 1.44$ 、すなわち 1 個から 2 個ぐらいが $N(0, 1)$ からの外れ値として $N(0, k^2)$ から期待される個数である。この分布の歪度と尖度は次の通りである。

$$\begin{aligned} \sqrt{\beta_1} &= 0 \\ \beta_2 &= \frac{3(1 - \alpha + \alpha k^4)}{(1 - \alpha + \alpha k^2)^2} \end{aligned} \quad (6)$$

スラッシュ分布 Slash distribution

$$f(x; \theta, \lambda) = \begin{cases} \frac{1 - e^{-(x-\theta)^2/\lambda^2}}{[(x-\theta)^2/\lambda^2] \sqrt{2\pi\lambda}} & x \neq \theta \\ \frac{1}{2\lambda\sqrt{2\pi}} & x = \theta \end{cases} \quad (7)$$

スラッシュ分布は、 $z \sim N(0, 1)$ 、 y は $(0, 1)$ の一様分布にしたがう確率変数のとき、 $x = z/y$ の分布として得られる。

最小好適分布 Least favorable distribution (l. f. d.)

$$f(x) = \begin{cases} \frac{1-\varepsilon}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} & |x| \leq k \\ \frac{1-\varepsilon}{\sqrt{2\pi}} e^{\frac{k^2}{2} - k|x|} & |x| > k \end{cases} \quad (8)$$

$$\sqrt{\beta_1} = 0$$

$$\beta_2 = \frac{\left\{ 2\phi(k) \left(k + \frac{12}{k} + \frac{24}{k^3} + \frac{24}{k^5} \right) + 3 \left[1 - 2\Phi(-k) \right] \right\}}{(1-\varepsilon) \left\{ 1 - 2\Phi(-k) + 4\phi(k) \left(\frac{1}{k} + \frac{1}{k^3} \right) \right\}^2}$$

$$\phi(k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{k^2}{2}}$$

$\Phi(\cdot)$ は標準正規変数の分布関数

これらの分布の確率密度関数を標準正規分布とともに図2から図7に示した。標準正規分布の分布関数を $\Phi(x)$ 、任意の分布関数を $F(x)$ とし、

$$P(x > t) = 1 - F(t) = \alpha$$

を満たす t の値を $\alpha = 0.25, 0.10, 0.05, 0.025, 0.01, 0.005$ について求め、右片側 α を与える確率の比

$$l = \frac{1 - F(x)}{1 - \Phi(x)}$$

をその長さを測る尺度として表1に示した。また $p = F(x)$ を満たす x の値 (変位値 quantile) を $p = 0.95, 0.975, 0.990, 0.995$ について計算した表が表2である。変位値は次式によって求めた。

ロジスティック

$$x = \theta - \lambda \log \left(\frac{1-p}{p} \right)$$

ラプラス

$$x = \theta - \lambda [\log 2 + \log(1-p)]$$

コーシー

$$x = \theta + \lambda \tan \left[\pi \left(p - \frac{1}{2} \right) \right]$$

表 1 累積確率 $F(x)=P(X \leq x)$ と上側確率の標準正規分布との比 l

分布	x	0.6745	1.2816	1.645	1.960	2.3263	2.5758
(1) 正 規	規	0.75	0.90	0.950	0.975	0.990	0.995
	l	1.0	1.0	1.0	1.0	1.0	1.0
(2) ロジスティック	規	0.6625	0.7827	0.8382	0.8765	0.9110	0.9293
	l	1.350	2.173	3.236	4.940	8.90	14.14
(3) ラ プ ラ ス	規	0.7453	0.8612	0.9035	0.9296	0.9512	0.9620
	l	1.019	1.388	1.930	2.816	4.88	7.60
(4) コ ー シ ー	規	0.6889	0.7891	0.8261	0.8498	0.8708	0.8821
	l	1.244	2.109	3.478	6.008	12.92	23.58
(5) 尺度混合正規							
① 0.05, 3	規	0.7419	0.8883	0.9379	0.9634	0.9795	0.9855
	l	1.032	1.117	1.242	1.464	2.045	2.903
② 0.10, 3	規	0.7334	0.8765	0.9258	0.9518	0.9691	0.9760
	l	1.064	1.235	1.484	1.927	3.090	4.805
③ 0.15, 3	規	0.7258	0.8648	0.9137	0.9402	0.9586	0.9665
	l	1.097	1.352	1.725	2.391	4.135	6.708
④ 0.05, 5	規	0.7402	0.8851	0.9340	0.9589	0.9745	0.9801
	l	1.039	1.149	1.321	1.645	2.554	3.982
⑤ 0.10, 5	規	0.7304	0.8701	0.9179	0.9427	0.9589	0.9652
	l	1.079	1.299	1.642	2.290	4.108	6.964
⑥ 0.15, 5	規	0.7205	0.8552	0.9018	0.9266	0.9434	0.9503
	l	1.118	1.448	1.963	2.935	5.662	9.946
⑦ 0.05, 10	規	0.7388	0.8826	0.9308	0.9551	0.9701	0.9753
	l	1.045	1.175	1.385	1.795	2.990	4.933
⑧ 0.10, 10	規	0.7277	0.8651	0.9115	0.9353	0.9502	0.9557
	l	1.089	1.349	1.769	2.589	4.980	8.867
⑨ 0.15, 10	規	0.7165	0.8477	0.8923	0.9154	0.9303	0.9360
	l	1.134	1.524	2.154	3.384	6.970	12.800
(6) スラッシュ	規	0.6297	0.7257	0.7702	0.8013	0.8300	0.8457
	l	1.481	2.744	4.597	7.950	17.001	30.851
(7) l.f.d.							
① 0, 1.345	規	0.75	0.90	0.9198	0.9475	0.9679	0.9771
	l	1.0	1.0	1.604	2.10	3.207	4.586
② 0.01, 1.945	規	0.7475	0.8960	0.9455	0.9703	0.9854	0.9910
	l	1.01	1.04	1.09	1.19	1.459	1.796
③ 0.1, 1.14	規	0.7250	0.8601	0.9075	0.9354	0.9575	0.9680
	l	1.1	1.400	1.85	2.583	4.253	6.400
④ 0.15, 0.98	規	0.7125	0.8407	0.8884	0.9181	0.9428	0.9552
	l	1.15	1.593	2.232	3.278	5.721	8.961

注(1) $l = \frac{1-F(x)}{1-\Phi(x)}$, $\Phi(x)$ は標準正規変数の分布関数。

(2) 尺度混合正規分布は α, k の値が示されている。

(3) l.f.d. は ϵ, k の値が示されている。

尺度混合正規分布

x の値はくりかえし計算で求めた。

最小好適分布 (l.f.d.)

表 2 変位値および95%, 99%区間幅

分布	p	0.95	0.975	0.990	0.995	95%区間幅	99%区間幅
(1) 正 規		1.6450	1.9600	2.3263	2.5758	3.9200	5.1516
(2) ロジスティック		2.9444	3.6636	4.5951	5.2933	7.3272	10.5866
(3) ラ プ ラ ス		2.3026	2.9957	3.9120	4.6052	5.9914	9.2104
(4) コ ー シ ー		6.3138	12.7062	31.8205	63.6567	25.4124	127.3134
(5) 尺度混合正規							
① 0.05, 3		1.7742	2.1944	2.9127	3.8614	4.3888	7.7228
② 0.10, 3		1.9322	2.5343	3.8527	4.9333	5.0686	9.8666
③ 0.15, 3		2.1292	2.9953	4.5026	5.5013	5.9906	11.0026
④ 0.05, 10		1.8774	2.5540	8.4097	12.8009	5.1080	25.6018
⑤ 0.10, 10		2.3184	6.7377	12.8105	16.4440	13.4754	32.8880
(6) l.f.d.							
① 0, 1.345		1.9962	2.5116	3.1928	3.7082	5.0232	7.4164
② 0.01, 1.945		1.6910	2.0494	2.5205	2.8769	4.0988	5.7538
③ 0.1, 1.14		2.1844	2.7924	3.5962	4.2042	5.5848	8.4084
④ 0.15, 0.98		2.4640	3.1713	4.1062	4.8135	6.3426	9.6270

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

図2～図7および表1, 表2をみることによってその長さ, 右片側確率および変位値の標準正規分布との相違を知ることができる。

たとえば尺度混合正規分布 $CN(0.15, 3)$ を例にとれば, 確率変数 X が2.3263より大きい値をとる確率は, 標準正規分布が0.01であるのに対して0.0414ある。正規分布のもとではただか1%の発生可能性しかもたない「外れ値」は, $CN(0.15, 3)$ のもとでは4.14%もあることがわかる(表1)。あるいは表2より右片側1%点は標準正規分布2.3263に対して $CN(0.15, 3)$ は4.5026であることがわかる。

いずれにしても, もし誤差項の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」と判断される値が, 必ずしも外れ値ではないということに注意しなければならない。

2 正規性の仮定と位置パラメータの推定

無作為標本 X_1, X_2, \dots, X_n にもとづいて期待値 μ を最小2乗法で推定すれば標本平均 \bar{X} が得られる。この最小2乗推定量 (OLSE) \bar{X} は, 正規母集団の仮定と正則条件のもとで, 期待値 μ の有効推定量となることはよく知られている。同じ結論は線形回帰モデルの回帰係数 β の OLSE $\hat{\beta}$ に対しても成り立つ (以下の \bar{X} に関する議論は $\hat{\beta}$ についても適用することができる)。

母集団分布がいかなる分布であろうと、したがって正規分布をしていなくても、OLSEは最良線形不偏推定量BLUEを与える。最小2乗法を適用することによって得られる \bar{X} や $\hat{\beta}$ が推定量の王座を占めてきた理由がここにある。

しかし X_1, \dots, X_n (あるいは線形回帰モデルの誤差項)の分布が正規分布より長い尾をもつ分布の場合には、標本データに非常に大きな(あるいは小さな)値が現われる可能性が大きくなる。すなわち、正規分布のもとでは「外れ値」とみなされる値は、長い尾をもつ分布のもとでは、十分大きな確率で発生し得る値であって決して「外れ値」ではない。そしてこの「外れ値」に対して \bar{X} や $\hat{\beta}$ は敏感に反応し、大きな影響を受け、分散が大きくなり、有効性が低下し、もはやOLSEは最良の推定量ではない。

正規母集団においてはOLSEとくらべて有効性の低下は少なく、長い尾をもつ母集団分布の場合には「外れ値」に対して敏感に反応せず頑健robustであり、 \bar{X} より高い有効性を与える推定量が頑健推定量である。

頑健推定に関する有名なプリンストン研究[1]のモンテカルロ実験の結果を示したのが表3である。表のHuber, Hampel, Andrewsとはそれぞれの ψ 関数による μ のM推定量を意味する。

Huber

$$\psi(x) = \begin{cases} x & |x| \leq H \\ H & |x| > H \end{cases} \quad (9)$$

$$H=1.5, s=MAD/0.6745$$

Hampel

$$\psi(x) = \begin{cases} x & |x| \leq a \\ a \cdot \text{sgn}(x) & a \leq |x| \leq b \\ \frac{a[c-|x|]}{c-b} & b \leq |x| \leq c \\ 0 & |x| > c \end{cases} \quad (10)$$

$$a=2.5, b=4.5, c=9.5, s=MAD$$

Andrews

$$\psi(x) = \begin{cases} \sin(x) & |x| \leq \pi \\ 0 & \text{その他} \end{cases} \quad (11)$$

$$s=2.1MAD$$

表3から次の点は明らかであろう。

- (1)平均 \bar{X} が高い有効性をもつのは母集団分布が正規分布の場合だけである。
- (2)母集団分布が正規分布より長い尾をもつ分布の場合、 \bar{X} の有効性は著しく低下する。
- (3)母集団分布が正規分布の場合のM推定量の有効性は \bar{X} より若干低くなるが、長い尾をもつ分布の場合には \bar{X} より有効性が高い。しかもこのM推定量の有効性は母集団分布が異なっても \bar{X} のように大きく崩れない。

表 3 モンテ・カルロ実験による μ の推定量の分散 ($n=20$)

母集団分布 \ 推定量	平均	メジアン	Huber	Hampel	Andrews _t
正規	1.00	1.498	1.050	1.046	1.070
CN (0.05, 3)	1.420	1.516	1.169	1.156	1.163
CN (0.15, 3)	2.259	1.747	1.474	1.487	1.467
CN (0.05, 10)	6.484	1.555	1.222	1.127	1.131
CN (0.10, 10)	11.084	1.804	1.491	1.259	1.245
ラプラス(二重指数)	2.10	1.37	1.55	1.58	1.54
t (3)	3.138	1.817	1.673	1.671	1.643
コーシー	12,548.0	2.9	4.5	3.7	3.5

出所: Andrews et al. [1] Exhibit 5 より作成。 t (3) は自由度 3 の t 分布

Huber の ϕ を用いる M 推定量を例にして、モンテカルロ実験ではなく、理論的にもう少し上記の点を検討しておこう。問題にするのは次の点である。

(1) 母集団分布が正規分布のとき、期待値 μ の M 推定量の有効性はどの程度損なわれるのか。M 推定量を T とし、正規分布を ϕ で表せば、 T の有効性の損失は

$$A = \frac{\text{Var}(T|\phi)}{\text{Var}(\bar{X}|\phi)} \quad (12)$$

によって測ることができる。 $A-1$ はプレミアム premium とよばれることもある (Anscombe [2])。 A が 1 に近いほど、あるいはプレミアムが小さいほど推定量 T の有効性の損失は小さい。

