

計量経済モデルのパラメータ推定における異常値検出の一方法

原 田 桂 一 郎

目 次

- 1 はじめに
- 2 計量経済モデルのパラメータ推定におよぼす観測値の影響
- 3 異常値検出の幾何学的方法
- 4 マスク効果の対策
- 5 異常値の検定
- 6 む す び

1 はじめに

経済現象を経済変数間の関数関係として捉えると、それは回帰モデルの型式を整えた計量経済モデルによって表現される。そして、パラメータの値は経済変数のデータを基にして統計的方法により推定される。

その推定には、推定量の統計的特性が良好であることから、最小二乗法が採用される。しかし、粗悪なデータは最小二乗推定に深刻な問題を発生させる。大標本の場合でも、唯一の異常値がパラメータの最小二乗推定値、標準誤差、 t 統計量の値を歪めてしまう。最小二乗法は変数データの中の異常値に対して頑健ではない。

粗悪なデータに対し、その推定量の特性が損われない頑健推定^①の方法が開発されているが、現在の研究段階では若干の問題が残されている。

偏りのない安定したパラメータの値を推定するには、変数データを吟味し、異常値を除去しておくことが必要である。

本稿では、まず、各々の観測値がモデル推定に与える影響を考察する。そし

計量経済モデルのパラメータ推定における異常値検出の一方法（原田）

て、その影響度を測る統計量が、異常値に反応を示す要素によって構成されていることから、直接データを診断し異常値を検出する方法を提示する。それは、個々の観測値と他の観測値集団の幾何学的距離を尺度とし、それに統計的検定を加えるものである。

注

- ① 頑健推定については、拙稿「線型回帰モデルの頑健推定に関する一考察」、政経論叢、59号、pp.1—20.

頑健推定を回帰分析に適用することへのコメントは Carroll〔4〕, Cook and Weisberg〔10〕に詳しい。

2 計量経済モデルのパラメータ推定におよぼす観測値の影響

計量経済モデルとして次の標準線型回帰モデルを採用する。

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (2-1)$$

\mathbf{y} : 従属変数ベクトル, $\mathbf{y} = (y_1 \cdots y_n)'$,

\mathbf{X} : 独立変数行列 ($n \times p$),

$\mathbf{X} = (\mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_n)'$, $\mathbf{x}_i (1 \times p)$, $\text{rank}(\mathbf{X}) = p$,

$\boldsymbol{\beta}$: 回帰係数ベクトル, $\boldsymbol{\beta} = (\beta_1 \cdots \beta_p)'$,

\mathbf{u} : 誤差ベクトル $\mathbf{u} = (u_1 \cdots u_n)'$,

そして、(2-1) が最小二乗推定されたものを次のように標記する。

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}, \quad (2-2)$$

\mathbf{y} : (2-1) に同じ,

\mathbf{X} : (2-1) に同じ,

\mathbf{b} : $\boldsymbol{\beta}$ の最小二乗推定量,

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

計量経済モデルのパラメータ推定における異常値検出の一方法（原田）

\mathbf{e} : 残差ベクトル, $\mathbf{e}=(e_1 \cdots e_n)'$

$$S^2 = \mathbf{e}'\mathbf{e}/n-p$$

モデルおよび設定された仮定の妥当性が認められると、パラメータの最小二乗推定におよぼす変数の観測値の影響が問題となる。推定量の統計的特性が優れており、また計算が簡便であるから、パラメータは一般に最小二乗推定される。しかし、この推定法は、所定の母集団分布とは異なる別の分布から抽出され、並外れた値となる観測値（以下では異常値と称す）の影響により、推定値に偏りが生ずる弱点を有している。

各々の観測値が推定値に与える影響^②は、それを除いて推定した回帰係数、その分散、あるいは予測値と、すべての観測値から得たそれらの推定値の差によって調べられる。

従属変数の異常値（これは残差の値を大とする）だけでなく、独立変数の異常値もパラメータの最小二乗推定値に影響を与える。

Andrews and Pregibon〔2〕は残差を最大とする観測値 (\mathbf{x}_i, y_i) でなく、残差を二番目に大とする観測値が推定値に最も大きな影響を及ぼしている事例を示している。これは独立変数の異常値が推定値を偏らせることを示唆するものである。

予測値 $\hat{\mathbf{y}}$ は

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}, \quad \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

