研究ノート

回帰分析における分散不均一性に関する一考察：標準世帯の支出行動に関する場合

橋　本　紀　子

1. 問題意識

回帰モデルの分析の際，標準的な線形回帰モデルの仮定[1)]が満足されている場合にはその最小二乗推定量は最良不偏線形推定量であることが知られている。しかしながらこれらの仮定のいずれか一つでもが満たされていない時には，最小二乗推定量は効率の悪い，精度の低い推定量になっている可能性がある。

本稿では，これらの最小二乗法をモデルにあてはめて分析していく上で問題となる諸仮定のうち，分散不均一性の問題について分析を行っていく。経済データを取り扱っていく際には分散均一性の仮定が損なわれていることがしばしば観察され，とりわけ階層化されたデータが用いられるクロスセクションデータ分析では頻繁に問題が生じることが知られている。分散不均一性が生じると，最小二乗法による推定量$\hat{\beta}$は依然不偏ではあるものの有効性を失い，その分散の推定値がバイアスを持つため$\hat{\beta}$に関する仮説検定も信頼性を失ってしまう。このため経済データを実証していく上では，分散不均一性の問題が生じていないかを検討するとともに，問題が生じている場合には最小二乗法に代わる手法により推定作業を試みる必要がある。

本稿の構成は以下の通りである。第２節で分散不均一性についてそれがどのような状態か，そのもとでは最小二乗推定量にどのような不都合が起きるか整理する。また経済データを扱っていく上でどのような形の分散不均一性が想定されるか考察し，問題を解決する推定手法としてどのようなものが考えられるかのべる。あわせて，分散不均一性の有無に関する検定方法についても検討を行う。第３節では第２節で検討した理論的な手法を実際

1）各誤差項についてその平均が０であり分散は一定であること，異なる誤差項が互いに相関しないこと，説明変数が確率変数でなく誤差項と相関しないこと，といった条件を指す。

の経済データに当てはめて分析を行う。日本の標準世帯のクロスセクションデータを用いて消費関数の推定を行い，分散不均一性が観察されるかどうか，観察された場合どのような手段で問題が解決できるか，検討を行う。最後に第4節で本稿での分析結果をまとめ，残された課題についてのべる。

2.　分散不均一性の問題と検出

次のような重回帰モデルを考える。

$$\boldsymbol{y}=\boldsymbol{x\beta}+\boldsymbol{u} \qquad \cdots\cdots(1)$$

ここで　$\boldsymbol{y}=\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},\quad \boldsymbol{x}=\begin{bmatrix} 1 & X_{11} & \cdots & X_{k1} \\ 1 & X_{12} & \cdots & X_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{1n} & \cdots & X_{kn} \end{bmatrix},\quad \boldsymbol{\beta}=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix},\quad \boldsymbol{u}=\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$

通常，誤差項には次のような仮定がおかれている[2)]。

$$E(\boldsymbol{u})=\boldsymbol{O} \qquad \cdots\cdots(2)$$

$$E(\boldsymbol{uu}^T)=\sigma^2\boldsymbol{I} \qquad \cdots\cdots(3)$$

(3)式の仮定が成立していない場合として，誤差項が互いに相関を持つ場合（系列相関），誤差項の分散が等しくない場合（分散不均一性）の2ケースがある。以下，分散不均一性に問題を限定し考察を進めていく。

このような分散不均一性の存在する場合に得られる最小二乗推定量$\hat{\boldsymbol{\beta}}$は，次の2点から望ましい推定量と考えることができないことが知られている[3)]。①$\hat{\boldsymbol{\beta}}$は依然として不偏推定量であり一致性は保有しているが，有効性は失っており，また漸近的に効率的でもない。すなわち$\hat{\boldsymbol{\beta}}$は最良不偏線形推定量（BLUE）ではない。②$\hat{\boldsymbol{\beta}}$は不偏推定量ではあるものの，その分散の推定値がバイアスを持つため$\hat{\boldsymbol{\beta}}$にもとづく$\boldsymbol{\beta}$に関する仮説検定あるいは信頼区間は誤った推論を導く可能性がある[4)]。

分散不均一性のもとでは推定量の標本分散が最小とならないという意味から最小二乗推定量は望ましい推定量とは言えず，別により望ましい推定量を探す必要が生じる。

2）以下，行列またはベクトルの右肩の上付文字Tはその行列（ベクトル）が転置されたものであることを示している。

3）一般的な回帰モデルにおける最小二乗推定量の問題については[26]（p.270-8）参照のこと。

4）ホワイトは，たとえ分散不均一性の構造が未知であっても，その標準誤差の修正を行うことにより，分散不均一性に対して頑健な，$\hat{\boldsymbol{\beta}}$の分散共分散行列の推定量が得られることを示した[29,39]。

誤差値に関する仮定を緩め，誤差項の分散共分散行列を次式のように表現することができる。

$$E(\boldsymbol{uu}^T)=\sigma^2\boldsymbol{\Omega} \quad \cdots\cdots(4)$$

ここで，$\boldsymbol{\Omega}$ は既知の n 次正値定符号行列

任意の正値定符号行列はある非特異行列 $\boldsymbol{P}$ により $\boldsymbol{PP}^T$ と表すことができるので，今 $\boldsymbol{\Omega}=\boldsymbol{PP}^T$ と表すことにする。(1)式のモデルの両辺に左から $\boldsymbol{P}^{-1}$ をかけ整理すると，(5)式が導出される。

$$\boldsymbol{y}^*=\boldsymbol{X}^*\boldsymbol{\beta}+\boldsymbol{u}^* \quad \cdots\cdots(5)$$

ただし，$\boldsymbol{y}^*=\boldsymbol{P}^{-1}\boldsymbol{y}$, $\boldsymbol{X}^*=\boldsymbol{P}^{-1}\boldsymbol{X}$, $\boldsymbol{u}^*=\boldsymbol{P}^{-1}\boldsymbol{u}$

(5)式の 誤差項 の平均，分散共分散行列 はそれぞれ $E(\boldsymbol{u}^*)=\boldsymbol{O}$, $E(\boldsymbol{u}^*\boldsymbol{u}^{*T})=\sigma^2\boldsymbol{I}$となるので，(5)式については標準的な回帰モデルの仮定が満足されている。この結果(5)式における $\boldsymbol{\beta}$ の最小二乗推定量 $\boldsymbol{b}$ は $\boldsymbol{\beta}$ の最良不偏線形推定量となる[5)]。

$$\boldsymbol{b}=(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^*\boldsymbol{y}^*=(\boldsymbol{X}^T\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Omega}^{-1}\boldsymbol{y} \quad \cdots(6)$$