(2) 母集団分布が正規分布より長い尾をもつ分布のとき、M 推定量 T の有効性は \bar{X} にくらべてどの程度大きくなるか。非正規分布を F で表せば、 T の有効性の増加は

$$B = \frac{\text{Var}(T|F)}{\text{Var}(\bar{X}|F)} \quad (13)$$

によって測ることができる。 $1-B$ は Anscombe [2] によってプロテクション protection とよばれた概念である。 B が 1 より小さいほど、あるいはプロテクションが正の大きな値をとるほど \bar{X} にくらべて T の有効性は大きい。

火災保険を例にとれば、掛金 (プレミアム) が安く、火災が発生したとき (外れ値が生じたとき)、補償範囲 (プロテクション) が高い保険 (推定量) ほど良いということになる。すなわち、 T が良い頑健推定量であるとは、 A が 1 より余り大きくなく (プレミアムが 0 に近く) かつ B が 1 より十分小さい (プロテクションが大きい) 場合である。

T を Huber の ϕ による M 推定量とすれば、真の分布が標準正規分布の場合の T の漸近的分散は次式で与えられる (蓑谷[26](63) 式) (以下 $n \times$ 分散が Var で示されている)。

$$\text{Var}(T|\phi) = \frac{-2H\phi(H) + 1 - 2\phi(-H) + 2H^2\phi(-H)}{[1 - 2\phi(-H)]^2} \quad (14)$$

ここで ϕ は標準正規分布の p. d. f., ϕ は d. f., H は調整定数である。

正規分布より長い尾をもつ分布として、l. f. d. が F のとき次式を得る (数学注参照)。

図 8 1.f.d. における \bar{X} と M 推定量 (Huber の ϕ) T の有効性

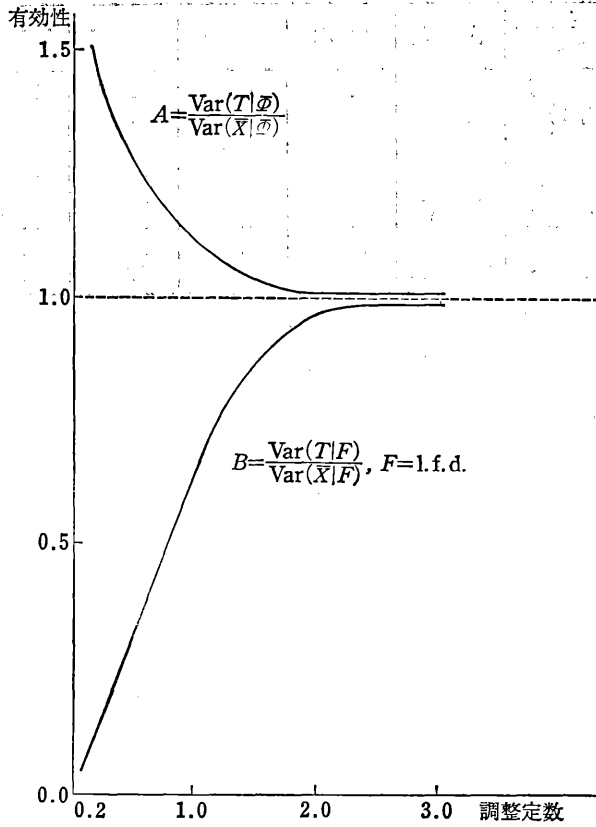


図 9 1.f.d. における Huber の ϕ 関数による M 推定量のプレミアムとプロテクション

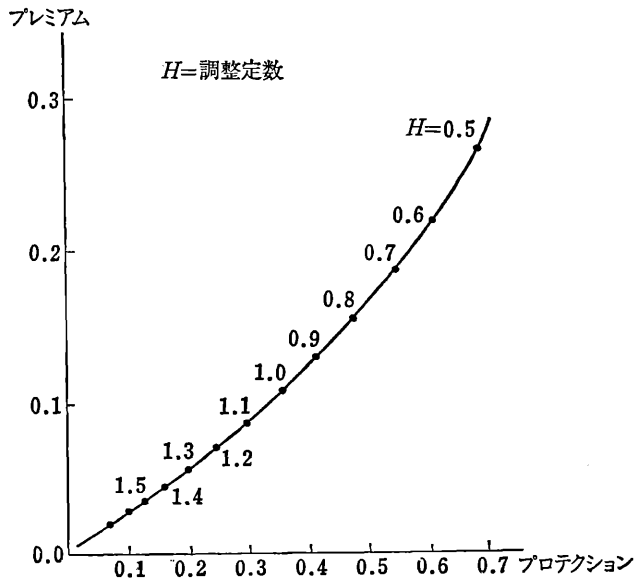


表 4 l.f.d.におけるM推定量 (Huber の ϕ) と \bar{X} の分散, 有効性, プレミアムおよびプロテクション

ε	H	$\text{Var}(\bar{X} F)$	$\text{Var}(T \phi)$	$\text{Var}(T F)$	$\frac{\text{Var}(T F)}{\text{Var}(\bar{X} F)}$	プレミアム	プロテクション
0.8753	0.100	200.0030	1.4923	12.5541	0.0628	0.4923	0.9372
0.7542	0.200	50.0131	1.4232	6.3084	0.1261	0.4232	0.8739
0.6401	0.300	22.2512	1.3624	4.2405	0.1906	0.3624	0.8094
0.5354	0.400	12.5502	1.3091	3.2171	0.2563	0.3091	0.7437
0.4417	0.500	8.0760	1.2625	2.6115	0.3234	0.2625	0.6766
0.3599	0.600	5.6610	1.2219	2.2149	0.3913	0.2219	0.6087
0.2899	0.700	4.2194	1.1866	1.9377	0.4592	0.1866	0.5408
0.2311	0.800	3.2971	1.1561	1.7352	0.5263	0.1561	0.4737
0.1825	0.900	2.6770	1.1298	1.5826	0.5912	0.1298	0.4088
0.1428	1.000	2.2445	1.1073	1.4648	0.6526	0.1073	0.3474
0.1109	1.100	1.9342	1.0880	1.3724	0.7095	0.0880	0.2905
0.0855	1.200	1.7070	1.0718	1.2989	0.7609	0.0718	0.2391
0.0655	1.300	1.5380	1.0580	1.2401	0.8063	0.0580	0.1937
0.0498	1.400	1.4106	1.0466	1.1926	0.8454	0.0466	0.1546
0.0376	1.500	1.3139	1.0371	1.1542	0.8785	0.0371	0.1215
0.0282	1.600	1.2400	1.0293	1.1231	0.9057	0.0293	0.0943
0.0211	1.700	1.1833	1.0229	1.0979	0.9278	0.0229	0.0722
0.0156	1.800	1.1397	1.0178	1.0774	0.9454	0.0178	0.0546
0.0115	1.900	1.1061	1.0137	1.0609	0.9592	0.0137	0.0408
0.0084	2.000	1.0803	1.0104	1.0477	0.9698	0.0104	0.0302
0.0061	2.100	1.0605	1.0078	1.0371	0.9779	0.0078	0.0221
0.0044	2.200	1.0454	1.0059	1.0286	0.9840	0.0059	0.0160
0.0032	2.300	1.0338	1.0043	1.0219	0.9885	0.0043	0.0115
0.0023	2.400	1.0251	1.0032	1.0167	0.9918	0.0032	0.0082
0.0016	2.500	1.0185	1.0023	1.0126	0.9942	0.0023	0.0058
0.0011	2.600	1.0135	1.0017	1.0094	0.9959	0.0017	0.0041
0.0008	2.700	1.0098	1.0012	1.0070	0.9972	0.0012	0.0028
0.0005	2.800	1.0071	1.0008	1.0051	0.9981	0.0008	0.0019
0.0004	2.900	1.0051	1.0006	1.0037	0.9987	0.0006	0.0013
0.0003	3.000	1.0036	1.0004	1.0027	0.9991	0.0004	0.0009
0.0002	3.100	1.0025	1.0003	1.0019	0.9994	0.0003	0.0006
0.0001	3.200	1.0018	1.0002	1.0014	0.9996	0.0002	0.0004
0.0001	3.300	1.0012	1.0001	1.0010	0.9997	0.0001	0.0003
0.0001	3.400	1.0009	1.0001	1.0007	0.9998	0.0001	0.0002
0.0000	3.500	1.0006	1.0001	1.0005	0.9999	0.0001	0.0001
0.0000	3.600	1.0004	1.0000	1.0003	0.9999	0.0000	0.0001
0.0000	3.700	1.0003	1.0000	1.0002	1.0000	0.0000	0.0000
0.0000	3.800	1.0002	1.0000	1.0001	1.0000	0.0000	0.0000
0.0000	3.900	1.0001	1.0000	1.0001	1.0000	0.0000	0.0000
0.0000	4.000	1.0001	1.0000	1.0001	1.0000	0.0000	0.0000

注(1) $F=l.f.d.$, $T=Huber$ の ϕ によるM推定量

(2) プレミアム $=\text{Var}(T|\phi)/\text{Var}(\bar{X}|\phi)-1$, プロテクション $=1-\text{Var}(T|F)/\text{Var}(\bar{X}|F)$

(3) $\text{Var}(\bar{X}|\phi)=1$ に基準化してある。

$$\text{Var}(\bar{X}|F) = (1-\varepsilon) \left\{ 1 - 2\phi(-k) + 4\phi(k) \left(\frac{1}{k} + \frac{1}{k^3} \right) \right\} \quad (15)$$

$$\text{Var}(T|F) = \frac{1}{(1-\varepsilon)[1-2\phi(-k)]} \quad (16)$$

F が $CN(\alpha, k)$ の場合には次式が得られる。

表 5 尺度混合正規分布におけるM推定量 (Huber の ϕ) と \bar{X} の分散, 有効性, プレミアムおよび
プロテクション

H	$E(\phi^2)$	$E(\phi)$	$\text{Var}(\bar{X} F)$	$\text{Var}(T F)$	$\frac{\text{Var}(T F)}{\text{Var}(\bar{X} F)}$	プレミアム	プロテクション
0.500	0.1915437291	0.3453414578	2.20000	1.60609	0.73004	0.26254	0.26996
0.600	0.2600015596	0.4075476004	2.20000	1.56538	0.71154	0.22193	0.28846
0.700	0.3330622085	0.4663362659	2.20000	1.53154	0.69615	0.18665	0.30385
0.800	0.4088429272	0.5213867506	2.20000	1.50396	0.68362	0.15612	0.31638
0.900	0.4857225692	0.5724710409	2.20000	1.48211	0.67369	0.12982	0.32631
1.000	0.5623459228	0.6194535897	2.20000	1.46550	0.66614	0.10727	0.33386
1.100	0.6376191845	0.6622875973	2.20000	1.45368	0.66076	0.08804	0.33924
1.200	0.7106975564	0.7010082527	2.20000	1.44623	0.65738	0.07176	0.34262
1.300	0.7809662827	0.7357235241	2.20000	1.44279	0.65581	0.05805	0.34419
1.400	0.8480166639	0.7666031726	2.20000	1.44299	0.65590	0.04659	0.34410
1.500	0.9116186833	0.7938666938	2.20000	1.44650	0.65750	0.03709	0.34250
1.600	0.9716918686	0.8177708750	2.20000	1.45300	0.66045	0.02928	0.33955
1.700	1.0282758904	0.8385975948	2.20000	1.46219	0.66463	0.02291	0.33537
1.800	1.0815022032	0.8566423956	2.20000	1.47377	0.66989	0.01777	0.33011
1.900	1.1315677838	0.8722042450	2.20000	1.48746	0.67612	0.01365	0.32388
2.000	1.1787117410	0.8855767755	2.20000	1.50299	0.68318	0.01039	0.31682
2.100	1.2231952922	0.8970411631	2.20000	1.52010	0.69095	0.00784	0.30905
2.200	1.2652853344	0.9068606933	2.20000	1.53853	0.69933	0.00585	0.30067
2.300	1.3052416120	0.9152769597	2.20000	1.55807	0.70821	0.00433	0.29179
2.400	1.3433072928	0.9225075645	2.20000	1.57847	0.71748	0.00317	0.28252
2.500	1.3797026279	0.9287451332	2.20000	1.59953	0.72706	0.00230	0.27294
2.600	1.4146212781	0.9341574221	2.20000	1.62106	0.73685	0.00165	0.26315
2.700	1.4482288423	0.9388882871	2.20000	1.64289	0.74677	0.00118	0.25323
2.800	1.4806631173	0.9430592828	2.20000	1.66486	0.75676	0.00083	0.24324
2.900	1.5120356349	0.9467716808	2.20000	1.68683	0.76674	0.00058	0.23326
3.000	1.5424340673	0.9501087229	2.20000	1.70868	0.77667	0.00040	0.22333
3.100	1.5719251500	0.9531379542	2.20000	1.73030	0.78650	0.00027	0.21350
3.200	1.6005578316	0.9559135155	2.20000	1.75160	0.79618	0.00019	0.20382
3.300	1.6283664281	0.9584783081	2.20000	1.77251	0.80568	0.00013	0.19432
3.400	1.6553736185	0.9608659705	2.20000	1.79296	0.81498	0.00009	0.18502
3.500	1.6815931734	0.9631026336	2.20000	1.81291	0.82405	0.00006	0.17595
3.600	1.7070323545	0.9652084418	2.20000	1.83231	0.83287	0.00004	0.16713
3.700	1.7316939602	0.9671988406	2.20000	1.85114	0.84143	0.00003	0.15857
3.800	1.7555780174	0.9690856458	2.20000	1.86937	0.84971	0.00002	0.15029
3.900	1.7786831421	0.9708779128	2.20000	1.88699	0.85772	0.00001	0.14228
4.000	1.8010075998	0.9725826310	2.20000	1.90398	0.86545	0.00001	0.13455
4.100	1.8225501060	0.9742052691	2.20000	1.92034	0.87288	0.00001	0.12712
4.200	1.8433104059	0.9757501951	2.20000	1.93607	0.88003	0.00001	0.11997
4.300	1.8632896729	0.9772209948	2.20000	1.95117	0.88689	0.00000	0.11311
4.400	1.8824907616	0.9786207084	2.20000	1.96564	0.89347	0.00000	0.10653
4.500	1.9009183462	0.9799520039	2.20000	1.97949	0.89977	0.00000	0.10023
4.600	1.9185789712	0.9812173015	2.20000	1.99273	0.90579	0.00000	0.09421
4.700	1.9354810364	0.9824188598	2.20000	2.00537	0.91153	0.00000	0.08847
4.800	1.9516347339	0.9835588363	2.20000	2.01743	0.91701	0.00000	0.08299
4.900	1.9670519506	0.9846393258	2.20000	2.02890	0.92223	0.00000	0.07777
5.000	1.9817461477	0.9856623858	2.20000	2.03982	0.92719	0.00000	0.07281

