Journal of the Operations Research
Society of Japan
Vol. 39, No. 4, December 1996

超楕円関数に基づくファジィロバスト回帰分析

藪内賢之　　和多田淳三
大阪工業大学

(受理 1994 年 12 月 21 日；再受理 1995 年 9 月 11 日)

和文概要　ファジィ回帰モデルでは，与えられた標本を包含するようにモデルを構成することでモデルの示すファジィ区間でシステムの可能性を表現している．このため，ファジィ回帰モデルでは，大きな誤差をもつデータが標本に混入しているとき，モデルが大きく歪む結果となる．モデルの形状を決定しているデータはデータ群の周辺部分に分布している．このため，本論文では，超楕円関数のパラメータを調整することにより周辺部分に存在するデータを抽出し，大きな誤差を含むデータについては距離概念で取り扱うことによって，大きな誤差を含むデータの影響を取り除いたファジィロバスト回帰モデルの構成方法を提案している．また，ファジィロバスト回帰モデルの応用としてアジア地域の経済状態と環境問題について分析し，その有効性を示している．

1. はじめに

L. A. Zadeh の拡張原理に基づいたファジィ多変量解析については，ファジィ線形回帰分析 [7]，ファジィ時系列モデル [12] や可能性線形モデル [6] などが既に提案されている．ファジィ線形回帰モデルでは，システムのもつ可能性を表現するモデル構築が目指されている．このため，ファジィ線形回帰モデルは，データを包含するようにモデルを構成することでファジィ区間によってデータのもつ可能性を表現している．このために可能性回帰モデルとも呼ばれる．

しかし，モデルはデータのもつ可能性のすべてを表現しようとするために，言い換えるとすべてのデータをファジィ区間に包含するように構成するために，データが著しく大きな誤差を含んでいる場合には，それらのデータをもシステムの可能性として処理することになる．結果として，現実のシステムよりもモデルのもつ可能性が広がりすぎたり，誤差をもったデータのために著しくモデルが歪むという弊害が生じる．

このため，誤差を含むデータ (以下，誤差データと呼ぶ) の影響を除くための方法論が求められてきた．石渕らは，可能性回帰モデルの幅を規定するに際して影響の大きいデータの幾つかを除くことで誤差の影響を排除する方法を議論しており，混合 0-1 整数計画問題による区間回帰分析 [1] を提案している．可能性回帰モデルの幅を規定する際に，データ群中心付近に特異データが存在してもモデルの幅に及ぼす影響はほとんどないが，特異データがモデルの上限や下限付近に分布している場合には特異データの影響によりモデルの幅が大きくなることがある．

このため，石渕モデルでは直接モデルに影響を与えているモデルの上限や下限付近に分布しているデータ (周辺データと呼ぶ) を取り除く方が特異データによる影響を効率的に取り除くことができるという事実に立脚した近似解法によって周辺データを除去し，誤差の影響

を排除した可能性回帰モデルを求めている.

文献[11]では,データに含まれる誤差が可能性回帰モデルの構成に及ぼす影響を最小にするために距離概念をファジィ回帰モデルに導入することにより,ファジィロバスト回帰モデルを構成する方法を提案している.本研究では,可能性概念で扱うべきデータと距離概念で扱うべきデータとを分離するための手段として超楕円体を用いたモデルの構築方法を提案する.

2. ファジィ回帰モデル

求めるべきファジィ回帰式は,次式で示される.

$$Y_j = A_1 x_{1j} + A_2 x_{2j} + \cdots + A_n x_{nj} = \mathbf{A}\mathbf{x}_j \qquad x_{1j} = 1 \tag{2.1}$$

ここで,回帰係数A_iは中心a_i,幅c_iとする三角型ファジィ数$A_i = (a_i, c_i)$であり,$\mathbf{A} = (\mathbf{a}, \mathbf{c})$とする.観測データ$(y_j, \mathbf{x}_j), \mathbf{x}_j = [x_{1j}, x_{2j}, \cdots, x_{nj}]'$におけるファジィ回帰式は,拡張原理より$\mathbf{A}\mathbf{x}_j = (\mathbf{a}, \mathbf{c})\mathbf{x}_j = (\mathbf{a}\mathbf{x}_j, \mathbf{c}|\mathbf{x}_j|)$であるため,ファジィ係数をもつファジィ回帰式(2.1)の出力もまたファジィ出力となることがわかる.ここで$|\mathbf{x}_j| = [|x_{1j}|, |x_{2j}|, \cdots, |x_{nj}|]'$である.この演算により,(2.1)式は,

$$Y_j = \mathbf{A}\mathbf{x}_j = (\mathbf{a}\mathbf{x}_j, \mathbf{c}|\mathbf{x}_j|) \tag{2.2}$$

と表現できる.ファジィ係数をもつファジィ回帰モデルは下限が$\mathbf{a}\mathbf{x}_j - \mathbf{c}|\mathbf{x}_j|$,中心が$\mathbf{a}\mathbf{x}_j$,上限が$\mathbf{a}\mathbf{x}_j + \mathbf{c}|\mathbf{x}_j|$である.可能性回帰モデルはデータを包含するように構成されている.しかし,データを包含するように構成したモデルでは,モデルの幅が大きければ大きいほど回帰式の意味する変数間の関係がぼやけたものとなる.結果として回帰式の表現は不明確なものとなる.モデルの曖昧さを少なくするようにモデルの幅を最小にすることによって回帰モデルを明確にする.ファジィ回帰モデルは次のLP問題に帰着させることができる.