このような 一般化最小二乗推定量を求めるためには $\boldsymbol{\Omega}$ の構造 がわかっていなくてはならない。分散が既知の場合には $\boldsymbol{\Omega}$ ひいては $\boldsymbol{P}$ がわかっていることから一般化最小二乗法の手順に従い，次のようなウェイト付最小二乗法（WLS）を行うことにより $\boldsymbol{\beta}$ の最良不偏線形推定量を得ることができる。

今，$Var(u_i)=\sigma_i^2$ $(i=1,2,\cdots,n)$ であることがわかっているとする[6)]。この時モデルの各変数に ウェイトをつけ 次のように 変換する：$Y_i^*=Y_i/\sigma_i$, $X_{ji}^*=X_{ji}/\sigma_i$, $u_i^*=u_i/\sigma_i$ $(i=1,2,\cdots,n,\ j=0,1,\cdots,k,\ X_{0i}=1\ for\ all\ i)$。次にこの変換された変数に対して最小二乗法を行うならば，変換された誤差項の分散は一定となっているので，$\boldsymbol{\beta}$ の最良不偏線形推定量が得られる。

このようにデータを適当に変換する[7)]ことができるならば，変換された方程式における

5）ただし $\boldsymbol{x}^*$ 行列の第１列の各要素は１とはならず，(5)式は定数項を持たない。このため，決定係数は意味を持たないことと推定されたパラメータの解釈の際に注意が必要である。

6）この時 $\boldsymbol{\Omega}$ および $\boldsymbol{P}$ はそれぞれ

$$\boldsymbol{\Omega}=\begin{bmatrix}\sigma_1^2 & 0 & \cdots\cdots & 0\\ 0 & \sigma_2^2 & \cdots\cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots\cdots & \sigma_n^2\end{bmatrix},\quad \boldsymbol{P}=\begin{bmatrix}1/\sigma_1 & 0 & \cdots\cdots & 0\\ 0 & 1/\sigma_2 & \cdots\cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots\cdots & 1/\sigma_n\end{bmatrix}$$ となっている。

7）ウェイト付最小二乗法では，原データにおいて誤差の大きい信頼性を欠く観測値については相対的に小さいウェイトが課せられ，直線のあてはめに際して割り引いて考えられている。

誤差項の分散は均一となり，結果として最良不偏線形推定量が得られる。しかしながら，一般には，誤差分散の相対的な大きさに関する情報は知られていない。

分散不均一性は経済現象に関わるデータではしばしば観察され，とりわけクロスセクションデータを用いた場合に観察されることが多い。クロスセクションデータを分析する際には，経験的な事実を利用して誤差分散に関する情報を想定することが可能な場合がある。以下第3節での日本のデータへの応用分析を踏まえ，分散不均一性が誤差項に生じている可能性の高い例として，所得階層別に分類されたデータを用いて家計の消費行動を推定する場合を考えてみる。

家計調査年報における所得階層別データでは，各階層値としてその階層に属する標本世帯の平均値があげられている。このため，実際に分析に用いることのできるデータは個別の家計に対応しているのではなく集計されたレベルのデータである。たとえ個別の家計データを適用した場合には誤差項の分散は均一であったとしても，集計されたデータに関しては誤差分散は各階層に含まれる集計世帯数に応じて変化するため等しくないと思われる。多数の世帯から得られた集計値は少数の世帯による値より信頼性が高いと考えられるからである。たとえば各階層の誤差項の分散が集計世帯数の逆数に比列すると想定し，ウェイトとして$\sqrt{n_i}$（n_iは第i階層の標本数）を用いるならば，標本世帯数の大きい集計値に大きなウェイトがつけられその情報を重視することとなり，分散不均一性の問題を解決する事ができる可能性がある[2]。

一方，経済データにおいては，誤差項の分散は説明変数と何らかの関係を持っていることが多いと考えられる[8)]。

説明変数の増加とともに残差の標準偏差が増加する傾向が見られることがあるが，家計消費の分析においては，誤差分散は所得の増加につれて増大すると想定される。すなわち，低額所得者の支出はその大部分が生活必需品であるため大きな変動はむずかしく規則的なものになると考えられる一方，高額所得者については多様性に富んだ消費パターンが可能であるため，相対的にみて高額所得者の支出パターンは大きく変動し，低額所得者の支出行動は安定していると考えられる。このような支出パターンの違いのため，消費支出の動きを説明するモデルでは高額所得者の家計に対する誤差分散は低額所得家計に対する

8）その定式化は加算的なもの，乗法的なものあわせて様々な形が考えられる。いくつかの代表的なパターンを列挙すると，$\sigma_i^2=\sigma^2X_i$，$\sigma_i^2=\sigma^2X_i^2$，$\sigma_i^2=\sigma^2(\delta_1+\delta_2X_i)$，$\sigma_i^2=\sigma^2\exp(\delta_1+\delta_2X_i)$，$\sigma_i^2=\sigma^2X_i^{\delta_i}e^{\nu_i}$，$\sigma_i^2=\sigma^2[E(Y_i)]^2$ などがあげられる（ここで，δは未知パラメータ，ν_iはホワイトノイズの誤差項を表す）[16, 35]。

場合よりも大きいと想定される。この時，所得変数に関する値[9)]の逆数をウェイトに用いてデータの変換を行えば分散不均一性の問題に対処することができると考えられる。

以上，家計に関する所得階層別データを用いて分析を進めていく上でどのような分散不均一性の形式が考えられるかについて検討を行った。このように何らかの形で分散不均一性の形式が想定される場合には，それらの情報を用いてウェイト付最小二乗法を行うことにより分散不均一性の問題に対処することができる[10)]。

次に，誤差項の分散が均一であるかどうかを判断する基準について述べる。分散不均一性の有無を検討する際には，まず手始めに，回帰を行った際の残差の動きを観察することが有益である。残差自身の動きに加え，推定された残差（の二乗値）の変動がいずれかの説明変数の動きに起因するかどうかを調べたり，残差を被説明変数の理論値に対してプロットすることで，分散不均一性の存在を発見することが可能である[16, 17, 30]。