注(1) $F=CN(0.15, 3)$, $T=Huber$ の ϕ によるM推定量。

(2) $\text{Var}(T|\phi)$ は表4に同じ。

(3) $\text{Var}(\bar{X}|\phi)=1$ に基準化してある。

$$\text{Var}(\bar{X}|F) = 1 - \alpha + \alpha k^2 \quad (17)$$

$$\text{Var}(T|F) = \frac{E(\phi^2)}{[E(\phi')]^2} \quad (18)$$

$$\begin{aligned} E(\phi^2) &= (1-\alpha)[-2H\phi(H) + 1 - 2\phi(-H)] \\ &\quad + \alpha k^2 \left[-2\left(\frac{H}{k}\right)\phi\left(\frac{H}{k}\right) + 1 - 2\phi\left(-\frac{H}{k}\right) \right] \\ &\quad + 2(1-\alpha)H^2\phi(-H) + 2\alpha H^2\phi\left(-\frac{H}{k}\right) \end{aligned}$$

$$E(\phi') = (1-\alpha)[1 - 2\phi(-H)] + \alpha \left[1 - 2\phi\left(-\frac{H}{k}\right) \right]$$

以上の結果を用いて、 $T = \text{Huber}$ の ϕ による M 推定量、 $F = \text{l. f. d.}$ のときの A, B 、プレミアムおよびプロテクションを計算したのが表 4、グラフが図 8、図 9 に示されている。同様に $F = \text{CN}(0.15, 3)$ の場合は表 5、 $\text{CN}(0.15, 3)$ と $\text{CN}(0.05, 5)$ のケースのグラフが図 10、図 11 に示されている。

まず $F = \text{l. f. d.}$ の場合をみてみよう (表 4, 図 8, 図 9)。l. f. d. のもとで Huber の ϕ を用いる M 推定量は最尤推定量である (証明は養谷 [26] をみよ)。表およびグラフより次のことがわかる。

(1) 調整定数 H の値いかにかわらず、 $A = \text{Var}(T|\phi) / \text{Var}(\bar{X}|\phi)$ の値は 1 以上であり、 T の有効性は \bar{X} より低い。しかし H の値が 1 以上となれば T の有効性の低下はわずかである。

(2) 調整定数 H の値いかにかわらず、 $B = \text{Var}(T|F) / \text{Var}(\bar{X}|F)$ の値は 1 以下であり、 $F = \text{l. f. d.}$ が正しいとき \bar{X} の有効性は T より低い。

(3) プレミアムとプロテクションはトレード・オフの関係にある。調整定数 H の値を大きくしてプレミアムを小さくしようとすればプロテクションも小さくなる。プロテクションを大きくしようとすれば H の値を小さくしなければならないが、そのときプレミアムも大きくなる。

次に $F = \text{CN}(\alpha, k)$ の場合をみてみよう (表 5, 図 10, 図 11)。 $F = \text{l. f. d.}$ のときの前述の (1), (2) については $F = \text{CN}(\alpha, k)$ においても同じであるが (3) は成立しない。図 11 からわかるように、調整定数 H の値を 0.5 から次第に大きくしていくことによってプレミアムは小さくなり、プロテクションは大きくなるという望ましい状態が続く。しかし $\text{CN}(0.15, 3)$ のとき $H = 1.3$ 、 $\text{CN}(0.05, 5)$ のとき $H = 1.6$ でプロテクションは最大になり、この値より大きくなるとプレミアムは小さくなるがプロテクションも小さくなる。

母集団分布が正規分布より長い尾をもつ分布であることが正規性の検定によって検出されても、われわれにはその長い尾をもつ分布が l. f. d. なのか $\text{CN}(\alpha, k)$ なのかあるいはそれ以外の分布なのかわからない。しかしこの $\text{CN}(\alpha, k)$ のような場合を考えると、Huber の ϕ を用いる M 推定において、調整定数 H の値を 1 から 2 の間で選択すると良いと言われてきたことが正当化される。

さて以上の検討によって明らかになったことは、 μ の推定量のなかで最小 2 乗推定量 \bar{X} が最良で

図10 尺度混合正規分布における Huber の ψ によるM推定量と \bar{X} の有効性

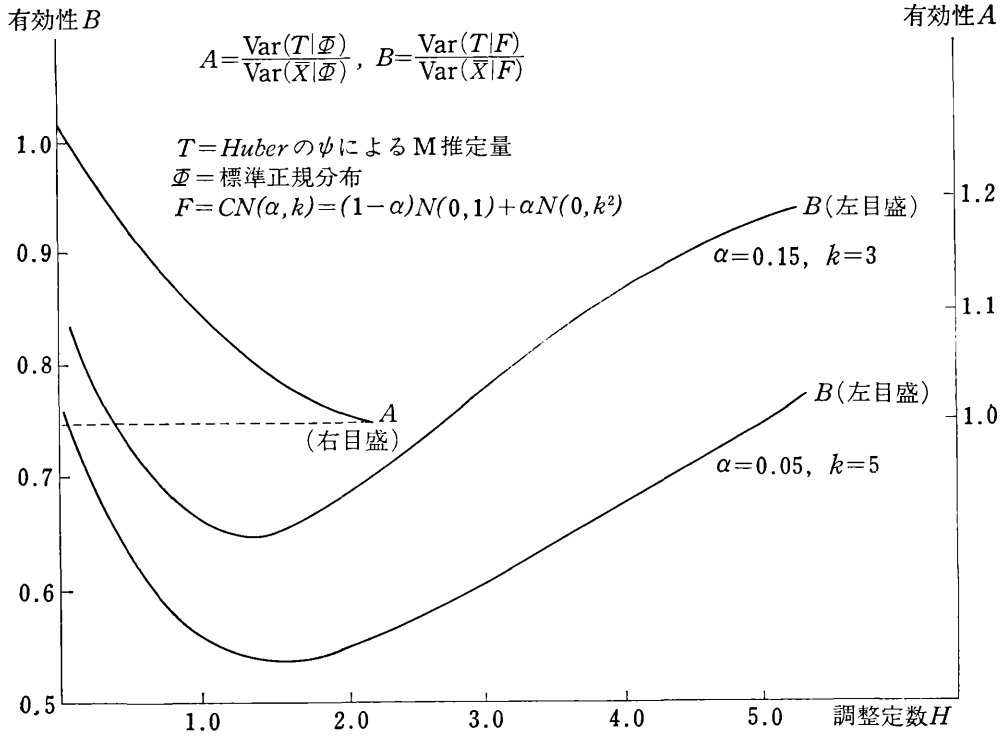
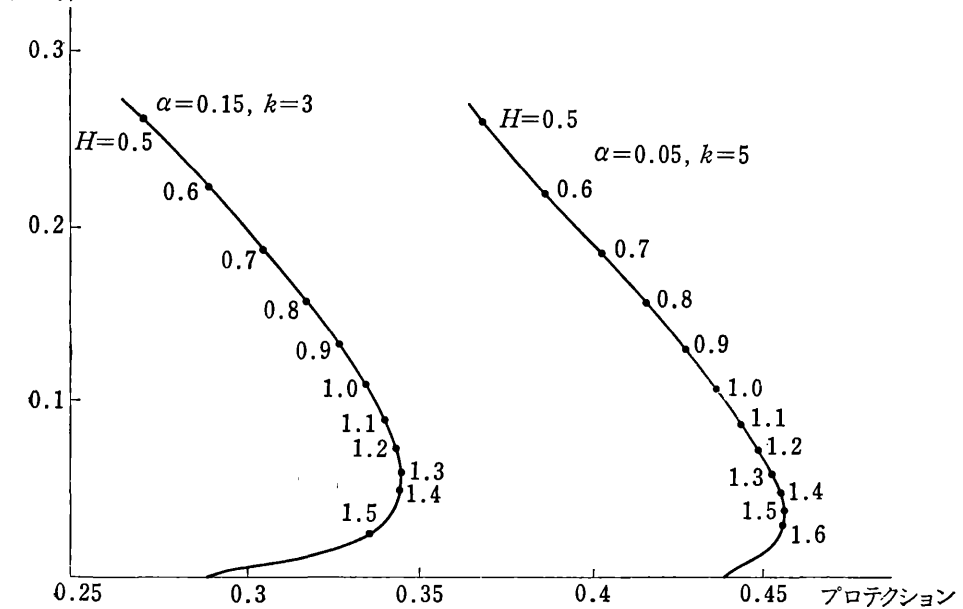


図11 尺度混合正規分布における Huber の ψ 関数によるM推定量のプレミアムとプロテクション



あるという特性は正規分布と密接に関連しているということである。l. f. d. のように p. d. f. $f(x)$ の範囲 range がパラメータ k に依存しており、クラメル・ラオ不等式の正則条件の 1 つが満たされない分布の場合には、 \bar{X} の分散より小さい分散をもつ不偏推定量が存在する。 μ の推定量を求めるとき正規分布の仮定と OLSE の \bar{X} にのみ抱泥しているべきではない。

\bar{X} に関して述べてきたことは線形回帰モデルの係数パラメータ β の OLSE である $\hat{\beta}$ についても成立する。Hill and Holland [20] は、回帰モデルの誤差項の分布に $CN(\alpha, k)$ を仮定し、モンテカルロ実験によって β の OLSE $\hat{\beta}$ 、最小絶対誤差 (LAR) 推定量 $\hat{\beta}_L$ および Andrews の ψ を用いる M 推定量 $\hat{\beta}_s$ の有効性を比較した。 $\hat{\beta}_j$ $j=L, S$ の非有効性 inefficiency を

$$\text{iff}(\hat{\beta}_j) = \frac{E\|\hat{\beta}_j - \beta\|}{E\|\hat{\beta} - \beta\|}$$

$\|\cdot\|$ はノルム

によって表し、 $\text{iff}(\hat{\beta}_j)$ $j=L, S$ が 1 より小さくなることを彼等は示した。彼等のモンテカルロ実験の主要な結果は次の 5 点である。

- ① $\hat{\beta}_L, \hat{\beta}_s$ とも $\text{iff}(\hat{\beta}_j)$ で判断して $\hat{\beta}$ よりきわめて良い。
- ② 分布の尾が長くなるほど $\hat{\beta}$ は $\hat{\beta}_L$ や $\hat{\beta}_s$ より著しく悪くなる。
- ③ 説明変数の高い作用点 high leverage point と結びついた大きな誤差 (外れ値) は $\hat{\beta}_L, \hat{\beta}_s$ の $\hat{\beta}$ に対する有効性を低下させる。
- ④ ほとんどのケースにおいて $\hat{\beta}_s$ の方が $\hat{\beta}_L$ より良い。
- ⑤ n が大きくなれば、 $\hat{\beta}_L, \hat{\beta}_s$ の外れ値に対する頑健性は容易に達成される。

位置パラメータの M 推定量のみを考察してきたが、 L 推定量の有効性も尾の長い分布のもとでは \bar{X} より高くなることは Rosenberger and Gasko [28] に詳しい。

3 正規性の仮定と位置パラメータに関する検定および信頼区間

正規母集団からの無作為標本 X_1, X_2, \dots, X_n によって母平均 μ に関する仮説検定や信頼区間を設定するとき

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1) \tag{19}$$

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

が用いられる。これまでの理論的研究およびモンテカルロ実験によって明らかとなった検定および信頼区間に関する次の 2 点は重要である。

(1)母集団分布が正規分布から若干離れても、 t 統計量の検定の大きさは大きく異ならない

たとえば Boos [9] はモンテカルロ実験によって、統計量 T の百分位点 $100(1-\alpha)\%$ 点を計算し、この百分位点がスラッシュ分布を除けば、母集団分布の型（ただし表6の非正規分布はすべて対称的な、長い尾をもつ分布である）によって大きく異ならないことを示した。

表 6 $T=(\bar{X}-\mu)/(s/\sqrt{n})$ の百分位点

	t_α	正 規	三 角	ロジス ティック	二重指数	$t(3)$	スラッシュ
$n=20$ $\alpha=0.025$	2.09	2.11	2.08	2.07	2.08	2.02	1.80
$n=20$ $\alpha=0.05$	1.73	1.73	1.71	1.72	1.73	1.70	1.55

注：(1) Boos[9]表2より作成。

(2) t_α は $df=19$ の右片側 $100\times\alpha\%$ 点。

この結果は理論的に確かめることができる。母集団分布が対称のとき、 $H_0:\mu=0$ のもとで T の分布は t 分布で近似することが可能であり、近似は次式で与えられる（柴田[32], p.145）。

$$P(T>t_\alpha)\approx\alpha-\kappa_4\frac{n}{n+2}\cdot\frac{n+1}{n+4}\times\frac{\Gamma(n/2)}{12(n-1)\sqrt{\pi(n-1)}\Gamma[(n-1)/2]}\cdot\frac{t_\alpha^3-3(n-1)t_\alpha/(n+1)}{\{1+t_\alpha^2/(n-1)\}^{(n+2)/2}} \quad (20)$$