で与えられる。行列 \mathbf{H} の対角要素を h_{ii} , 非対角要素を h_{ij} と標記すると、予測値 $\hat{\mathbf{y}}$ の各要素は

$$\begin{aligned} \hat{y}_i &= h_{ii}y_i + \sum_{j \neq i} h_{ij}y_j, \\ h_{ii} &= \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'_i, \\ h_{ij} &= \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'_j, \end{aligned} \tag{2-3}$$

計量経済モデルのパラメータ推定における異常値検出の一方法（原田）

となる。また残差の各要素は

$$\begin{aligned} e_i &= y_i - \hat{y}_i \\ &= (1 - h_{ii})y_i - \sum_{j \neq i} h_{ij}y_j \end{aligned} \quad (2-4)$$

であるから、残差は h_{ii} の減少関数である。 h_{ii} が大となるほど異常値による残差を大とする効果を減ずる。

Huber [16] は行列 \mathbf{H} の対角要素 h_{ii} が大であるポイントが独立変数行列の異常値（行）に該当することを指摘している。さらに、Hoaglin and Welsch [13] は h_{ii} が独立変数の異常値の検出に適した量であることを示している。

回帰モデルの推定結果におよぼす観測値の影響を、 \mathbf{H} 行列の情報と \mathbf{y} ベクトルの情報を結びつけた量によって調べる。

ここで、 (i) は観測値 (\mathbf{x}_i, y_i) を除いた推定値を表すものとすれば、次なる予測値の差

$$\begin{aligned} \hat{y}_i - \hat{y}_i(i) &= \mathbf{x}_i \mathbf{b} - \mathbf{x}_i \mathbf{b}(i) \\ &= \mathbf{x}_i [\mathbf{b} - \mathbf{b}(i)] \\ &= e_i (h_{ii} / (1 - h_{ii})), \\ \therefore \mathbf{b} - \mathbf{b}(i) &= \frac{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i' e_i}{1 - h_{ii}} \end{aligned} \quad (2-5)$$

は、観測値 (\mathbf{x}_i, y_i) が回帰モデルの推定に与える影響の大きさを測る量となる。

Cook [5] は観測値の影響の大きさを測る量として、

$$C_i = [\mathbf{b} - \mathbf{b}(i)]' \mathbf{X}' \mathbf{X} [\mathbf{b} - \mathbf{b}(i)] / p S^2 \quad (2-6)$$

を与え、さらに、この量が

計量経済モデルのパラメータ推定における異常値検出の一方法（原田）

$$C_i = p^{-1} t_i^2 (h_{ii}/1 - h_{ii}), \quad (2-7)$$

$$t_i = e_i / S(1 - h_{ii})^{1/2}$$

となることを示している^②。

(2-5), (2-7) のいずれも、従属変数の異常値に反応する e_i と独立変数の異常値に反応する $h_{ii}/1 - h_{ii}$ (h_{ii} が大ならばこの比も大) の積の型である。両変数の観測値 (\mathbf{x}_i, y_i) のどちらか、あるいは双方が異常値であるとき、これらの統計量の値は大となる。それは、異常値が回帰モデルのパラメータの推定値を偏らせるからである。

注

- ② 推定量への観測値の影響は理論的には influence function で表現されるが、現実
にこれを構築することは困難である。実用的な型式は Cook and Weisberg [9] で
提示されている。
- ③ Cook の与えた統計量は、誤差項に正規性を仮定すると、F 統計量になる。

3 異常値検出の幾何学的方法

Andrews and Pregibon [2] は従属変数ベクトル \mathbf{y} を独立変数行列 \mathbf{X} に付加した、 $\mathbf{Z} = [\mathbf{X}, \mathbf{y}]$ 行列に基づいて、両変数の異常値の影響を考察している。まず、 \mathbf{y} ベクトルの異常値 y_i のポイントに該当する \mathbf{Z} 行列の行 (\mathbf{x}_i, y_i) の削除は、残差平方和 $\mathbf{e}'\mathbf{e}$ の値を大きく減じる。 $\mathbf{e}'\mathbf{e}$ は \mathbf{y} の異常値を含む行 (\mathbf{x}_i, y_i) を検出する統計量である。次に、 \mathbf{X} 行列の等 i 行 \mathbf{x}_i を削除した場合、正方向行列 $\mathbf{X}'\mathbf{X}$ のスカラー量 $|\mathbf{X}'\mathbf{X}|$ が変化する。 \mathbf{Z} 行列の異常値に該当する行 (\mathbf{x}_i, y_i) はパラメータの最小二乗推定値 \mathbf{b} およびその分散 $\text{Var}(\mathbf{b})$ に大きく影響する。