不均一分散を検定する統計量に関しては多くのものが提唱されている[11)]。それらは誤差分散の定式化に関して何らかの明示的な仮定をおくもの[12)]とおかないもの[13)]に大別され

9）プライス＝ハウタッカーは大量の家計データを世帯人員別，所得階層別に分割して分析した結果，消費支出額を説明するモデルにおいてその誤差分散は近似的に所得変数の二乗に比例することを見い出した[34]。

また，カトーナは，所得階層別の貯蓄関数に関して，同様に，その誤差分散がそれぞれの階層の平均所得の二乗に比例することを見い出している[24]。

10）一般には，分散不均一性が存在することはわかっていてもその形は不明である場合が多い。そのような場合直接誤差項の分散構造を推定することは不可能であるので，何らかの他の情報源を探す，適当な分散構造を想定し２段階の推定手法[5, 28, 37]を採用する，等の方法をとらなくてはならない。

11）これらの検定統計量についての包括的な議論については[6, 8, 11, 22, 23, 25]参照のこと。

12）たとえば，残差eの絶対値が説明変数X，または$\sqrt{X}$，$1/X$等の１次関数であると考え分散不均一性の有無を検討するグレイサー検定[９]，$\ln e^2$が$\ln X$の関数であると考えるパーク検定[33]などがあげられる。

13）多くの検定手法が提唱されているが，代表的なものとしてゴールドフェルト＝クォント（以下，ＧＱと略す）検定[12]，ブルーシュ＝ペーガン（以下，ＢＰと略す）検定[４]，ホワイト検定[39]がある。

ＧＱ検定は，まず観測値をある説明変数X_iの大きさの順に並べ，観測値をX_iの大きな値に対応するグループと小さな値に対応するグループとに分け，それぞれのグループについて回帰を行いＦ検定によりそれらの誤差分散の同一性を検定する方法であ

る[14)]。

本稿では後者の手法のうち，尤度比検定により分散不均一性の検出を行った[15)]。

今，観測値を2つのグループに分け，それぞれの標本数を T_1，T_2 とする。各グループ内の誤差分散の推定値を $\hat{\sigma}_i^2$ $(i=1,2)$，観測値全体の誤差分散の推定値を $\hat{\sigma}^2$と表す。この時 $\lambda=(\hat{\sigma}_1)^{T_1}(\hat{\sigma}_2)^{T_2}/(\hat{\sigma})^T$ と定義すると，(7)式で与えられる $-2\ln\lambda$は自由度1のカイ二乗分布に従う。

$$
\begin{aligned}
-2\ln\lambda &= T\cdot\log\hat{\sigma}^2-T_1\cdot\log\hat{\sigma}_1^2-T_2\cdot\log\hat{\sigma}_2^2\\
&= T\cdot\ln(SSR/(T-K-1))-T_1\cdot\ln(SSR_1/(T_1-K-1))\\
&\quad -T_2\cdot\ln(SSR_2/(T_2-K-1)) \qquad \cdots\cdots(7)
\end{aligned}
$$

る。この検定の際には観測値の全てを用いる必要はなく，中央のいくつかの観測値を除いて誤差分散を判別することにより検定力を高めることをゴールドフェルト＝クォントは勧めている。

ＢＰ検定は，ゴドフレー検定[10]と同じくラグランジュ乗数（ＬＭ）検定法であり，誤差項が正規分布に従うという仮定の下で導出されている。ＢＰ検定は漸近的な検定手法であり，コエンカーはＢＰ検定がとりわけ小標本においては正規性の仮定に強く依存していることを示した[27]。小標本におけるＬＭ検定統計量のバイアスの修正については[21]を参照のこと。

ホワイト検定は非常に広い範囲の分散不均一性に対応することができる手法であるが，そのために有限標本においてはあまり検定力が強くないことが知られている。ＢＰ検定とホワイト検定の関連性については[38]参照のこと。

なお，これらの手法を含む代表的な分散不均一性に関する検定統計量のふるまいについての比較研究に[1,7,15]がある。

14）脚注12)，13)でのべたパラメトリックな検定以外にノンパラメトリックな検定手法も提唱されている。一例としてピーク検定[13]があげられる。とりわけ誤差項の分布が正規性を満足していない場合にはこれらの手法の適用を考える必要がある。

15）第3節での分析の際には尤度比検定に加え，脚注13）にあげた検定手法についても適用を試みたが，手法により得られた結論は大きく異なった。本稿では残差等のプロットもあわせて考慮した結果，今回用いたデータの動きを判断する方法として尤度比検定の手法を選んだ。第3節での実証分析における検定の際の最大の問題点はその対象とする標本数が少なかったことであると考えられるが，尤度比検定の小標本における検定力はあまり強くないと考えられる[８]。これらの問題，すなわち手法により大きく検定結果が異なった点，小標本においていずれの検定手法を適用していくのがよいか等の点については今後の検討課題としたい。なお，尤度比検定の問題点，他のテストとの比較については[11]参照のこと。

ここで $T=T_1+T_2$, SSR は残差二乗和，その添字はそれぞれ対応するグループを表す。

第3節では，本節で考察を行った様々な理論的手法を日本の家計の所得階層別データに適用して，その消費行動を分析していく際分散不均一性が見られるか，見られた場合どのような手だてにより推定結果が改善されるか，検討を行った。解決策として各種のウエイトを用いた推定を行ったが，対数変換により分散不均一性が軽減される場合があることが知られているのであわせてその方法も試みた。またデータの動きを検討した際外れ値と考えられる値が存在した[20]のでその点についても考慮を行った。

3. 実証例における分散不均一性の検討

本節では，標準世帯（4人世帯で有業人員が1人である世帯）の消費支出行動に関する回帰分析における分散不均一性の問題について検討を行っていく。

先に『家計調査年報』の標準世帯（勤労者世帯）のデータを用いて，消費支出（CS）に対して定期的な収入（YR），ボーナスを中心とする臨時性の高い収入（YI）とその他の収入（YO）が与える影響が異なるかについて考察を行った[18]。本稿ではその分析の際分散不均一性が存在していなかったか，存在していた場合どのようにすれば問題の改善を図ることができるか検討していく。

分析に用いたデータは家計の年収により階級分けがなされており[16]，1963年から91年までの29年間にわたる年次データが利用可能である。

考察対象とするモデルは

$$\boldsymbol{CS}=\beta_0+\beta_1\boldsymbol{YR}+\beta_2\boldsymbol{YI}+\beta_3\boldsymbol{YO} \qquad \cdots\cdots(8)$$