ここで κ_4 は 4 次のキュムラントであり

$$\kappa_4=\mu_4-3\mu_2^2=\mu_2^2(\beta_2-3)$$

である。

対称的かつ $\kappa_4>0$ すなわち $\beta_2>3$ の急尖的分布のとき、1%の t_α はつねに $\sqrt{3}$ より大きく、5%の t_α も自由度18以下では $\sqrt{3}$ より大きい。このとき (20) 式から明らかなように

$$P(T>t_\alpha)<\alpha$$

となる。たとえば尺度混合正規分布 $CN(0.15, 3)$ を例にとってみよう。 $CN(0.15, 3)$ は $\beta_2=8.058$ ($\kappa_4=24.48$) の急尖的分布である。 $n=20, \alpha=0.05$ の場合 $CN(0.15, 3)$ の (20) 式右辺第 2 項の値は -0.000001437 にしかならないから $CN(0.15, 3)$ が母集団分布であったとしても、正規分布の仮定のもとで導かれる t 統計量の $P(T>t_\alpha)=0.05$ よりわずかに小さいだけである。いいかえれば、第 I 種の過誤確率 $P(I)$ は、母集団分布が対称的な急尖的分布のとき、正規母集団の仮定のもとで t 統計量を用いて設定される有意水準 α より若干小さくなる。 $P(I)$ が α をこえるならば、 $P(I)$ をコントロールできなくなることを意味するから深刻であるが、そうではなく、 α よりわずかに小さくなるのであるから仮説検定の有意水準が非正規性のもとでくずれてしまうという心配はない（妥当性の頑健 robustness of validity）。

他方、 $|t_\alpha|>\sqrt{3}$ であっても $\kappa_4<0$ すなわち緩尖的分布の場合には

$$P(T > t_\alpha) > \alpha$$

となる。しかし急尖的あるいは緩尖的いずれの分布であっても対称分布であれば $P(T > t_\alpha)$ と α との違いは小さく、 n が30ぐらいあればほとんど無視できる大きさである。

(2)母集団分布が正規分布より長い尾をもつ分布のとき、極端に大きいあるいは小さい値が観測される可能性は高く、これらの値によって \bar{X} の分散が大きくなる。そのため正規分布の仮定のもとで得られる μ の $(1-\alpha) \times 100\%$ 信頼区間

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \quad (21)$$

の区間幅は広くなり、 μ の推定量 \bar{X} の信頼性は低くなる。

このことを仮説検定の観点からみれば、 t 統計量の検定力は小さくなるということである。 $H_0: \mu=0$ を $H_1: \mu \neq 0$ に対して検定する場合を考えてみよう。信頼区間の幅が広がるから、 $H_1: \mu \neq 0$ が正しいにもかかわらず、区間内に0が含まれる可能性は高くなる。すなわち H_1 が正しいときに H_0 を棄却しないという第II種の過誤確率が高くなる、つまり検定力が小さくなる。

パラメータ推定の観点から正規分布より長い尾をもつ母集団分布においては、外れ値に対して頑健なM推定量を用いる方が良いということは前述した。信頼区間あるいは検定の観点から \bar{X} とM推定量を比較してみよう。

影響関数 $\psi(x)$ が非減少関数、左連続、 x の大きな正 (あるいは負) の値に対して $\psi(x) > 0$ (あるいは < 0)、位置パラメータ θ に関して対称そして $\psi(x) = -\psi(-x)$ であるならば (ほとんどの $\psi(x)$ はこれらの条件を満たしている)、

$$T = -\frac{\sum_{i=1}^n \psi(X_i - \theta)}{n} / (s / \sqrt{n}) \quad (22)$$

は $df = n-1$ の t 分布で近似できる (Boos[9])。

ここで

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n \psi^2(X_i - \hat{\theta}) \quad (23)$$

は $E\psi^2(X-\theta)$ の推定量であり、 θ のM推定量 $\hat{\theta}$ は

$$\frac{1}{n} \sum_{i=1}^n \psi(X_i - \theta) = 0 \quad (24)$$

の解である。

$$\lambda(c) = -\frac{1}{n} \sum_{i=1}^n \psi(X_i - c) \quad (25)$$

とおけば、(22)式は

$$T = \lambda(\theta) / (s / \sqrt{n}) \quad (26)$$

と表すことができるから、 θ の近似的な $(1-\alpha) \times 100\%$ 信頼区間は

$$\left[\lambda^{-1}\left(-t_{\alpha/2} \frac{s}{\sqrt{n}}\right), \lambda^{-1}\left(t_{\alpha/2} \frac{s}{\sqrt{n}}\right) \right] \quad (27)$$

で与えられる。

たとえば最小 2 乗推定量を考えよう。最小 2 乗法の ϕ 関数は

$$\phi(X_i - c) = X_i - c$$

で与えられるから

$$\lambda(c) = c - \bar{X} = z$$

ゆえに

$$c = \bar{X} + z = \lambda^{-1}(z)$$

すなわち

$$\lambda^{-1}(z) = \bar{X} + z$$

したがって

$$\lambda^{-1}(0) = \bar{X}$$

$$\lambda^{-1}\left(-t_{\alpha/2} \frac{s}{\sqrt{n}}\right) = \bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$\lambda^{-1}\left(t_{\alpha/2} \frac{s}{\sqrt{n}}\right) = \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$

この結果はいうまでもなく (21) 式を与える。

Huber の ϕ 関数

$$\phi(X_i - c) = \begin{cases} X_i - c & |X_i - c| \leq r \\ r & X_i - c > r \\ -r & X_i - c < -r \end{cases}$$

の場合には

$$\lambda^{-1}(0) = \bar{X}_1 + \frac{n_2}{n_1} r - \frac{n_3}{n_1} r = \hat{\mu}_M$$

$$\lambda^{-1}(z) = \hat{\mu}_M + \frac{n}{n_1} z$$

となる。ここで

$n_1 = |X_i - c| \leq r$ を満たす X_i の個数

$n_2 = X_i - c > r$ "

$n_3 = X_i - c < -r$ "

$$\bar{X}_1 = \frac{1}{n_1} \sum_{|X_i - c| \leq r} X_i$$

したがって次式が得られる。

$$\lambda^{-1}\left(-t_{\alpha/2}\frac{S}{\sqrt{n}}\right)=\hat{\mu}_M-\frac{n}{n_1}t_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$\lambda^{-1}\left(t_{\alpha/2}\frac{S}{\sqrt{n}}\right)=\hat{\mu}_M+\frac{n}{n_1}t_{\alpha/2}\frac{S}{\sqrt{n}}$$

さて(7)式を用いれば、信頼区間幅

$$CI=\lambda^{-1}\left(t_{\alpha/2}\frac{S}{\sqrt{n}}\right)-\lambda^{-1}\left(-t_{\alpha/2}\frac{S}{\sqrt{n}}\right) \quad (8)$$

の期待値は μ の期待 $(1-\alpha)\times 100\%$ 信頼区間幅を与える。表 7 は $\sqrt{n}CI$ の期待値 (期待95%信頼区間幅) が、 μ の推定量として \bar{X} 、Huber の ϕ による M 推定量 (調整定数 $r=1.5$ の $Hr=1.5$ と $r=1.0$ の $Hr=1.0$) を用いるとき母集団分布 (正規、三角、ロジスティック、二重指数、 $t(3)$ 、スラッシュ) によってどのように異なるかを示している。Boos [9] の行ったモンテカルロ実験の結果である。

表 7 μ の期待95%信頼区間幅 ($n=20$) ($\sqrt{n}CI$)

母集団分布 推定量	正 規	三 角	ロジス ティック	二重指数	$t(3)$	スラッシュ
\bar{X}	4.13	3.37	4.05	4.03	6.34	149.53
$Hr=1.5$	4.26	3.59	3.97	3.58	5.33	11.94
$Hr=1.0$	4.42	3.76	4.01	3.44	5.25	11.03

注：Boos[9] 表3より作成。

このモンテカルロ実験の結果は、母集団分布が正規分布より長い尾をもつロジスティック、二重指数、 $t(3)$ 、スラッシュ等々の分布の場合には、推定量 \bar{X} の信頼区間幅は、外れ値に頑健な Huber の M 推定量の信頼区間幅より広くなり、推定量 \bar{X} の信頼性が低くなること、 t 検定の検定力が小さくなるという前述したことを裏付けている。

位置パラメータの期待値 μ の最小 2 乗推定量 \bar{X} および線形回帰モデルの回帰係数 β の最小 2 乗推定量 $\hat{\beta}$ のすぐれた特性と t 検定の検定力は正規母集団のもとでのみ保証されるといっても過言ではない。しかし実際には誤差項の分布が正規分布よりもっと長い尾をもつ分布に従っている例が多いということがわかってきた。長い尾をもつ分布のもとでは「外れ値」が発生する可能性は高い。ところで計量経済学において「外れ値」とは一体何なのか。正規性の検定へ入る前に「外れ値」について考えてみよう。

4 外れ値について

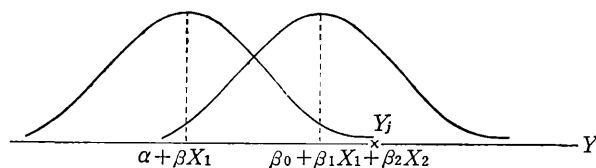
統計学では外れ値自体を研究することがある。入学試験においてきわめて高得点で外れ値とみなさざるを得ない受験生が 2, 3 人いたとしよう。この場合には入試問題が漏洩したのではないかと

(1)
 疑いがかけられ、これらの外れ値自体を綿密に調査しなければならない。しかし計量経済学においては外れ値自体に特別の関心があることはほとんどない。計量経済学において外れ値が関心あるのは、次の点からである。

①外れ値と考えられる大きな残差 e はモデルが不適切であることを示しているのではないか。この場合、不適切なモデルという場合次の2つを区別しなければならない。残差 e が大きいのは、モデルの定式化が間違っているからではなく、均一分散の仮定が成立しておらず、たとえば X の水準が大きくなるにしたがって誤差分散が大きくなるという不均一分散のためである。このケースと区別されるべきは、重要な系統的要因が欠落しているために大きな残差が現われているという場合である。この場合には大きな残差は複数個あり、モデルの説明力も低いであろう。

大きな、複数個の残差はモデル定式化の不適切さを示しているとすれば、頑健推定ではなく、モデルを変更してさらに系統的要因を考えるべきである。たとえば $E(Y) = \alpha + \beta X_1$ のとき、図12の Y_j をモデルは過小推定し、残差 $e_j = Y_j - \hat{Y}_j$ は正の大きい値をとり、 e_j は外れ値とみなされるであろう。しかしモデル $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ のもとで e_j は小さくなり、もはや外れ値ではない。

図12 不適切なモデルによって残差 e_j が外れ値になる場合



均一分散の仮定が妥当なとき、われわれはなぜ被説明変数の観測データ Y_j ではなくモデルに疑いを向けるのであろうか（データの入力ミス（転記ミス、パンチミス、キーボードからの入力ミス等々）は考えないことにしよう）。観測された経済データをもたらししたのは、経済主体の行動を貫いている何らかの合理性の結果である。さらに個人の行動結果ではなく、集計された人間行動の結果が観測データであり、そこには人間行動の合理性および慣性が働いており、意外な、全く説明しきれない異常な動きが経済変数に現れるとは考えられない。われわれは合理的人間行動を背後にもたない偶然変動、たとえばサイコロ投げで1が10回続けて現れるような偶然をみつめているのではない。観測データ Y_1, \dots, Y_n ではなく、残差 e_1, \dots, e_n をチェックするのはつねにモデルとの関連でのみ外れ値を問題にしているからである。

②大きな残差は、モデルで示されている Y と説明変数の関係では説明することができない突発的原因によるものかも知れない。この原因を明らかにすることができれば、この大きな残差はダミー変数で処理することができる。

注(1) その他、外れ値自体に特別の関心ある例として Barnett et al. [6] に示されている Hadlum 氏の離婚届け（軍務のため国を離れて349日目に夫人に赤ちゃんが生まれた。349日という懐胎期間は外れ値ではないか？）がある。

③外れ値とみなさざるを得ない大きな残差は、モデルの不適切さや突発的事態によるものではなく、正規分布より長い尾をもつ誤差項分布から発生したと考えられる場合がある。頑健推定を試みるべき状況はこの場合である。

結局、回帰モデルにおける大きな残差をすべてモデルが不適切であるからと言うこともできないし、モデルは適切であるが、誤差項が正規分布より長い尾をもつ分布に従っているからであると言うこともできない。外れ値はつねにモデルとの関連でのみ問題になるのであって、モデルを離れて絶対的な外れ値の水準があるわけではない。しかし計量経済学は③の観点から外れ値を解釈し、OLSに代って頑健推定を行うことを怠ってきたことは間違いない。③の場合を考慮に入れるためには、説明力も高く、誤差項は均一分数であり、突発的原因によるYと説明変数との間の基本的関係がくずれることもなかったと思われる関数に対して正規性の検定を行ない、とくに「外れ値」の現れやすい長い尾をもつ分布と判断されたときには頑健推定を行ない、OLSによる推定値をチェックすべきである。そこで次に正規性の検定について代表的な方法とその特徴、検定力について述べよう。