これら二つの効果を結びつけるには、観測値 (\mathbf{x}_i, y_i) の削除がもたらす次の量

$$\mathbf{e}'\mathbf{e} / |\mathbf{X}'\mathbf{X}|$$

計量経済モデルのパラメータ推定における異常値検出の一方法（原田）
 の変化の大きさを計算することである。また

$$|\mathbf{Z}'\mathbf{Z}| = \mathbf{e}'\mathbf{e} |\mathbf{X}'\mathbf{X}| \quad (3-1)$$

であるから④， \mathbf{Z} 行列において異常値を含む行 $(\mathbf{x}_i, \mathbf{y}_i)$ の削除は $|\mathbf{Z}'\mathbf{Z}|$ の値を大きく減じる。

独立変数と従属変数の観測値を並べた \mathbf{Z} 行列を直接診断することにより両変数の異常値を検出する方法が考えられる。

Andrews and Pregibon [2] は，観測値 $(\mathbf{x}_i, \mathbf{y}_i)$ の削除による相対的变化を測るため，

$$R(i) = \frac{|\mathbf{Z}'(i)\mathbf{Z}(i)|}{|\mathbf{Z}'\mathbf{Z}|} \quad (3-2)$$

を与えている。なお， $\mathbf{Z}(i)$ は \mathbf{Z} 行列から第 i 行を削除したものである。 $R(i)$ が小であるほど，観測値 $(\mathbf{x}_i, \mathbf{y}_i)$ はパラメータの推定値に大なる影響を与えるものと判定される。

ところで， \mathbf{Z} 行列の第 i 行（ベクトル） $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i)$ と，この \mathbf{z}_i を除いた他のすべての行ベクトルについての平均ベクトル $\bar{\mathbf{z}}(i)$ の距離 $d(\mathbf{z}_i, \bar{\mathbf{z}}(i))$ によって，各々の観測値と他との相異の程度を測ることができる。両変数データのいずれか，あるいは両方に異常値が存在すれば，この距離が大となる。

ここで， \mathbf{Z} 行列の観測値 $(\mathbf{x}_i, \mathbf{y}_i)$ の平均をとったベクトルを

$$\bar{\mathbf{z}} = \left(\frac{1}{n} \sum_{i=1}^n x_{i1}, \frac{1}{n} \sum_{i=1}^n x_{i2}, \dots, \frac{1}{n} \sum_{i=1}^n x_{ip}, \frac{1}{n} \sum_{i=1}^n y_i \right)$$

とし，それぞれの行の要素とこの平均ベクトルの要素の差（平均値からの偏差）を並べた行列

$$\begin{aligned} \tilde{\mathbf{Z}} &= [(\mathbf{z}_1 - \bar{\mathbf{z}})(\mathbf{z}_2 - \bar{\mathbf{z}}) \cdots (\mathbf{z}_i - \bar{\mathbf{z}}) \cdots (\mathbf{z}_n - \bar{\mathbf{z}})]' \\ &= (\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2, \dots, \tilde{\mathbf{z}}_i, \dots, \tilde{\mathbf{z}}_n)' \end{aligned}$$

をつくる。

\tilde{Z} 行列から、その第 i 行を除いたものを、 $\tilde{Z}(i)$, $(n-1) \times (p+1)$ 行列と標記する。 $\tilde{Z}(i)$ 行列の行ベクトルの平均を、 $\tilde{z}(i)$, $(p+1)$ ベクトルと標記する。 $\tilde{Z}(i)$ 行列のそれぞれの行の要素と $\tilde{z}(i)$ ベクトルの要素の差を並べた行列を $\tilde{\tilde{Z}}(i)$ と標記する。また、 $\tilde{Z}(i)$ の共分散行列は $\tilde{Z}'(i)\tilde{Z}(i)$ である。

\tilde{z}_i ベクトルと $\tilde{z}(i)$ ベクトルの Mahalanobis 汎距離は、

$$M(i) = (n-2)(\tilde{z}_i - \tilde{z}(i))[\tilde{Z}'(i)\tilde{Z}(i)]^{-1}(\tilde{z}_i - \tilde{z}(i))' \quad (3-3)$$