である。

各年の所得階層別データに対して最小二乗法による推定を行った結果，尤度比検定統計量からほとんどの年で分散不均一性が観察されると判断されることが示された[17]（表1，②欄）。

このことを残差のプロットから観察してみる。図1は分散不均一性が検出されなかった1985年についての，図2は分散不均一性が観察された1991年についての(ア)残差（e_i）の動き，(イ)e_i^2 と $\hat{Y}_i$（被説明変数 CS の予測値）の関連，(ウ)$e_i/\Sigma|e_i|$ と $\hat{Y}_i/\Sigma\hat{Y}_i$ の関連を示したものである。それぞれの年の(ア)図を比較すると，85年では第1階級が他とは大きく異な

16）なお年により階級の数は異なる。詳細は表1，①欄参照のこと。

17）表1には $T_1=T_2$ の場合をあげた。T_1, T_2 の値を変化させた際いずれの場合にも分散不均一性が認められなかったのは85，89年のみであった。

表1　尤度比検定の結果（$T_1=T_2$のケース）

年	① 階級数	② OLS	③ WLS ($\sqrt{n_i}$)	④ WLS ($^1/_{TY}$)	⑤ 対数変換	⑥ 外れ値	⑦ OLS（外れ値）
63	16		*			13 16	
64	15				**	12 13	
65	15					12 15	
66	15					13 14	**
67	15				*		
68	15		*			1 15	**
69	16				**	1	
70	16					16	*
71	15		**			2	
72	14					1 14	
73	15		**			13 14	
74	16			**	*	14 16	**
75	16					16	
76	16		*	**	**	14 16	
77	16	**	**	**	**	14 16	*
78	16	**	**	*		3 16	**
79	17			**	**	16	*
80	17			**		15	**
81	17				**	16	
82	17				**	15 16	
83	17		**			1	
84	17		*			15	*
85	17	**	**			1	**
86	17			*	**	17	**
87	17	**	**	**	**	14 17	
88	16		**		**	1	**
89	16	**	**	*	*	14	*
90	16		*	**	*	13	*
91	17			**	**	15 17	**

**は「誤差項の分散が均一である」との仮説が有意水準5％で棄却されなかったことを示す。

*は同じ仮説が1％で棄却されなかったことを示す。

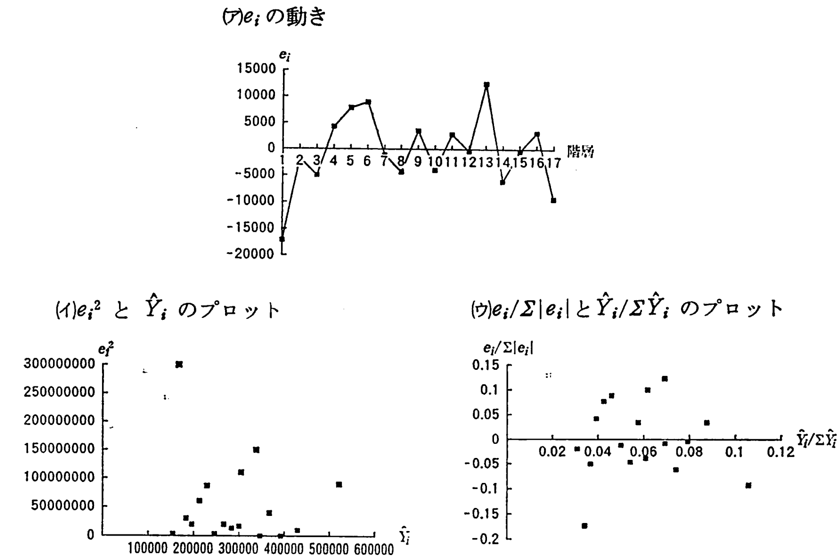

図1　1985年の OLS の結果

る動きを見せた以外は残差に傾向的な動きは見られなかった一方，分散不均一性が検出された91年では階級が大きくなるにつれ残差も大きくなる傾向が見られた。それぞれの(イ)図（あるいは91年の場合その一部を拡大した(イ)'図）より85年では外れ値の第1階級を除けば e_i^2 と $\hat{Y}_i$ に明示的な関係は見られないが，91年では外れ値を除いたデータについても $\hat{Y}_i$ が大きくなるにつれ e_i^2 が大きくなるという傾向が見てとれる。(ウ)図を見ても85年では基準化された残差と $\hat{Y}$ の間には関連は見られなかったが，91年では $\hat{Y}$ が大きくなるにつれ残差の広がりも大きくなっていることがわかる。

このように分散不均一性が検出された年では残差の動きに傾向的な動きが観察されたが，このような問題を解決するために 1) 各誤差項の分散が各階級に含まれる標本世帯数（n_i）の逆数に比例する場合，2)各誤差項の分散が総収入（TY）[18]の2乗値に比例する場

18) ここで，$TY=YR+YI+YO$ である。なお，各収入の項の2乗値に誤差分散が比例するケースについても推定を試みたが，残差のプロットの結果からも明らかなように，定期収入（YR）以外の収入項の場合は分散不均一性の問題が解決されるような結果はほとんど得られなかった。また，YR 項をウェイトに用いた場合の結果は表には示さなかったが，概ね TY をウェイトに用いた場合と結果は変わらなかった。誤差分散がTYに比例する場合についても検討を行ったが，この場合は分散不均一性が解消されるケースはほとんど見られなかった。