5 正規性の検定(1)——正規確率プロット

まず正規確率プロットあるいはランキット rankit プロットは正規性検定のきわめて有用な方法であることを示そう。

Y_1, Y_2, \dots, Y_n を平均 μ 、分散 σ^2 をもつ、独立な正規確率変数とすれば

$$Z_i = \frac{Y_i - \mu}{\sigma} \sim N(0, 1)$$

である。順序づけられた Y_i を

$$Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$$

とすれば

$$\frac{Y_{(1)} - \mu}{\sigma} \leq \frac{Y_{(2)} - \mu}{\sigma} \leq \dots \leq \frac{Y_{(n)} - \mu}{\sigma}$$

が成り立つから、 Z_i の順序

$$Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$$

$$\text{ここで } Z_{(i)} = \frac{Y_{(i)} - \mu}{\sigma}$$

は $Y_{(i)}$ の順序と同じである。

平均 0、分散 1 の標準正規分布から n 個の標本をくりかえし無作為抽出したときの i 番目の標本順序統計量の期待値を $\mu_{(i)}$ 、すなわち順序標準正規変数 $Z_{(i)}$ の期待値を $\mu_{(i)}$ とすれば

$$\mu_{(i)} = E[Z_{(i)}] = E\left[\frac{Y_{(i)} - \mu}{\sigma}\right]$$

であるから

$$E[Y_{(i)}] = \mu + \sigma \mu_{(i)} \quad (29)$$

である。

したがって観測値 Y_1, Y_2, \dots, Y_n が正規分布に従うならば, $Y_{(i)}$ の $\mu_{(i)}$ への回帰は切片 μ , 勾配 σ の直線とならなければならない。この $\mu_{(i)}$ は正規得点 normal score あるいはランキットとよばれ, 次のようにして求めることができる。

$Z_{(i)}$ の p. d. f. は

$$i \binom{n}{i} [\Phi(t)]^{i-1} [1-\Phi(t)]^{n-i} \phi(t) \quad (30)$$

で与えられるから (Dudewicz[15]p. 281), $Z_{(i)}$ の期待値 $\mu_{(i)}$ は次式となる。

$$\mu_{(i)} = i \binom{n}{i} \int_{-\infty}^{\infty} t \{\Phi(t)\}^{i-1} \{1-\Phi(t)\}^{n-i} \phi(t) dt \quad (31)$$

ここで

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

$$\Phi(t) = \int_{-\infty}^t \phi(z) dz$$

である。そしてこの $\mu_{(i)}$ を求めるために近似式

$$\mu_{(i)} \doteq \Phi^{-1}(a_i) \quad (32)$$

$$\text{ここで } a_i = \frac{i - \frac{3}{8}}{n + \frac{1}{4}}$$

を使うことができる (David [14] p. 78 Table 4.4)²⁾。私が採用した $\mu_{(i)}$ の求め方は, Harter [19] にもとづき

$$\mu_{(i)} = \Phi^{-1}(a_i)$$

$$a_i = \frac{i - \alpha}{n - 2\alpha + 1} \quad (33)$$

$$\alpha = 0.327511 + 0.058212 \log_{10} n$$

$$- 0.007909 (\log_{10} n)^2$$

である。ただし $n \geq 9$ であつ $i=1$ あるいは $i=n$ のときのみ $\mu_{(i)}$ を

$$\Phi^{-1} \left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}} \right) \quad (34)$$

によって求めている。⁷

$\mu_{(i)}$ の真の値と3つの近似式

注(2) a_i の選択は上記以外にも可能であり, *BMDP* は $a_i = (i-1/3)/(n+1/3)$ を用いている。Mage [25] は a_i の選択について議論している。 $\mu_{(i)}$ の厳密な求め方は Royston[29] をみよ。

表 8 標準正規母集団からの順序統計量の期待値と3つの近似式

n=5					n=20					n=30				
i	$\mu_{(i)}$	(1)	(2)	(3)	i	$\mu_{(i)}$	(1)	(2)	(3)	i	$\mu_{(i)}$	(1)	(2)	(3)
3	0.000	0.000	0.000	0.000	11	0.062	0.062	0.062	0.063	16	0.041	0.041	0.041	0.042
4	0.495	0.495	0.497	0.524	12	0.187	0.187	0.186	0.189	17	0.125	0.125	0.124	0.126
5	1.163	1.172	1.180	1.282	13	0.315	0.315	0.314	0.319	18	0.209	0.209	0.208	0.210
					14	0.448	0.448	0.447	0.454	19	0.294	0.294	0.294	0.297
					15	0.590	0.590	0.589	0.598	20	0.382	0.382	0.382	0.385
					16	0.745	0.745	0.744	0.755	21	0.473	0.473	0.472	0.477
					17	0.921	0.921	0.919	0.935	22	0.568	0.568	0.567	0.573
					18	1.131	1.131	1.128	1.150	23	0.669	0.669	0.668	0.674
					19	1.408	1.408	1.404	1.440	24	0.777	0.777	0.775	0.784
					20	1.867	1.879	1.869	1.960	25	0.894	0.894	0.893	0.903
										26	1.026	1.026	1.024	1.036
										27	1.179	1.179	1.176	1.192
										28	1.365	1.365	1.361	1.383
										29	1.616	1.616	1.610	1.645
										30	2.043	2.054	2.041	2.128

(注) $\mu_{(i)} = E(Z_{(i)})$ は CRC Handbook (1966) [11]

- (1) $\Phi^{-1}\left(\frac{i-\alpha}{n-2\alpha+1}\right)$, $\alpha = 0.327511 + 0.058212 \log_{10} n - 0.007909(\log_{10} n)^2$
- (2) $\Phi^{-1}\left(\frac{i-3/8}{n+1/4}\right)$
- (3) $\Phi^{-1}\left(\frac{i-1/2}{n}\right)$

(1) $\Phi^{-1}\left(\frac{i-\alpha}{n-2\alpha+1}\right)$, α の値は前述の通り

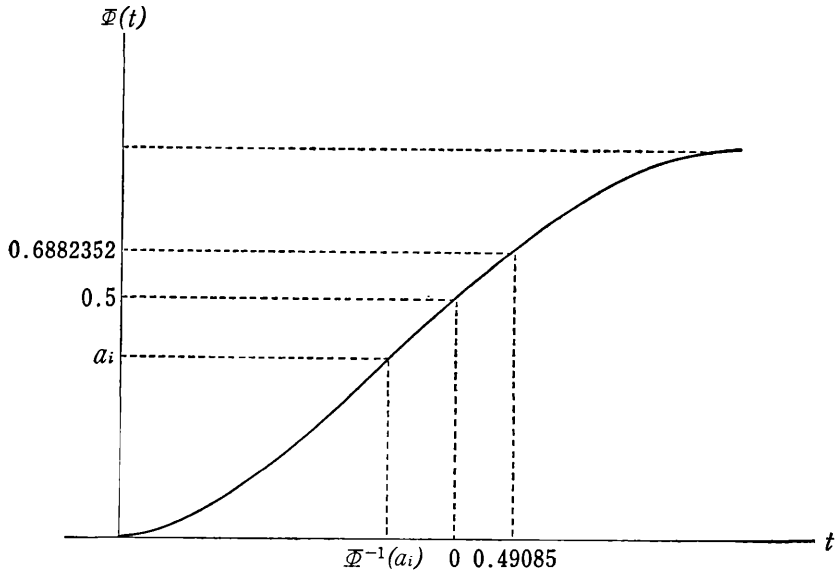
(2) $\Phi^{-1}\left(\frac{i-\frac{3}{8}}{n+\frac{1}{4}}\right)$

(3) $\Phi^{-1}\left(\frac{i-\frac{1}{2}}{n}\right)$

によるランキッツの値は表8をみよ。 $n \leq 8$ のとき(1), $n \geq 9$ で, $i=1$ あるいは $i=n$ のときには(2), $i=2, \dots, n-1$ のときには(1)の近似式を用いるというのが私が採用したランキッツ計算方法である。理由は表8から明らかであろう。

たとえば $n=21, i=11$ のとき

図13 標準正規変数の変位値



$$a_{11} = \frac{11 - \frac{3}{8}}{21 + \frac{1}{4}} = 0.5, \quad \Phi^{-1}(0.5) = 0$$

$n=21, i=15$ のとき

$$a_{15} = \frac{15 - \frac{3}{8}}{21 + \frac{1}{4}} = 0.6882352$$

$$\Phi^{-1}(0.6882352) = 0.49085$$

となる (図13参照)。

線形回帰モデルの誤差項 u_i が平均 0, 分散 σ^2 の正規分布に従うならば

$$v_i = \frac{u_i}{\sigma} \sim N(0, 1)$$

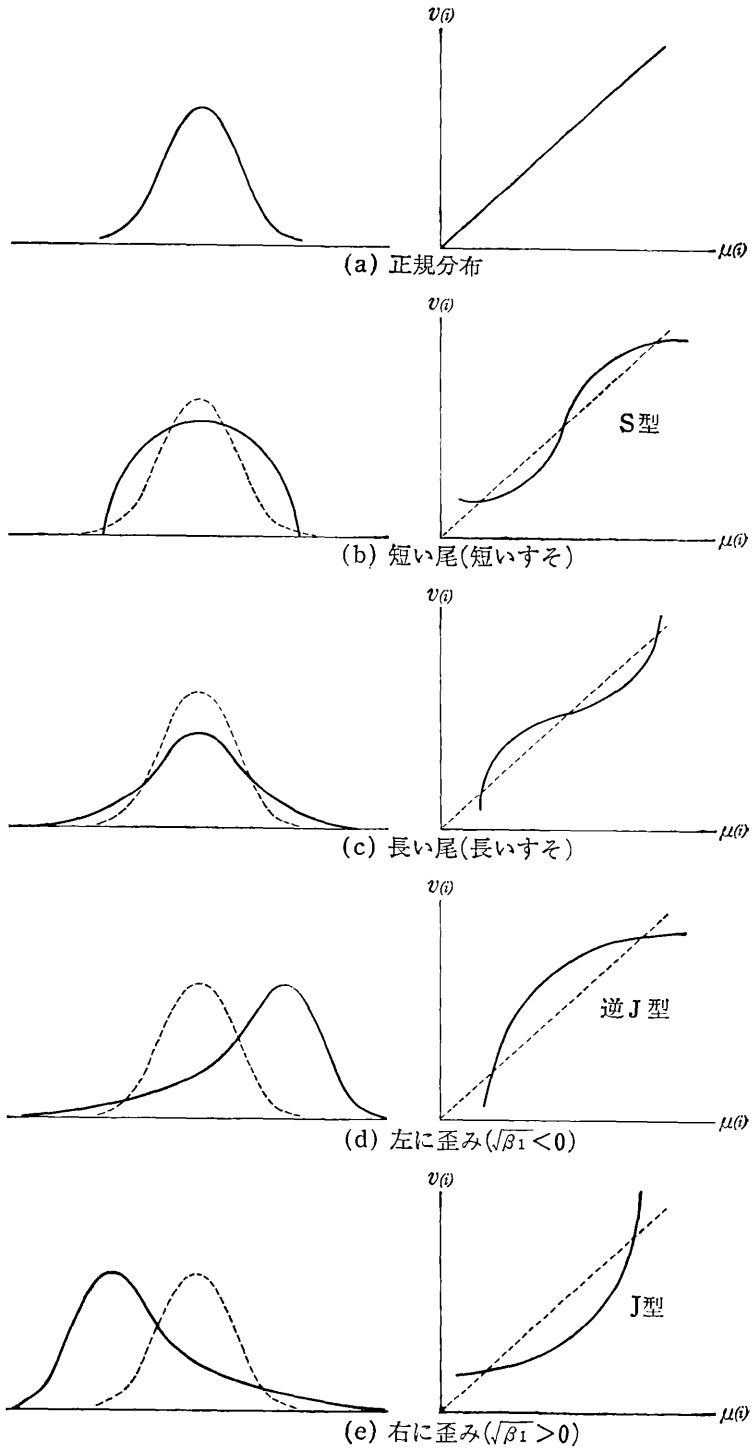
であるから, 順序化された標準正規変数 $v_{(i)}$ とランキット $\mu_{(i)}$ とのプロット $(v_{(i)}, \mu_{(i)})$ は平均 0, 勾配 1 の直線になる。したがって $(v_{(i)}, \mu_{(i)})$ をプロットして, このプロットが直線かどうか, あるいはどのようなパターンを示しているかによって u_i の正規性を検定する方法が正規確率プロットあるいはランキットプロットである。 $(v_{(i)}, \mu_{(i)})$ の典型的なパターンは図 14 に示した。図(a)は正規分布, (b)~(e)はいずれも非正規分布である。(b)~(e)の点線は正規分布を表している。

線形回帰モデルにおける v_i の標本対応はスチューデント化残差

$$r_i = \frac{e_i}{s(1-h_{ii})^{1/2}} \quad (35)$$

あるいは

図14 種々の分布とランキットプロット



$$t_i = \frac{e_i}{s_{(i)}(1-h_{ii})^{1/2}} \quad (36)$$

である。ここで

e_i = 最小 2 乗残差

h_{ii} = ハット行列の (i, i) 要素

s^2 = 誤差分散の不偏推定量

$s^2_{(i)}$ = i 番目の観測値を除いたときの誤差分散の不偏推定量

したがって r_i の順序化された値 $r_{(i)}$ あるいは t_i の順序化された値 $t_{(i)}$ とランキット $\mu_{(i)}$ をプロットすることによって、線形回帰モデルの誤差項の正規性を検定することができる。

ただし正規確率プロットを正規性の検定に用いるとき次の点に注意すべきである。

(1) 正規分布を示す直線からの乖離は、誤差項が非正規である場合に生ずることはいうまでもないが、モデルの特定化の誤りからも生ずる。正しいモデルからみれば不必要な説明変数が含まれていたり、必要な変数が除外されているとき、正規性の検定を残差にもとづいて行ない、正規確率プロットで直線からの乖離が生じたとしても、それはモデルの特定化の誤りにもとづいているのかも知れない。したがってパラメータ推定値の理論との整合性（符号条件、大きさ）および統計的有意性、説明力、誤差項の均一分散、自己相関なしの仮定、これらを吟味した上で「正しい」と思われるモデルの残差についてのみ正規性検定の意義がある。