のように与えられる (Belsley et al. [3], p. 27)。これは \tilde{z}_i と $\tilde{z}(i)$ のユークリッド距離の二乗を一般化したもので、逆行列 $[\tilde{Z}'(i)\tilde{Z}(i)]^{-1}$ によって相関のある変数の成分の影響を除去している。

すべての観測値についてこの $M(i)$ を計算し、その値が大となるポイントの観測値 (x_i, y_i) が異常値と考えられる。

注

$$\begin{aligned} \textcircled{4} \quad \tilde{X}'e &= \tilde{y} - \tilde{X}\tilde{b} \\ &= [I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}']\tilde{y}, \\ [I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'] &\text{は巾等行列であるから,} \\ e'e &= \tilde{y}'[I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}']\tilde{y}. \\ Z &= [\tilde{X} \quad \tilde{y}], \\ Z'Z &= \begin{bmatrix} \tilde{X}'\tilde{X} & \tilde{X}'\tilde{y} \\ \tilde{y}'\tilde{X} & \tilde{y}'\tilde{y} \end{bmatrix}, \\ |Z'Z| &= |\tilde{X}'\tilde{X}| \cdot |\tilde{y}'\tilde{y} - \tilde{y}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}| \\ &= |\tilde{X}'\tilde{X}| \cdot |\tilde{y}'[I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}']\tilde{y}| \\ &= |\tilde{X}'\tilde{X}| e'e \end{aligned}$$

4 マスク効果の対策

複数の異常値が存在すると、それらが連帯してパラメータの推定値を偏らせる。この場合、各々の観測値について推定値への影響の度を調べていく方法

では、異常値が検出できないことがある。

異常値が複数の場合、一行ずつ削除して診断する方法では、一行（異常値）の削除による推定値の変化の大きさが、他の異常値の潜在的効果により減少し、異常値検出は不可能となる。一つの異常値の削除による推定値の変化が、他の異常値の効果によって相殺される、これがマスク効果である。

Mahalanobis 汎距離による異常値の検出法においては、 $\tilde{\mathbf{Z}}$ から複数個の行を削除して距離を計算することにより、マスク効果に対処する。

$\tilde{\mathbf{Z}}$ 行列で複数の行が異常値ベクトルと考えられる場合、それらを一組として I と標記する。共分散行列と列平均の計算から I を削除し、 $\tilde{\mathbf{Z}}'(I)\tilde{\mathbf{Z}}(I)$ および $\tilde{\mathbf{z}}(I)$ を算出する。 $\tilde{\mathbf{z}}_i$ と $\tilde{\mathbf{z}}(I)$ の Mahalanobis 汎距離は

$$M(i, I) = (n-2)(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}(I))[\tilde{\mathbf{Z}}'(I)\tilde{\mathbf{Z}}(I)]^{-1}(\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}(I))' \quad (4-1)$$

となる。 $M(i, I)$ を用いて異常値を検出するには次のような反復計算を行う (Belsley et al. [3], p.38)。

まず、 $M(i)$ においてその二つの最大値を生じた二行が構成する組を $I_2^{(1)}$ と標記し、これを “starting set” とする。 $M(i, I_2^{(1)})$ により、最も大なる値二つを生じた二行を $I_2^{(2)}$ と標記すると、 $I_2^{(1)} = I_2^{(2)}$ となれば反復を止める。 $I_2^{(1)} \neq I_2^{(2)}$ であれば $M(i)$ においてマスク効果が発生したと判断できる。 $M(i, I_2^{(2)})$ を計算し、その二つの最大値を生ずる二行、 $I_2^{(3)}$ が、 $I_2^{(2)} = I_2^{(3)}$ であれば、ここで反復を止める。 $I_2^{(2)} \neq I_2^{(3)}$ ならば $M(i, I_2^{(2)})$ においてもマスク効果があったことになる。一般に $M(i, I_2^{(k)})$ においての二最大値を示す二行、 $I_2^{(k+1)}$ を得るまで反復を行って、 $I_2^{(k+1)} = I_2^{(k)}$ となったときに停止する。

5 異常値の検定

前節までに述べた Mahalanobis 汎距離、 $M(i)$ もしくは $M(i, I_2^{(k)})$ では、その値を大とする $\tilde{\mathbf{Z}}$ 行列の行（観測値）が異常値と考えられるが、合理的な判