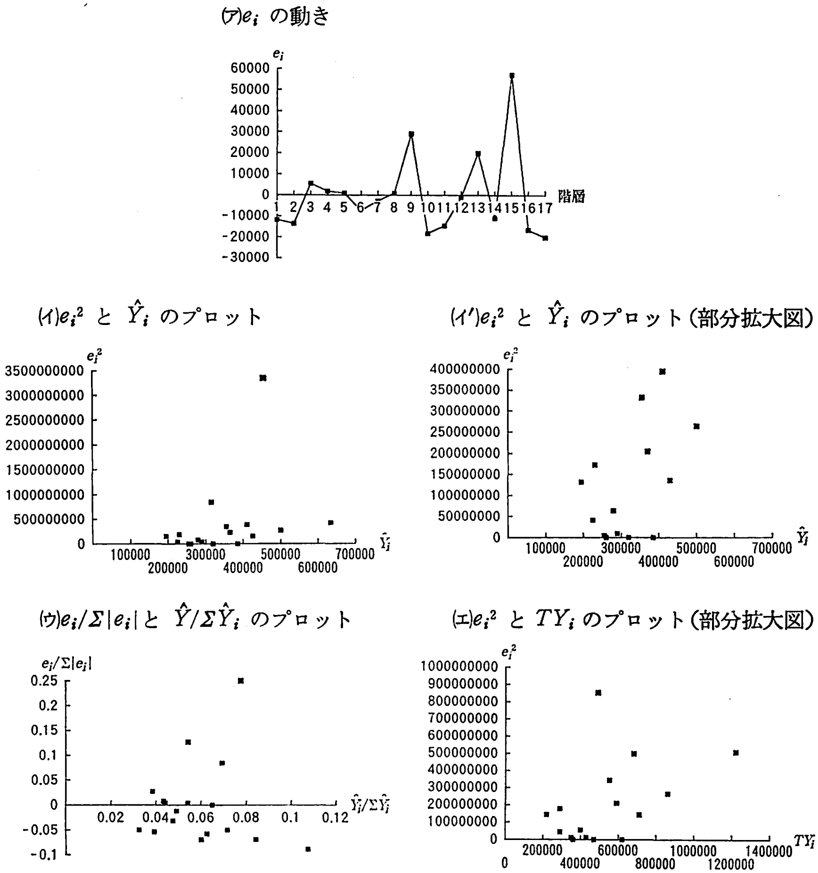

図2　1991年の OLS の結果

合，3)モデルを対数線形型に変形した場合を考えた。結果はそれぞれ表1の③，④，⑤に掲げた。

ウェイト付最小二乗法により推定したいずれの場合も最小二乗法の場合に比べ分散不均一性が解決されたケースが多くなっている。$\sqrt{n_i}$をウェイトに用いた場合（表1，③欄）には29年中14年で分散が均一であると仮定しても問題がないとの結果が得られた。図3に標本世帯数を考慮することにより分散不均一性の問題が改善された例として，1971年の場合をあげた。ウェイト付最小二乗法を用いた場合，最小二乗法では外れ値として観察された階級がウェイト付けによりさほど特異な動きを見せなくなっていること，残差と説明変

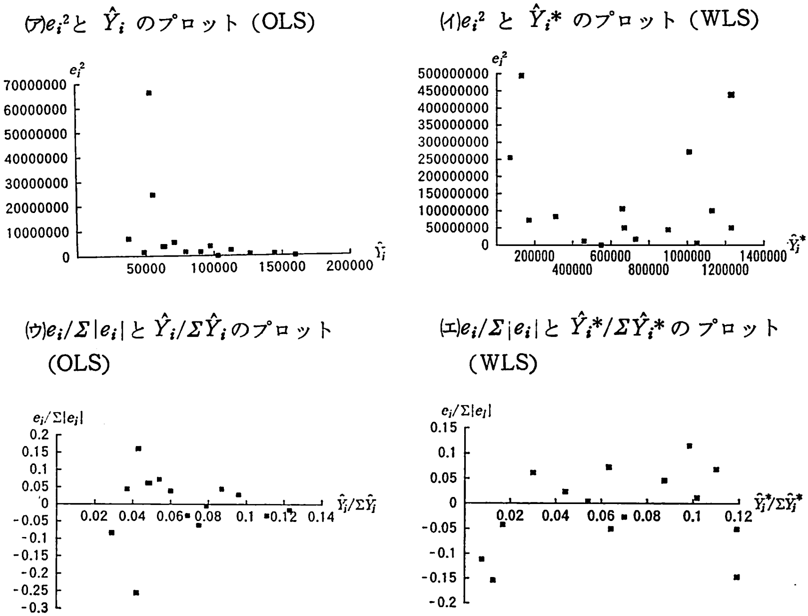

図 3　1971年の e_i^2 と $\hat{Y}$ のプロット

数の予測値の間の傾向的な動きが弱まっていることが観察される。

誤差分散が TY^2 に比例すると考えた場合は，標本世帯数を考慮した場合に比ベケース数としてはあまり改善が見られなかった[19]（表１，④欄）。TY^2 について考慮して改善が見られた1991年の残差の動きを見ると，外れ値である第15階級以外の階級値は収入変数の増加に応じて増加していることがわかる（図２，(エ)）。このような残差の動きが見られた

19）このことに関連して，ホワイト検定（(ア)$e_i^2=\beta_1+\beta_2TY$，(イ)$e_i^2=\beta_1+\beta_2TY+\beta_3TY^2$，(ウ)$e_i^2=\beta_1+\beta_2YR+\beta_3YI+\beta_4YO$)，グレイサー検定（(エ)$|e_i|=\beta_1+\beta_2TY$，(オ)$|e_i|=\beta_1+\beta_2\sqrt{TY}$，(カ)$|e_i|=\beta_1+\beta_2(1/TY)$）を用いて，誤差分散が収入変数に依存しているかどうかについての検定を行った。その結果 TY や TY^2，$\sqrt{TY}$ に関する検定の場合には29ケースのうち４～７ケースのみで，各収入変数を項とする検定(ウ)や $1/TY$ についての検定(カ)では２～４ケースのみでしか分散不均一性は検出されなかった。

これらの結果は，収入変数をウエイトに用いた場合にさほど分散不均一性の問題が改善されなかったことの裏付けと考えられる。

年では，収入変数[20]の二乗値に誤差分散が比例すると考えてウエイト付最小二乗法を行った場合分散不均一性の問題が解決されることが観察された。

一方，対数変換を行った場合にもかなりの改善が見られ（表1，⑤欄），29年中15年で誤差分散が均一であるとの仮定が棄却されなかった。分散不均一性の問題が解決された場合には，図4にあげた1964年に見られるように，原データを最小二乗法で推定した場合には誤明変数の増加とともに残差が指数的に増える傾向が強く見られたが，対数変換を行うことによりこの傾向は弱まった[21]。