(2) Daniel and Wood [12] の Appendix 3 A に示されているプロットによれば標本の大きさが小さい ($n \leq 16$) とき、正規確率プロットはしばしば線形性から著しく乖離することがある。大標本 ($n \geq 32$) においてはプロットの検出力はよい。正規確率プロットによって正規性を検定する場合、通常約 20 ぐらいの n が必要である。

(3) 線形回帰モデルの誤差項が正規分布をしていなくても、この非正規性が残差を用いる正規確率プロットには現れないことがある。残差 e_i は単純な確率変数ではなく、パラメータ推定を行なった後の残余であり

$$e_i = u_i - \sum_{j=1}^n h_{ij} u_j \quad (37)$$

であるから、誤差項 u_i の非正規性がそのまま e_i に反映されない。

(4) 正規確率プロットは外れ値の検出に用いることもできる。1 個あるいは 2 個の外れ値があれば、その外れ値は直線から著しく離れたところに位置している。

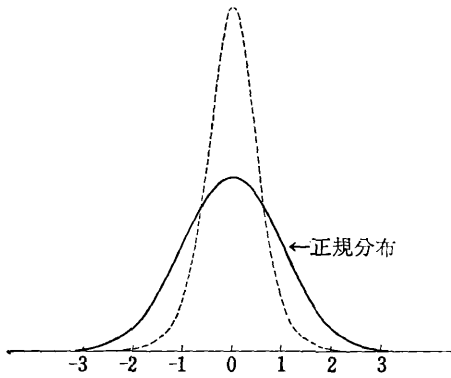
6 正規確率プロットの実例

例 1. 正規母集団あるいは非正規母集団から得られた標本値 X_i を

$$Z_i = \frac{X_i - \bar{X}}{s}$$

図15 ジョンソン無限分布 SU(0,2)

(a) 確率分布



(b) 標本正規確率プロット

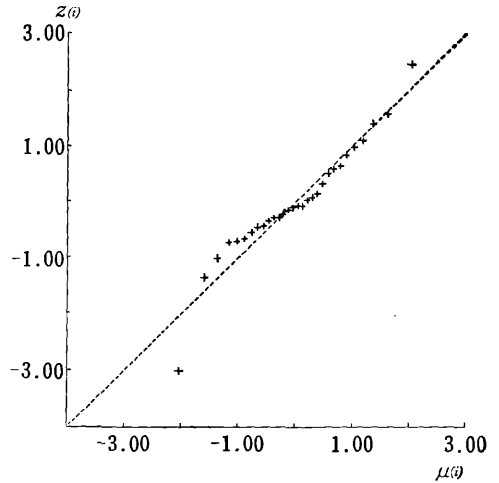
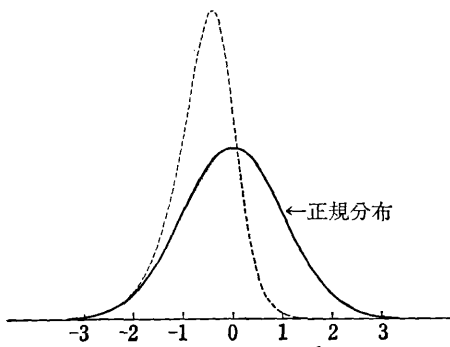
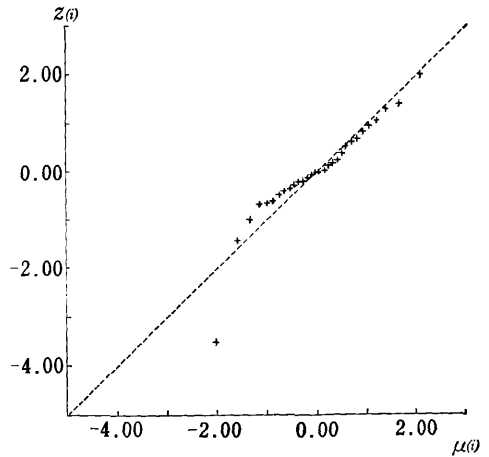


図16 ジョンソン無限分布 SU(1,2)

(a) 確率分布



(b) 標本正規確率プロット



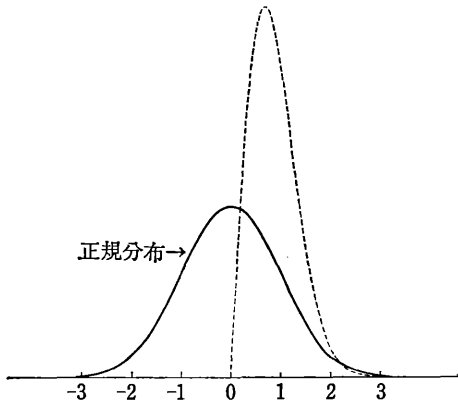
$$\text{ここで } s = \left[\frac{\sum (X_i - \bar{X})^2}{(n-1)} \right]^{\frac{1}{2}}$$

と標準化し、 Z_i を順序化した $Z_{(i)}$ とランキット $\mu_{(i)}$ をプロットして正規確率プロットの非正規性検出力を調べてみよう。何百組もの標本にもとづくモンテカルロ実験ではなく、1標本にもとづく実験にすぎないから、この実験結果にもとづいて正規確率プロットの検出力について結論を下すことは早計であるが、分布の型と正規確率プロットの型の実例としての意義はある。

考察した非正規分布は表9に示されている6種類である。一様分布、指数分布は周知の分布であ

図17 ワイブル分布 ($\beta=1, \gamma=2$)

(a) 確率分布



(b) 標本正規確率プロット

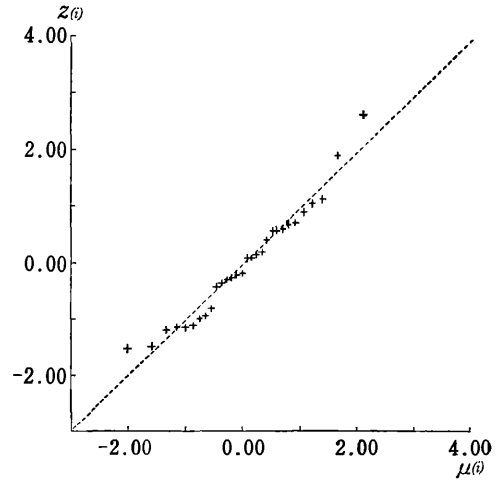


図18 一様分布からの標本正規確率プロット

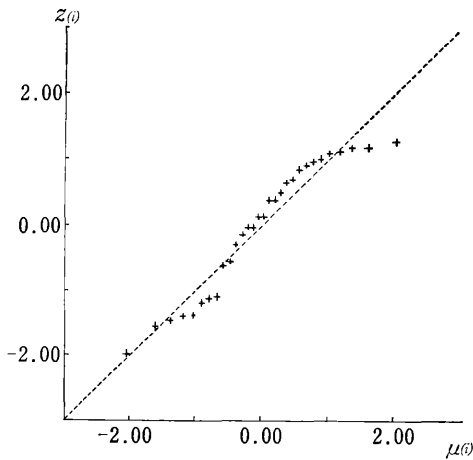
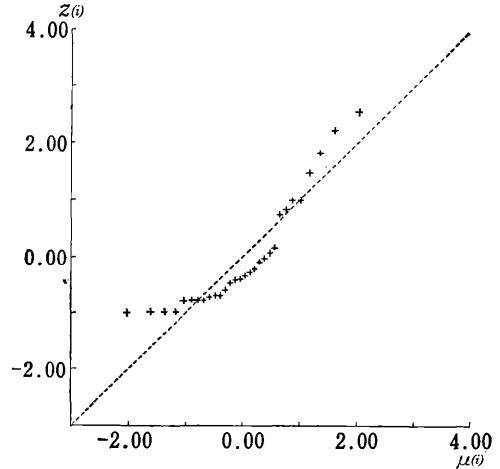


図19 指数分布からの標本正規確率プロット



り、ロジスティック分布は図2に示されているので、ジョンソン無限分布とワイブル分布を図15(a), 図16(a), 図17(a)に示した。それぞれの母集団からの標本データは D'Agostino and Stephens [13] の Appendix 3 のデータ30個を使用した。これらの標本データから得られた $Z_{(i)}$ と $\mu_{(i)}$ のプロットは図18から図20および図15(b), 図16(b), 図17(b)に示されている。

この標本データから得られた正規確率プロットを図14のパターンと照し合わせてみると、図18は(b), 図19は(e), 図15(b), 16(b), 20は(c)と識別できる。図17(b)は直線からの乖離はみられるがパターンを特定化することはできない。図18や図19の正規確率プロットをみて、母集団分布がそれぞれ一様分布、指数分布と識別することは全く不可能であるが、非正規性を検出することはできる。図20, 図15(b), 図16(b)からそれぞれロジスティック, $SU(0, 2)$, $SU(1, 2)$ と母集団分布を特定化するこ

図20 ロジスティック分布からの標本正規確率プロット

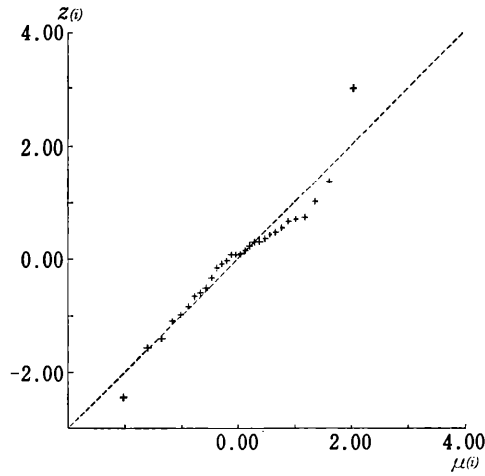


表 9 6種類の非正規分布

母集団分布	確率密度関数	$\sqrt{\beta_1}$	β_2	標本データ $n=30$				
				データ番号	\bar{X}	s	$\sqrt{b_1}$	b_2
(1) 一様	$\frac{x}{10}, 0 < x < 10$	0	1.80	1-30	6.134	2.941	-0.413	1.843
(2) 指数	$\frac{1}{5}e^{-x/5}, x > 0$	2	9	31-60	5.257	5.215	1.137	3.234
(3) ロジスティック	$\frac{\exp[(x-100)/10]}{10\{1+\exp[(x-100)/10]\}^2}$ $-\infty < x < \infty$	0	4.2	31-60	101.78	14.48	0.276	4.867
(4) ジョンソン無限 $\gamma=0, \delta=2$ SU(0,2)	$\frac{\delta}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)(1+z^2)^{-1/2}$ $z = \gamma + \delta \sin h^{-1}x$ $-\infty < x < \infty$	0	4.51	31-60	-0.062	0.451	-0.248	4.91
(5) ジョンソン無限 $\gamma=1, \delta=2$ SU(1,2)		-0.87	5.59	31-60	-0.628	0.535	-1.16	6.57
(6) ワイブル $\beta=1, \gamma=2$	$(\gamma x^{\gamma-1}/\beta^\gamma) \exp[-(x/\beta)^\gamma]$	0.63	3.25	1-30	0.868	0.474	0.509	3.03

とはできないが、これらの正規確率プロットはいずれも母集団分布は $\beta_2 > 3$ である長い尾をもつ分布であることを正しく示している。図17(b)のように非正規性を明確に検出できない場合ももちろんあるが、後述する検定統計量のように1つの数値に標本データの情報を縮約してしまわない“目による検定”，である正規確率プロットは正規性検定の有力な方法であり、計量経済学の分野においても積極的に導入すべき方法である。

例2. 次に線形回帰モデルの誤差項の正規性を正規確率プロットを用いて検定してみよう。

例としてとりあげたのは次の3つの関数である。

貨幣賃金率関数

$$WDOT2 = -4.6369 + 16.375 \left(\frac{1}{RU2} \right) + 0.88961 CPIDOT2 \quad (38)$$

(-4.186) (8.208) (10.864)

$$\bar{R}^2 = 0.933, \quad s = 1.679, \quad DW = 1.312$$

推定期間 1965年度～1988年度

ここで

$WDOT2 = w$ の対前年度変化率 (%)

$$w = \frac{\text{雇用者所得}}{\text{雇用者数}}$$

$RU2 =$ 完全失業率 (%)

$CPIDOT2 =$ 消費者物価対前年度変化率 (%)

アメリカから EC 向け輸出関数

$$\log QXUEC\$ = 1.3397 + 1.3721 \log GDPEC\$88 - 0.68646 \log \left(\frac{PXUSA88}{WPIEC\$88} \right)_{-1} \quad (39)$$