定は統計的仮説検定に依る。

その検定のために、二つの母集団分布から抽出された標本の平均の間の差を検定する統計量として Hotelling T^2 を用いる。この場合の帰無仮説は、一つの正規母集団の平均と他の正規母集団の平均が等しいとするものである。ここで、二つの正規母集団の共分散行列は等しく未知であると仮定する。

いま、 p 次元ベクトル $u_j^{(k)}$ の標本、 $u_1^{(k)} \dots u_{N_k}^{(k)}$ は $N(\mu^{(k)}, \Sigma)$, $k=1, 2$, から抽出されたものとし、帰無仮説 $H_0: \mu^{(1)} = \mu^{(2)}$ を検定する。平均ベクトル $\bar{u}^{(k)}$ は

$$\bar{u}^{(k)} \sim N[\mu^{(k)}, (1/N_k) \Sigma] \quad (5-1)$$

であるから、

帰無仮説の下では

$$\sqrt{N_1 N_2 / (N_1 + N_2)} (\bar{u}^{(1)} - \bar{u}^{(2)}) \sim N(0, \Sigma) \quad (5-2)$$

である。さらに、

$$S = \frac{1}{N_1 + N_2 - 2} \left\{ \sum_{j=1}^{N_1} (u_j^{(1)} - \bar{u}^{(1)})(u_j^{(1)} - \bar{u}^{(1)})' + \sum_{j=1}^{N_2} (u_j^{(2)} - \bar{u}^{(2)})(u_j^{(2)} - \bar{u}^{(2)})' \right\} \quad (5-3)$$

とおけば、 $(N_1 + N_2 - 2) S$ は分散 $\sum_{j=1}^{N_1 + N_2 - 2} V_j V_j'$ として分布し、 V_j は $N(0, \Sigma)$ にしたがって分布する。かくして、分散比として Hotelling T^2 は

$$T^2 = \frac{N_1 N_2}{N_1 + N_2} (\bar{u}^{(1)} - \bar{u}^{(2)})' S^{-1} (\bar{u}^{(1)} - \bar{u}^{(2)}) \quad (5-4)$$

と与えられる（Anderson [1], p.109）。ここで、

計量経済モデルのパラメータ推定における異常値検出の方法（原田）

$$D^2 = (\bar{u}^{(1)} - \bar{u}^{(2)})' S^{-1} (\bar{u}^{(1)} - \bar{u}^{(2)})$$

は二標本平均間の Mahalanobis 汎距離であるから (Wilks [20], p.560),

$$T^2 = \frac{N_1 N_2}{N_1 + N_2} D^2 \quad (5-5)$$

となり, Hotelling T^2 検定は Mahalanobis D^2 検定になる。

ところで, T^2 の分布は

$$\{T^2 / (N_1 + N_2 - 2)\} \{(N_1 + N_2 - p - 1) / p\} \sim F_{p, N_1 + N_2 - p - 1}$$

であるから,

$$\left(\frac{N_1 + N_2 - p - 1}{p} \right) \left\{ \frac{N_1 N_2}{(N_1 + N_2 - 2)(N_1 + N_2)} \right\} D^2 \sim F_{p, N_1 + N_2 - p - 1} \quad (5-6)$$

を得る。(Rao [17], p.566)。

有意水準を α とすれば,

$$D^2 \geq \left(\frac{p}{N_1 + N_2 - p - 1} \right) \left\{ \frac{(N_1 + N_2 - 2)(N_1 + N_2)}{N_1 N_2} \right\} F(\alpha)_{p, N_1 + N_2 - p - 1} \quad (5-7)$$

の範囲に D^2 の値が入れば, 帰無仮説 $\mu^{(1)} = \mu^{(2)}$ は棄却される。

ここで, \tilde{Z} が $(p+1)$ 次元正規分布からの n 個の独立な標本から成っている
とすれば,

$$M(i) \geq \left(\frac{p+1}{n-p} \right) \left\{ \frac{(n-2)n}{n-1} \right\} F(\alpha)_{p+1, n-p} \quad (5-8)$$

計量経済モデルのパラメータ推定における異常値検出の一方法（原田）

の範囲に $M(i)$ もしくは $M(i, I_2^{(k)})$, $i=1\cdots n$, の値が入れば, 第 i 観測値 (x_i, y_i) は別の母集団に属するものと判定される。