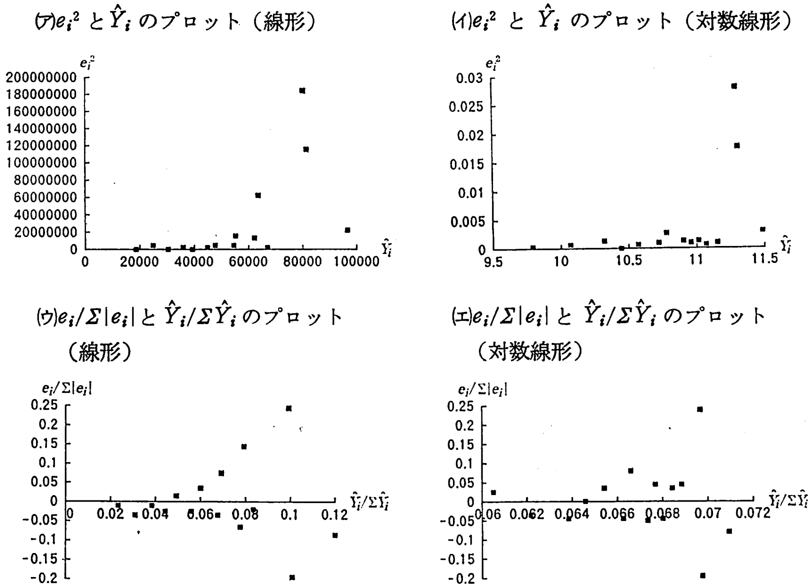

図4　1964年の OLS の結果

20）ここでは総収入 TY について考慮した結果をあげたが，プライス＝ハウタッカー[33]は「消費支出は $E[TY]$ の2乗値に比例する」と考えている。この点について2段階の推定法を用いて考慮を行った[28]。結果は TY に関するウエイトを用いた場合に比べ概ね同じであったが，④の結果に加え84，85年で誤差分散が均一であると仮定してもよいとの結論が得られた。

21）1964年について対数変換を行ったモデルでは，プロットを見ても明らかであるが，第13，14階級値が大きな外れ値として検出された。これらの点を除いて再推定を行うと誤差分散のプロットにはほぼ規則性が見られなくなった。

さていくつか示した残差のプロットからも明らかなように，今回用いた標準世帯の階級データにはほとんどの年について外れ値と考えられる値がいくつか含まれていた[20]。次にこの問題を考慮して分析を行っていく。表１の⑥欄にはそれぞれの年で(8)式のモデルを推定した場合に外れ値と判断された階級値をあげた22)。これらの階級を除いて最小二乗法で推定した際に誤差分散が均一と考えられるかどうかについての判断結果を表１，⑦欄にあげた。これより，外れ値を推定対象から取り除くことにより分散不均一性の問題が解決されるケースがかなり見られることがわかる。図５には1966年の全ての階級値を用いた場合，外れ値と考えられた階級を除いた場合の残差の動きを示した。全階級を用いた場合には規則性が見られる一方，外れ値を除いた場合には残差の動きはランダムなものとなっていることがわかる。

さて，外れ値とは（Y軸方向に測った）残差の大きな観測値のことをさすが，回帰分析の結果にはX軸方向に測った際に中心から大きく乖離した点（作用点）も影響を与える。

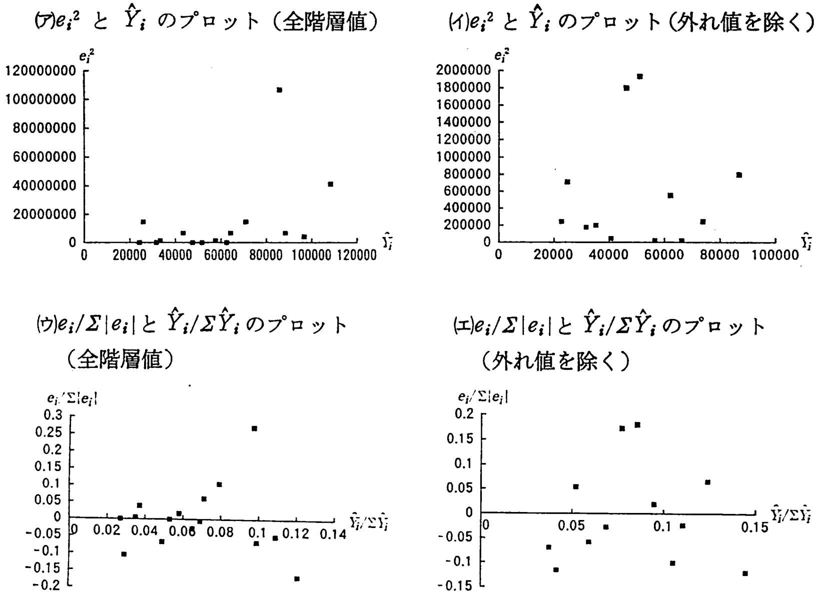

図５　1966年の OLS の結果

22）外れ値については[20]で行った検討結果を用いた（p. 215，表1）。なお，この表の作用点の欄について誤りがあったため，今回付表として訂正を行ったものを再掲した。

すなわち回帰直線はこれらの作用点に引き寄せられるように推定されてしまうことが知られている[30]。外れ値のみならず作用点あるいは影響点（付表参照のこと）についても考慮した場合，いっそう分散不均一性の問題が解決される場合が見られた。例えば，図6にあげる1983年の場合，外れ値である第1階級のみを除いても分散不均一性は解決されなかったが，作用点である第15，17階級をも除いた場合大きな改善が見られた。

さて，これまで誤差分散の構造に着目して分析してきたが，それぞれの手法により得られた推定結果の妥当性という観点から検討を行ってみる。例として1991年の推定結果について考察を行うことにする。表2は表1で検討を行った様々な手法による推定結果である。これらの結果は対応するモデルが異なるので比較するのが困難ではあるが，各モデルの各収入項に着目して考察を行ってみる。表2の2種類の最小二乗法，標本世帯数でウェイト付けしたウェイト付最小二乗法の場合，収入項の係数はいずれも当該収入の消費性向を示している。この場合推定値は一般には0と1の間の値をとると考えられる。この基準から検討すると標本世帯数でウェイト付けしたウェイト付最小二乗法の場合のみで妥当な推定結果が得られていると考えられる。