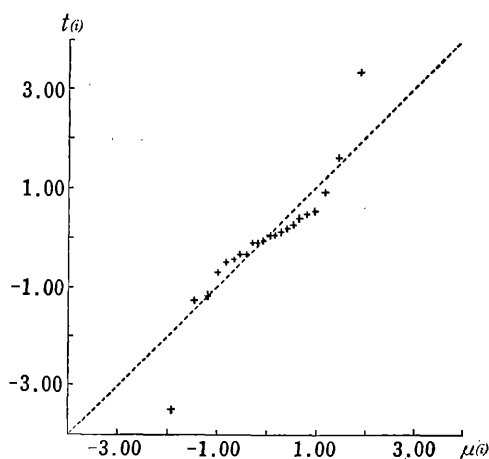
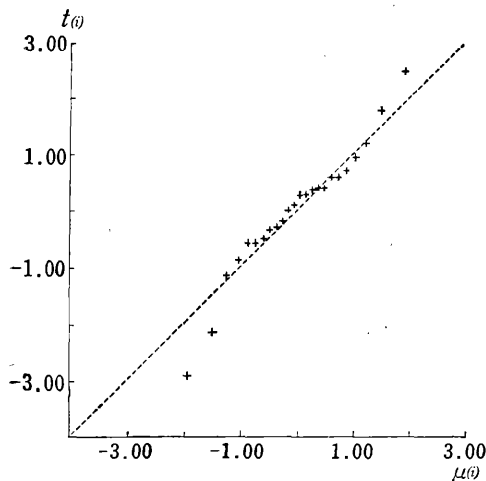
(1.869) (16.892) (-5.197)

$$\bar{R}^2 = 0.932, \quad s = 0.0702, \quad DW = 1.415$$

推定期間 1967年～1988年

図21 正規確率プロット ((38)式)

図22 正規確率プロット ((39)式)



ここで

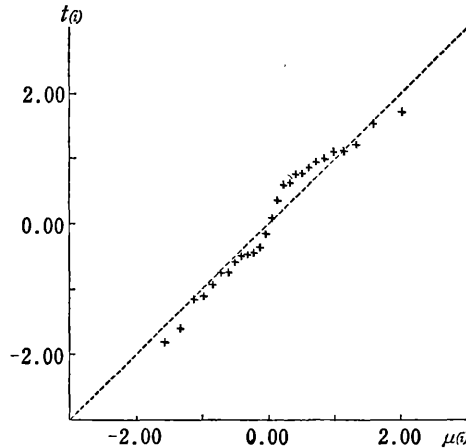
$QXUEC\$ =$ アメリカから EC 向け輸出数量 (1988年基準, 億ドル)

$GDPEC\$88 =$ EC の GDP (1988年基準, 億ドル)

$PXUSA88 =$ アメリカの輸出価格指数 (1988年=100)

WPIEC\$88=EC の WPI (GDP による加重平均, ドル建て, 1988年=100)

図23 正規確率プロット ((40) 式)



アメリカの日本からの輸入関数

$$\log MUJR2 = -6.9062 + 3.3451 \log \sum_{i=0}^3 \left(\frac{4-i}{10}\right) YRU_{-i} - 1.1982 \log \sum_{i=1}^8 w_i RPUJ4_{-i} \quad (40)$$

(-6.311) (23.140)
(-9.246)

$$w_1 = w_8 = \frac{1}{20}, \quad w_2 = w_7 = \frac{2}{20}, \quad w_3 = w_6 = \frac{3}{20}, \quad w_4 = w_5 = \frac{4}{20} \quad (\text{逆V分布})$$

$$\bar{R}^2 = 0.975, \quad s = 0.05249, \quad DW = 1.724$$

推定期間 1982/I ~ 1988/IV

ここで

$MUJR2$ = アメリカの日本からの輸入 (1985年価格, 10億ドル)

YRU = アメリカの実質 GNP (1985年価格, 100億ドル)

$RPUJ4$ = 日本の輸出価格 / アメリカの消費者物価 (1985年=100)

説明変数, 被説明変数とも季節変動調整済み四半期データである。

推定に用いたデータは表10, 表11, 表12に示した。(38), (39), (40)式の残差, スチューデント化残差, ランキッツは表13, 表14, 表15に, 順序化されたスチューデント化残差 $t(i)$ による正規確率プロットは図21, 図22, 図23に示されている。

図21, 図22は図14(c)のパターンを示しており, 誤差項は正規分布より長い尾をもつ分布に従っていることを示唆している。図23は図14(b)のパターンに近く, 誤差項の分布は正規分布より短い尾をもっているのではないかとと思われる。

実際, 誤差項が正規分布より長い尾をもつ分布に従っているとと思われる(38)式, (39)式の残差の箱型

表10 賃金率関数のデータ

	<i>RU2</i> (%)	<i>CPIDOT2</i> (%)	<i>WDOT2</i> (%)
1965	+1.290000D+00	+6.470000D+00	+1.0620669D+01
1966	+1.300000D+00	+4.560000D+00	+1.1061940D+01
1967	+1.220000D+00	+4.360000D+00	+1.3085118D+01
1968	+1.120000D+00	+4.740000D+00	+1.3318160D+01
1969	+1.110000D+00	+6.380000D+00	+1.6384806D+01
1970	+1.180000D+00	+7.500000D+00	+1.7020774D+01
1971	+1.310000D+00	+5.810000D+00	+1.3961900D+01
1972	+1.340000D+00	+5.710000D+00	+1.5373316D+01
1973	+1.270000D+00	+1.559000D+01	+2.1626076D+01
1974	+1.510000D+00	+2.068000D+01	+2.7175516D+01
1975	+1.950000D+00	+1.040000D+01	+1.2733039D+01
1976	+1.970000D+00	+9.580000D+00	+1.0773495D+01
1977	+2.070000D+00	+6.770000D+00	+9.9240842D+00
1978	+2.200000D+00	+3.810000D+00	+6.3601807D+00
1979	+2.030000D+00	+4.890000D+00	+5.9052366D+00
1980	+2.080000D+00	+7.630000D+00	+5.9696431D+00
1981	+2.220000D+00	+3.950000D+00	+6.2449259D+00
1982	+2.460000D+00	+2.600000D+00	+3.8389011D+00
1983	+2.660000D+00	+1.900000D+00	+2.3958818D+00
1984	+2.670000D+00	+2.180000D+00	+3.8871677D+00
1985	+2.640000D+00	+1.930000D+00	+3.9355676D+00
1986	+2.840000D+00	+0.000000D+00	+2.9671797D+00
1987	+2.780000D+00	+4.980000D-01	+2.6342295D+00
1988	+2.420000D+00	+7.930000D-01	+4.3272539D+00

表11 アメリカから EC 向け輸出関数のデータ

	$\ln(GDPEC \$ 88)$ (100億ドル)	$\ln(PXUWPIEC) (-1)$ (*88CY=100)	$\ln(QXUEC \$)$ (億ドル(*88))
1967	+5.4685706D+00	+4.7026509D+00	+5.4433610D+00
1968	+5.5191462D+00	+4.7290636D+00	+5.6923473D+00
1969	+5.5777089D+00	+4.7695137D+00	+5.7783796D+00
1970	+5.6233350D+00	+4.7651082D+00	+5.8863373D+00
1971	+5.6540984D+00	+4.7522450D+00	+5.8400310D+00
1972	+5.6900756D+00	+4.7075048D+00	+5.8690709D+00
1973	+5.7480609D+00	+4.6265816D+00	+6.0556757D+00
1974	+5.7717775D+00	+4.5784714D+00	+6.0870701D+00
1975	+5.7542950D+00	+4.6470997D+00	+6.0116176D+00
1976	+5.8038299D+00	+4.6778180D+00	+6.0862655D+00
1977	+5.8275329D+00	+4.7271722D+00	+6.0877243D+00
1978	+5.8618056D+00	+4.6690473D+00	+6.1873808D+00
1979	+5.8977251D+00	+4.5923467D+00	+6.3123916D+00
1980	+5.9115877D+00	+4.5603523D+00	+6.4942998D+00
1981	+5.9257840D+00	+4.5739830D+00	+6.3087212D+00
1982	+5.9317968D+00	+4.7794765D+00	+6.2357393D+00
1983	+5.9443231D+00	+4.8405301D+00	+6.1506241D+00
1984	+5.9732624D+00	+4.8984154D+00	+6.1911368D+00
1985	+5.9979559D+00	+4.9671077D+00	+6.1718133D+00
1986	+6.1029947D+00	+4.9587955D+00	+6.3042459D+00
1987	+6.1293116D+00	+4.7367211D+00	+6.4182426D+00
1988	+6.1621857D+00	+4.5983183D+00	+6.5700843D+00

注：データはすべて日本貿易振興会（JETRO）計量分析チームの好意による。

表12 アメリカの日本からの輸入関数データ

	$Ln(SW(YRU)(0,3))$ (100億ドル)	$Ln(IV(RPUJ4)(1,8))$ (%)	$Ln(MUJR2)$ (10億ドル)
1982.01	+4.4895988D+00	+4.9174522D+00	+2.2975826D+00
1982.02	+4.4838316D+00	+4.9028608D+00	+2.2550846D+00
1982.03	+4.4770425D+00	+4.8809214D+00	+2.2674138D+00
1982.04	+4.4743129D+00	+4.8545622D+00	+2.1901216D+00
1983.01	+4.4769049D+00	+4.8251086D+00	+2.2658796D+00
1983.02	+4.4789732D+00	+4.7944467D+00	+2.3241509D+00
1983.03	+4.5024340D+00	+4.7654251D+00	+2.3484185D+00
1983.04	+4.5191906D+00	+4.7417881D+00	+2.5599368D+00
1984.01	+4.5395708D+00	+4.7240992D+00	+2.5942841D+00
1984.02	+4.5569217D+00	+4.7117395D+00	+2.6533828D+00
1984.03	+4.5693057D+00	+4.7048300D+00	+2.7974642D+00
1984.04	+4.5785592D+00	+4.7010893D+00	+2.6873031D+00
1985.01	+4.5876974D+00	+4.6959245D+00	+2.8457230D+00
1985.02	+4.5956026D+00	+4.6889781D+00	+2.9038911D+00
1985.03	+4.6045290D+00	+4.6806292D+00	+2.8924806D+00
1985.04	+4.6128675D+00	+4.6661721D+00	+2.9249342D+00
1986.01	+4.6238840D+00	+4.6485874D+00	+3.0518763D+00
1986.02	+4.6301359D+00	+4.6299511D+00	+3.0730636D+00
1986.03	+4.6343697D+00	+4.6160295D+00	+3.1146258D+00
1986.04	+4.6377148D+00	+4.6140650D+00	+3.1213512D+00
1987.01	+4.6435829D+00	+4.6281902D+00	+3.0236878D+00
1987.02	+4.6528264D+00	+4.6571785D+00	+3.0589418D+00
1987.03	+4.6634863D+00	+4.6927162D+00	+3.0455695D+00
1987.04	+4.6762808D+00	+4.7289702D+00	+3.0878556D+00
1988.01	+4.6894102D+00	+4.7618786D+00	+3.0288285D+00
1988.02	+4.7009712D+00	+4.7896769D+00	+3.0053855D+00
1988.03	+4.7109256D+00	+4.8135330D+00	+3.0570625D+00
1988.04	+4.7190344D+00	+4.8335440D+00	+3.1539746D+00

注：データはすべて日本貿易振興会（JETRO）計量分析チームの好意による。

表13 賃金率関数(38)式の順序化残差, ランキッツ

	残 差	スチューデント 化残差 $r(i)$	スチューデント 化残差 $t(i)$	RANKITS	順序化 $r(i)$ の絶対値	順序化 $t(i)$ の絶対値	Half Normal Scores
1	-4.0537	-2.5088	-2.9257	-1.9473	0.0050	0.0049	0.0322
2	-3.1919	-1.9904	-2.1565	-1.5037	0.1012	0.0988	0.0831
3	-1.8745	-1.1460	-1.1551	-1.2392	0.1939	0.1894	0.1351
4	-1.4242	-0.8932	-0.8887	-1.0409	0.2708	0.2648	0.1874
5	-0.9538	-0.6004	-0.5910	-0.8768	0.2831	0.2768	0.2404
6	-0.8821	-0.5855	-0.5761	-0.7334	0.3068	0.3001	0.2940
7	-0.8134	-0.5102	-0.5010	-0.6038	0.3391	0.3318	0.3485
8	-0.4996	-0.3391	-0.3318	-0.4837	0.3862	0.3783	0.4041
9	-0.4936	-0.3068	-0.3001	-0.3702	0.3894	0.3814	0.4609
10	-0.3060	-0.1939	-0.1894	-0.2613	0.4092	0.4009	0.5193
11	-0.0082	-0.0050	-0.0049	-0.1556	0.5102	0.5010	0.5796
12	0.1646	0.1012	0.0988	-0.0516	0.5798	0.5704	0.6420
13	0.4213	0.2708	0.2648	0.0516	0.5855	0.5761	0.7071
14	0.4518	0.2831	0.2768	0.1556	0.5955	0.5861	0.7753
15	0.5939	0.3862	0.3783	0.2613	0.6004	0.5910	0.8473
16	0.6278	0.3894	0.3814	0.3702	0.7056	0.6969	0.9240
17	0.6529	0.4092	0.4009	0.4837	0.8932	0.8887	1.0065
18	0.9303	0.5798	0.5704	0.6038	0.9366	0.9337	1.0965
19	0.9379	0.5955	0.5861	0.7334	1.1460	1.1551	1.1965
20	1.1086	0.7056	0.6969	0.8768	1.1740	1.1853	1.3100
21	1.4922	0.9366	0.9337	1.0409	1.6825	1.7652	1.4435
22	1.8383	1.1740	1.1853	1.2392	1.9904	2.1565	1.6091
23	2.5710	1.6825	1.7652	1.5037	2.2105	2.4626	1.8363
24	2.7105	2.2105	2.4626	1.9473	2.5088	2.9257	2.2301