ここにおいて, 異常値は別の母集団分布から抽出された観測値と解釈される。

6 む す び

経済変数間の関係を回帰モデルで表わし, そのパラメータを最小二乗推定する場合, 変数データの中の異常値によって推定結果に偏りが生じる。異常値を含んだデータから安定した経済関係（パラメータ）を推定することはできない。

本稿では, 変数データを直接診断して異常値を検出する方法を提示している。Cook 統計量による診断法では, それぞれの観測値が推定結果にどの程度影響しているかを測っているが, 本稿では, 直接, 観測値と他との差を Mahalanobis 汎距離によって測り異常値を検出している。

変数データの中から, 所定の母集団ではない別のものから抽出された異常値を検出し除去することは, 回帰モデルの誤差項の正規性の仮定を侵すことなく, パラメータの推定量の特性を良好に保ち, また検定結果を歪めることもなく, 推定結果を信頼あるものとなす。

計量経済モデルのパラメータ推定に採用される最小二乗推定法は, データの異常値に対して頑健な推定方式ではない。経済変数データから異常値を検出し, 除去しておくことは安定した経済関係を推定するために必要な措置である。

（1987年8月）

REFERENCES

- [1] Anderson, T. W. : *An Introduction to Multivariate Statistical Analysis*. New York : John Wiley & Sons, 1958.

- [2] Andrews, D. F., and D. Pregibon : "Finding the Outliers that Matter," *Journal of the Royal Statistical Society*, B 40 (1978), 85-93.
- [3] Belsley, D. A., E. Kuh and R. E. Welsch : *Regression Diagnostics ; Identifying Influential Data and Sources of Collinearity*. New York : John Wiley & Sons, 1980.
- [4] Carroll, R. J. : "Comment" on Minimax Aspect of Bounded-Influence Regression by P. J. Huber. *Journal of the American Statistical Association*, 78 (1983), 78-79.
- [5] Cook, R. D. : "Detection of Influential Observations in Linear Regression," *Technometrics*, 27 (1977), 15-18.
- [6] ——— : "Influential Observations in Linear Regression," *Journal of the American Statistical Association*, 74 (1979), 169-174.
- [7] ——— : "Assessment of Local Influence," *Journal of the Royal Statistical Society*, B 48 (1986), 133-169.
- [8] Cook, R. D., and S. Weisberg. : "Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression," *Technometrics*, 22 (1980), 495-508.
- [9] ——— : *Residuals and Influence in Regression*. New York, London : Chapman and Hall, 1982.
- [10] ——— : "Comment" on Minimax Aspect of Bounded-Influence Regression by P. J. Huber. *Journal of the American Statistical Association*, 78 (1983), 74-75.
- [11] Draper, N. R., and J. A. John : "Influential Observations and Outliers in Regression," *Technometrics*, 23 (1981), 21-26.
- [12] Gray, J. B., and R. F. Ring : "K-Clustering as a Detection Tool for Influential Subsets in Regression," *Technometrics*, 26 (1984), 305-330.
- [13] Hoaglin, D. C., and R. E. Welsch : "The Hat Matrix in Regression and ANOVA," *The American Statistician*, 32 (1978), 17-22.
- [14] Hocking, R. R. : "Developments in Linear Regression Methodology : 1959-1982," *Technometrics*, 25 (1983), 219-249.
- [15] Huber, P. J. : "Robust Regression : Asymptotic, Conjectures and Monte Carlo," *The Annals of Statistics*, 1 (1973), 799-821.
- [16] ——— : "Robustness and Design," *A Survey of Statistical Design and Linear Models* : J. N. Strisstra ed., Amsterdam : North Holland, 1975, 281-301.
- [17] Rao, C. R. : *Linear Statistical Inference and Its Applications*, 2nd ed.,

計量経済モデルのパラメータ推定における異常値検出の方法（原田）

New York : John Wiley & Sons, 1973.

- [18] Velleman, P. F., and R. E. Welsch : “Efficient Computing of Regression Diagnostics,” *The American Statistician*, 35 (1981), 234-242.
- [19] Weisberg, S. : *Applied Linear Regression*, 2nd ed., New York : John Wiley & Sons, 1985.
- [20] Wilks, S. S. : *Mathematical Statistics*, 2nd ed., New York : John Wiley & Sons, 1962.