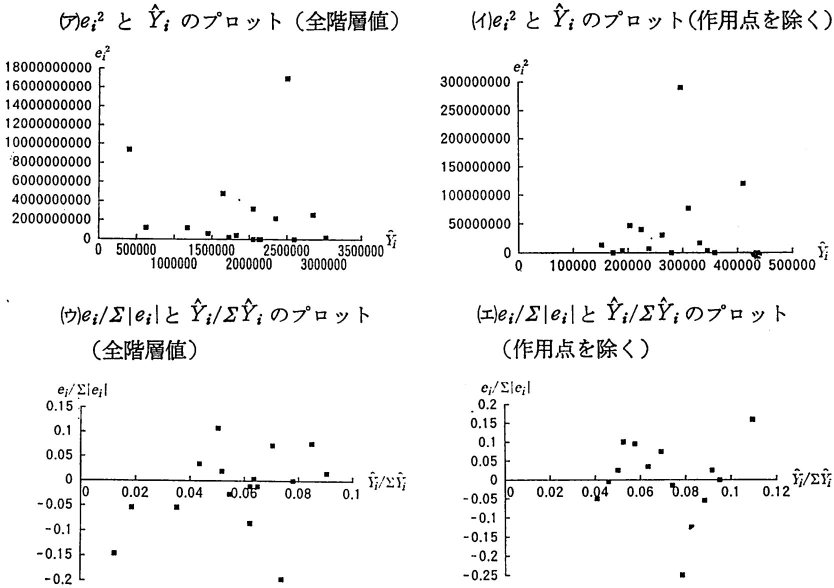

図6　1983年の OLS の結果

表2　各推定手法による推定結果

	OLS	WLS($\sqrt{n_i}$)	WLS($^1/_{TY}$)#	対数変換#	外れ値#
定数項	109946*	83810.6	98903*	3.2901**	69508
YR項	0.3801	0.4732	0.3517	0.6005**	0.4852
YI項	0.7159	0.5013	0.76	0.1058	0.3598
YO項	-0.4224	0.8019	0.5509	0.0507	2.2703

#はその推定手法を用いた際誤差分散が均一であるとの仮説が棄却されなかったことを示す。

*はその推定値が有意水準1％で，**は5％で有意であったことを示す。

総収入でウェイト付けしたウェイト付最小二乗法の場合，各収入項の係数は総収入にしめる各収入項目の比率がいかに総収入にしめる消費支出の比率に影響しているかを示している。また，対数変換されたモデルでは各収入項の係数は消費支出に対する当該収入の弾力性を示している。いずれの場合も，今回用いた大きな収入項目を考えるならば，マイナスの値や1より非常に大きな値はとりにくい考えられるが，表2の結果はこれらの推論に合致するものである。

しかしながら，それら2種類のウェイト付最小二乗法，対数変換したモデルといった推定値が一応の基準を満たしたかに見えた手法を用いても，ほとんどの場合に得られた推定値は有意ではなく，信頼に足る安定した推定結果とは考えにくい。

以上，本稿では，分散不均一性の問題については理論的測面から検討を行い，その結果を標準世帯データに適用して実証的に検討したが，今回用いた実証例においては分散不均一性の問題を解決するという点からすると一定の効果をあげたものの，妥当な推定結果を得るという観点からすると満足のいく結果をあげることはできなかった。

4.　結語にかえて

本稿では，回帰分析における分散不均一性の問題について考察を行った。

分散不均一性の問題は経済データとりわけクロスセクションデータを扱っていく上で問題となることが多いが，分散不均一性が存在する場合に推定結果にどのような問題が生じるか，それを解決するためにはどのような手だてを講ずればよいか，さらに分散不均一性を検出する方法にどのようなものがあるかについて検討した。また，これらの手法を具体的に標準世帯の支出データに適応し，分散不均一性に関する様々な検討を行った。

第3節の最後にも述べたように，今回用いた実証例では分散不均一性の検出，およびそ

れへの対処という点からは一定の成果をみたが，信頼にたる推定結果を得るという点からは必ずしも満足のいく結果を得ることはできなかった[23)]。この理由としては以下のことが考えられる。

今回考えた実証例の枠組みは家計の階層データを用いての消費関数の計測であったため，分散不均一性の構造に関し事前情報があるものとして非常に限定した考慮を行った。また，最小二乗法に代わる推定法としても非常に限られた形のものしか考察し得なかった[24)]。

さらに分散不均一性の検出という面からしても，本稿の分析は非常に初歩的，試験的な段階のものであり，扱い得ていない検討課題が数多く残った。とりわけ今回用いたような小標本データにおいては検討すべき多くの課題が残っていると考えられる[25)]。

しかしながら，ある程度誤差項の分散不均一性の問題が解決されたにも関わらず推定結果が芳しくなかったことの背景にはより根元的な問題，たとえば誤差項の分布が正規性を満足していないといった問題の可能性があるとも考えられる。この点についての検討，さらにはその場合に応じての頑健な（robust）推定手法[22, 31]についての考察については，記して今後の課題としたい。

参考文献

〔1〕Ali, M. M. and C. Giaccotto (1984) A study of several new and existing tests for heteroscedasticity in the general linear model, *Journal of Econometrics*, vol. 26, 355-73.

〔2〕伴金美・中村二朗・跡田直澄（1988）『エコノメトリックス』，有斐閣.

〔3〕Box, G. E. P. and D. R. Cox (1964) An analysis of transformations, *Journal of the Royal Statistical Society* (*B*), vol. 26, pp. 211-52.

23）この点については[18]（pp. 153-5）でも問題にしている。[18]では分散不均一性の見られる可能性が高いであろうと想定して直接ウェイト付最小二乗法の適用を行ったが，結果はあまり芳しくなかった。本稿では分散不均一性についてより詳細に検定を行い，前回の結果からは一定の改善を見たものの，良好な推定結果を得るところまでにはいたらなかった。

24）たとえば予備検定統計量[14, 32, 26]については，取り扱い得なかった。また，対数変換の問題についても，ボックス＝コックス変換[3]を始めとする線形か対数線形かの関数形の選択の問題についても今回はふれることができなかった[28, chap 5.6]。

25）特に，小標本におけるバイアスの問題について考慮していく必要があると考えられる。たとえば，[21]参照のこと。

〔4〕 Breusch, T. S. and A. R. Pagan (1979) A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, vol. 47, pp. 1287-94.

〔5〕 Chatterjee, S. and B. Price (1991) *Regression Analysis by Example* (2nd ed.), John Wiley and Sons.

〔6〕 Davidson, R. and J. J. MacKinnon (1993) *Estimation and Inference in Econometrics,* Oxford Univ. Press.

〔7〕 Evans, M. A. and M. L. King (1988) A further class of tests for heteroscedasticity, *Journal of Econometrics,* vol. 37, 265-76.