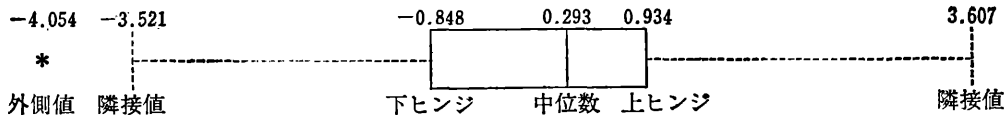
表14 アメリカから EC 向け輸出関数(39)式の順序化残差, ランキッツ

	残 差	スチューデント 化残差 $r(i)$	スチューデント 化残差 $t(i)$	RANKITS	順序化 $r(i)$ の絶対値	順序化 $t(i)$ の絶対値	Half Normal Scores
1	-0.1714	-2.7594	-3.4695	-1.9101	0.0390	0.0380	0.0351
2	-0.0797	-1.2398	-1.2587	-1.4584	0.0720	0.0700	0.0906
3	-0.0679	-1.1369	-1.1462	-1.1881	0.0782	0.0761	0.1474
4	-0.0462	-0.6833	-0.6734	-0.9845	0.0804	0.0782	0.2046
5	-0.0333	-0.4909	-0.4809	-0.8152	0.0868	0.0845	0.2626
6	-0.0289	-0.4379	-0.4284	-0.6665	0.1475	0.1437	0.3214
7	-0.0222	-0.3341	-0.3261	-0.5313	0.2045	0.1993	0.3814
8	-0.0216	-0.3333	-0.3253	-0.4053	0.2860	0.2790	0.4429
9	-0.0055	-0.0868	-0.0845	-0.2855	0.3333	0.3253	0.5061
10	-0.0051	-0.0804	-0.0782	-0.1697	0.3341	0.3261	0.5713
11	-0.0027	-0.0390	-0.0380	-0.0563	0.4143	0.4051	0.6392
12	0.0048	0.0720	0.0700	0.0563	0.4379	0.4284	0.7101
13	0.0053	0.0782	0.0761	0.1697	0.4909	0.4809	0.7847
14	0.0101	0.1475	0.1437	0.2855	0.5020	0.4919	0.8641
15	0.0123	0.2045	0.1993	0.4053	0.5632	0.5528	0.9492
16	0.0183	0.2860	0.2790	0.5313	0.6833	0.6734	1.0419
17	0.0264	0.4143	0.4051	0.6665	0.9236	0.9198	1.1445
18	0.0331	0.5020	0.4919	0.8152	1.1369	1.1462	1.2608
19	0.0382	0.5632	0.5528	0.9845	1.2398	1.2587	1.3972
20	0.0598	0.9236	0.9198	0.1881	1.5484	1.6122	1.5661
21	0.1022	1.5484	1.6122	1.4584	2.6972	3.3418	1.7970
22	0.1741	2.6972	3.3418	1.9101	2.7594	3.4695	2.1965

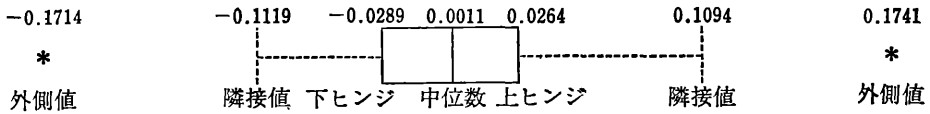
表15 アメリカの日本からの輸入関数(40)式の順序化残差, ランキッツ

	残 差	スチューデント 化残差 $r(i)$	スチューデント 化残差 $t(i)$	RANKITS	順序化 $r(i)$ の絶対値	順序化 $t(i)$ の絶対値	Half Normal Scores
1	-0.0963	-1.9174	-2.0341	-2.0122	0.0905	0.0887	0.0277
2	-0.0892	-1.7411	-1.8198	-1.5818	0.1600	0.1569	0.0712
3	-0.0744	-1.5656	-1.6152	-1.3267	0.3570	0.3506	0.1157
4	-0.0576	-1.1535	-1.1616	-1.1370	0.3675	0.3611	0.1605
5	-0.0537	-1.1010	-1.1059	-0.9812	0.4451	0.4379	0.2057
6	-0.0456	-0.9304	-0.9278	-0.8462	0.4758	0.4683	0.2512
7	-0.0379	-0.7513	-0.7445	-0.7250	0.4936	0.4860	0.2974
8	-0.0374	-0.7434	-0.7366	-0.6137	0.5989	0.5910	0.3441
9	-0.0275	-0.5989	-0.5910	-0.5096	0.6098	0.6020	0.3917
10	-0.0250	-0.4936	-0.4860	-0.4107	0.6343	0.6265	0.4401
11	-0.0242	-0.4758	-0.4683	-0.3158	0.7434	0.7366	0.4897
12	-0.0220	-0.4451	-0.4379	-0.2236	0.7513	0.7445	0.5404
13	-0.0186	-0.3675	-0.3611	-0.1334	0.7762	0.7698	0.5927
14	-0.0081	-0.1600	-0.1569	-0.0443	0.7816	0.7754	0.6465
15	0.0046	0.0905	0.0887	0.0443	0.8626	0.8580	0.7024
16	0.0179	0.3570	0.3506	0.1334	0.9304	0.9278	0.7605
17	0.0308	0.6098	0.6020	0.2236	0.9532	0.9514	0.8213
18	0.0325	0.6343	0.6265	0.3158	0.9990	0.9990	0.8853
19	0.0372	0.7762	0.7698	0.4107	1.0920	1.0964	0.9531
20	0.0388	0.7816	0.7754	0.5096	1.1010	1.1059	1.0257
21	0.0427	0.8626	0.8580	0.6137	1.1032	1.1082	1.1041
22	0.0460	0.9532	0.9514	0.7250	1.1535	1.1616	1.1899
23	0.0495	0.9990	0.9990	0.8462	1.2054	1.2170	1.2855
24	0.0559	1.0920	1.0964	0.9812	1.5030	1.5441	1.3945
25	0.0564	1.1032	1.1082	1.1370	1.5656	1.6152	1.5232
26	0.0609	1.2054	1.2170	1.3267	1.6539	1.7172	1.6836
27	0.0663	1.5030	1.5441	1.5818	1.7411	1.8198	1.9046
28	0.0779	1.6539	1.7172	2.0122	1.9174	2.0341	2.2887

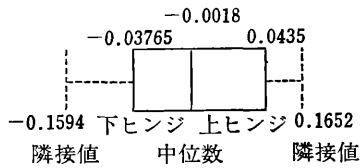
図24 残差の箱型図



(a) (38)式



(b) (39)式



(c) (40)式

(3)
 図 boxplot を描くと 図24(a), (b) のようになり, 「外れ値」(図の外側値) が 38 式で 1 個, 39 式で 2 個 発生していることがわかる。このような長い尾をもつ分布に対しては頑健推定が望ましい。(40) 式の 箱型図は外れ値の発生を示していない。

数学注

(1) x が l. f. d.

$$f(x) = \begin{cases} \frac{1-\varepsilon}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} & |x| \leq k \\ \frac{1-\varepsilon}{\sqrt{2\pi}} e^{\frac{k^2}{2} - k|x|} & |x| > k \end{cases}$$

にしたがうとき

$$\int_{-k}^k x^2 f(x) dx = (1-\varepsilon) [-2k\phi(k) + 1 - 2\Phi(-k)]$$

$$\int_{-\infty}^{-k} x^2 f(x) dx = \int_k^{\infty} x^2 f(x) dx$$

$$= \left(k + \frac{2}{k} + \frac{2}{k^3} \right) e^{-k^2}$$

であるから次式を得る。

(注3) 渡部他 [35] 参照。

$$\text{Var}(x) = (1-\varepsilon) \left\{ 1 - 2\Phi(-k) + 4\phi(k) \left(\frac{1}{k} + \frac{1}{k^3} \right) \right\}$$

ここで

$$\phi(k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{k^2}{2}}$$

$$\Phi(-k) = \int_{-\infty}^{-k} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

他方M推定量の漸近的分散は

$$\frac{E(\psi^2)}{[E(\psi')]^2}$$

で与えられるから、真の分布が l. f. d. のとき、Huber の ψ を用いるM推定量 T の分散を求めるために、 $E(\psi^2)$ と $E(\psi')$ を計算すると次式となる。

$$\begin{aligned} E(\psi^2) &= \int_{-\infty}^{-k} \frac{k^2(1-\varepsilon)}{\sqrt{2\pi}} e^{\frac{k^2}{2}+kx} dx + \int_{-k}^k \frac{x^2(1-\varepsilon)}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + \int_k^{\infty} \frac{k^2(1-\varepsilon)}{\sqrt{2\pi}} e^{\frac{k^2}{2}-kx} dx \\ &= (1-\varepsilon)k\phi(k) + (1-\varepsilon)[-2k\phi(k) + 1 - 2\Phi(-k)] + (1-\varepsilon)k\phi(k) \\ &= (1-\varepsilon)[1 - 2\Phi(-k)] \end{aligned}$$

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

$\text{Var}(T|F)$ は以上の結果を用いれば得られる。

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] Anscombe, F. J. : Rejection of outliers, *Technometrics*, 2, 1960, 123-147.
- [3] Atkinson, A. C. : Two graphical displays for outlying and influential observations in regression, *Biometrika*, 68, 1981, 13-20.
- [4] Atkinson, A. C. : Regression diagnostics, transformations and constructed variables, *Journal of the Royal Statistical Society, B*, 44, No. 1, 1982, 1-36.
- [5] Atkinson, A. C. : *Plots, Transformations, and Regression—An Introduction to Graphical Methods for Diagnostic Regression Analysis*, Oxford University Press, 1985.
- [6] Barnett, V. and Levis, T. : *Outliers in Statistical Data*, 2nd ed., John Wiley & Sons, 1984.
- [7] Beasley, J. D. and Springer, S. G. : Algorithm ASIII. The percentage points of the normal distribution, *Applied Statistics*, 26, 1977, 118-121.
- [8] Beckman, R. J. and Cook, R. D. : Outlier……s, *Technometrics*, 25, 1983, 119-149.
- [9] Boos, D. D. : A new method for constructing approximate confidence intervals from M-estimators, *Journal of the American Statistical Association*, 75, 1980, 142-145.
- [10] Cook, R. D. and Weisberg, S. : *Residuals and Inference in Regression*, Chapman and Hall, 1982.
- [11] *CRC Handbook of Tables for Probability and Statistics*, 2nd ed., The Chemical Ruber Co., 1966.

- [12] Daniel, C. and Wood, F. S.: *Fitting Equations to Data; Computer Analysis of Multifactor Data*, 2nd ed., John Wiley & Sons, 1980.
- [13] D'Agostino, R. B. and Stephens, M. A. ed.: *Goodness-of-Fit Techniques*, Dekker, 1986.
- [14] David, H. A.: *Order Statistics*, 2nd ed., John Wiley & Sons, 1981.
- [15] Dudewicz, E. J. and Mishra, S. Y.: *Modern Mathematical Statistics*, John Wiley & Sons, 1988.
- [16] Galton, F.: *Natural Inheritance*, Macmillan and Co., 1889.
- [17] Geary, R. C.: The ratio of the mean deviation to the standard deviation as a test of normality, *Biometrika*, 27, 1935, 310-332.
- [18] 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).
- [19] Harter, H. L.: Expected values of normal order statistics, *Biometrika*, 48, 1961, 151-165.
- [20] Hill, R. W. and Holland, P. W.: Two robust alternatives to least-squares regression, *Journal of the American Statistical Association*, 72, 1977, 828-833.
- [21] 廣松毅, 池田実, 藤原直哉, 若林芳雄: 計量経済分析再考——より信頼性の高いモデル作りのための推定手続き——, 「経済分析」, 第112号, 経済企画庁経済研究所, 1988.
- [22] Hoaglin, D. C.: Using quantiles to study shape, in Hoaglin, D. C., Mosteller, F. and Tukey, J. W., *Exploring Data Tables, Trends, and Shapes*, John Wiley & Sons, 1985.
- [23] Hocking, R. R.: Developments in linear regression methodology: 1959-1982, *Technometrics*, 25, 1983, 219-249.
- [24] 岩田暁一: 『先物とオプションの理論』, 東洋経済新報社, 1989.
- [25] Mage, D. T.: An objective graphical method for testing normal distributional assumptions using probability plots, *The American Statistician*, 36, 1982, 116-120.
- [26] 養谷千鳳彦: “計量経済学における頑健推定(1)”, 「三田学会雑誌」82巻3号, 1989.
- [27] Pearson, E. S. and Hartley, H. O.: *Biometrika Tables for Statisticians*, Vols. I and II, Cambridge University Press, 1972.
- [28] Rosenberger, G. L. and Gasko, M.: Comparing location estimators: trimmed means, medians, and trimean, in Hoaglin, D. C., Mosteller, F. and Tukey, J. W. ed., *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, 1983.
- [29] Royston, J. P.: Expected normal order statistics (exact and approximate), Algorithm AS 177, *Applied Statistics*, 31, 1982, 115-124.
- [30] Shapiro, S. S. Wilk, M. B. and Chen, H. J.: A comparative study of various tests for normality, *Journal of the American Statistical Association*, 1968, 1343-1372.
- [31] Sparks, D. N.: Half normal plotting, *Applied Statistics* 19, 1970, 192-196.
- [32] 柴田義貞: 『正規分布——特性と応用』, 東京大学出版会, 1981.
- [33] 高木信二: 『為替レート変動と国際通貨制度』, 東洋経済新報社, 1989.
- [34] 竹内啓: 『確率分布と統計解析』, 日本規格協会, 1975.
- [35] 渡部洋, 鈴木規夫, 山田文康, 大塚雄作: 『探索データ解析入門——データの構造を探る』, 朝倉書店, 1985.
- [36] Weisberg, S.: *Applied Linear Regression*, 2nd ed., John Wiley & Sons, 1985.
- [37] Wetherill, G. B.: *Regression Analysis with Applications*, Chapman and Hall, 1986.
- [38] Wilk, M. B. and Gnanadesikan, R.: Probability plotting methods for analysis of data, *Biometrika*, 55, 1968, 1-17. (経済学部教授)