〔8〕 Fomby, T. B., Hill, R. C. and S. R. Johnson (1984) *Advanced Econometric Methods,* Springer-Verlag.

〔9〕 Glejser, H. (1969) A new test for homoscedasticity, *Journal of the American Statistical Association,* vol. 48, pp. 817-38.

〔10〕 Godfrey, L. G. (1978) Testing for multiplicative heteroskedasticity, *Journal of Econometrics,* vol. 8, pp. 227-36.

〔11〕 Godfrey, L. G. (1988) *Misspecification Test in Econometircs,* Cambridge Univ. Press.

〔12〕 Goldfeld, S. M. and R. E. Quandt (1965) Some tests for homoscedasticity, *Journal of the American Statistical Association* vol. 60, pp. 539-47.

〔13〕 Goldfeld, S. M. and R. E. Quandt (1972) *Nonlinear Methods in Econometrics,* North-Holland.

〔14〕 Greenberg, E. (1980) Finite sample moments of a preliminary test estimator in the case of possible heteroscedasticity, *Econometrica,* vol.48, pp. 1805-13.

〔15〕 Griffiths, W. E. and K. Surekha (1986) A Monte Carlo evaluation of the power of some tests for heteroscedasticity, *Journal of Econometrics,* vol. 31, 219-31.

〔16〕 Gujarati, D. (1988) *Basic Econometrics* (2nd ed.), McGraw-Hill.

〔17〕 Gujarati, D. (1992) *Essentials of Econometrics,* McGraw-Hill.

〔18〕 橋本紀子 (1993)「臨時性収入の家計消費への影響に関する一考察：標準世帯の場合」, 関西大学『経済論集』43, pp. 613-626.

〔19〕 橋本紀子 (1994)「回帰分析における多重共線性の検出と処理：標準世帯の支出行動に関する実証の場合」, 関西大学『経済論集』44, pp. 59-79.

〔20〕 橋本紀子 (1994)「回帰分析における外れ値についての一考察：標準世帯の支出行動に関する場合」, 関西大学『経済論集』44, pp. 207-30.

〔21〕 Honda, Y. (1988) A size correction to the Lagrange multiplier test for heteroskedasticity, *Journal of Econometrics,* vol.38, pp. 375-86.

〔22〕 Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and T.-C.Lee (1985) *The Theory and Practice of Econometrics* (2nd ed.), John Wiley and Sons.

〔23〕 Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. and T.-C. Lee (1988) *Introduction to the Theory and Practice of Econometrics* (2nd ed.), John Wiley and Sons.

〔24〕 Katona, G., et. al. (1954) *Contributions of Survey Methods to Econometrics,* Columbia Univ. Press.

〔25〕 Kennedy, P. (1992) *A Guide to Econometrics* (3rd ed.), Blackwell.

〔26〕 Kmenta, J. (1986) *Elements of Econometrics* (2nd ed.), Macmillan.

〔27〕 Koenker, R. (1981) A note on studentizing a test for heteroscedasticity, *Journal of Econometrics,* vol. 17, pp. 107-12.

〔28〕 Maddala, G. S. (1988) *Introduction to Econometrics,* Macmillan.

〔29〕 MacKinnon, J. G. and H. White (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties, *Journal of Econometrics,* vol. 29, pp. 305-25.

〔30〕 蓑谷千凰彦（1992）『計量経済学の新しい展開』多賀出版

〔31〕 蓑谷千凰彦（1992）『計量経済学における頑健推定』多賀出版

〔32〕 Ohtani, K. and T. Toyoda (1980) Estimation of regression coefficients after preliminary test for homoscedasticity, *Journal of Econometrics,* vol. 12, pp. 151-9.

〔33〕 Park, R. (1966) Estimation with heteroskedastic error terms, *Econometrica,* vol. 34, pp. 888.

〔34〕 Prais, S. J. and H. S. Houthakker (1955) *The Analysis of Family Budgets,* Cambridge Univ. Pres.

〔35〕 Stewart, J. (1991) *Econometrics,* Philip Allan.

〔36〕 Taylor, W. E. (1978) The heteroscedastic linear model: exact finite sample results, *Econometrica,* vol. 46, pp. 663-75.

〔37〕 Theil, H. (1971) *Principles of Econometrics,* North-Holland.

〔38〕 Waldman, D. M (1983) A note on algebraic equivalence of White's test and a variation of the Godfrey/Breusch-Pagan test for heteroskedasticity, *Economics Letters,* vol. 13, pp. 197-200.

〔39〕 White, H. (1980) A heteroskedasticity consistent covariance matrix estimator and a direct test of heteroskedasticity, *Econometrica,* vol. 48, pp. 817-38.

付表 回帰診断・影響分析の結果

年	階級数	外れ値		作用点		影響点	
63	16		(11) 13 16		13 16		11 13 15 16
64	15		12 13				11 12 13 15
65	15		12 15	1	10		12 15
66	15		13 14 15		14 15		(12) 13 14 15
67	15			1	15		13 (15)
68	15	1	15	1	14 15	1	12 (14) 15
69	16	1			14 16	1	15
70	16		16	1	15 16	1	15 16
71	15	2			14 15	1, 2	14 15
72	14	1 (2)	14			1, 2	14
73	16		13 14		15		13 14 15
74	16		14 16		16		14 16
75	16		16		16		13 16
76	16		14 16		13		13 14 16
77	16		14 16		15 16	2	14 15 16
78	16	3	16			3	11 13 16
79	17		16 (17)		17		14 (15) 16 17
80	17		15		13 17		13 15 17
81	17		(15) 16	1	17	1	16 17
82	17		15 16 (17)		16 17		14 15 16 17
83	17	1			15 17	1	(15) 17
84	17		15		16 17		13 15 (16) 17
85	17	1			15 16	1	13 16 17
86	17		17		14 16		14 16 17
87	17		14 17		17		14 17
88	16	1	(14)	1		1	13 14
89	16		14		15 16		11 14 16
90	16		13		14		13
91	17		15 17		17		15 17

（ ）内の階級は，一部の指標で問題があると判断されたことを示す。
外れ値，作用点の欄の太字は，その階級が影響点と判断されていることを示す。
（注）[20]で今回用いたのと同じデータを用いて外れ値，作用点，影響点の検出を行った。その結果をまとめた表1（[20]，p.215）の作用点の欄に誤りがあったため，今回ここに，訂正したものをあげる。