$$\begin{aligned} &\min_{\mathbf{a},\mathbf{c}} \sum_{j=1}^{m} \mathbf{c}|\mathbf{x}_j| \\ &subject\ to \\ &\quad y_j \leq \mathbf{a}\mathbf{x}_j + \mathbf{c}|\mathbf{x}_j| \\ &\quad y_j \geq \mathbf{a}\mathbf{x}_j - \mathbf{c}|\mathbf{x}_j| \\ &\quad \mathbf{c} \geq \mathbf{0} \quad (j = 1, 2, \cdots, m) \end{aligned} \tag{2.3}$$

このLP問題を解くことによって得られる解は,例えば図1に示す回帰モデルになる.

3. ファジィロバスト回帰モデル

現実の世界では観測データの中に,観測方法の誤り,観測値の読み間違い,あるいは観測装置の誤動作などによって観測値に誤差が含まれることがある.このようなデータは,システム全体のもつ可能性を歪めて表現させる原因になる.誤差データがシステム構築に影響を与えることは好ましくない.特異データのような,システムがもつべき可能性とは異なった要因がモデル構築に影響する危険性を取り除くことは,システムを正しく表現するためには必要である.本研究ではそのような誤差データの影響を最小にする方法を提案する.

本研究では距離概念を導入して特異データの混入によるモデルの歪みを排除した可能性回帰モデルを構成することを検討する.

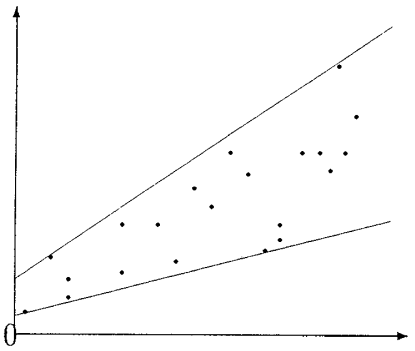


図 1　可能性回帰モデル

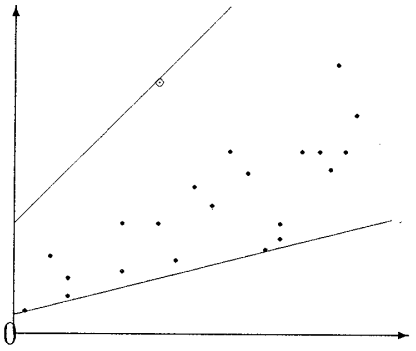


図 2　特異点⊙を含む可能性回帰モデル

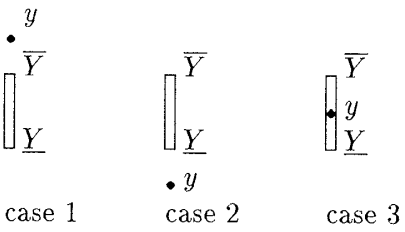


図 3　モデルとデータの関係

システムが本来もつ可能性を表現している可能性回帰モデルは図 1 に示す通りであるとする．この図 1 に特異データ⊙が混入したとき，モデル (2.1) は，図 2 に示すように，システムがもつ本来の可能性から見ると大きく歪んだ形になる．この特異データの存在が大きく回帰モデルを不明確なものにする．さらに誤差を含んだデータについては誤差と可能性を切り離すことが望まれる．このため距離概念を導入し，データとモデルとの誤差とモデルのあいまいさとを別々に評価するのがよいであろう．つまり，モデルとデータとの距離を最小にすることをも考慮する．モデルの幅を最小にし，モデルと特異データとの距離を最小にすることによってモデルの上限および下限を決定するのである．

特異データとモデルとの位置関係は，図 3 に示すようになる．

モデルとデータの関係は，図 3 より，

case 1：特異データがモデルよりも上側に位置する，

case 2：特異データがモデルよりも下側に位置する，

case 3：データがモデルの内側に存在する，

の 3 つの場合に分けることができる．特に，case 3 についてはそのデータを特異データとはみなさない．このため通常のデータとして取り扱われる．一方，case 1 については，誤差データとシステムとの距離はシステムの上限までと考えることが適当であろう．case 2 についても case 1 と同様に，誤差データとシステムとの距離はシステムの下限までとなる．この結果，データ D_j と可能性モデルとの誤差，つまり，可能性モデルと誤差データとの距離を r_j とおくと次のように表現できる．

$$r_j = \begin{cases} y_j - \overline{Y_j} & ; \quad \overline{Y_j} < y_j \\ 0 & ; \quad \underline{Y_j} \le y_j \le \overline{Y_j} \\ \underline{Y_j} - y_j & ; \quad y_j < \underline{Y_j} \end{cases} \tag{3.1}$$

ここで，$\overline{Y_j}$ はモデルの上限，$\underline{Y_j}$ はモデルの下限を示している．距離関数 (3.1) を用いて可能

性回帰モデルの評価関数を次のように定義する．

$$J = \sum_{j=1}^{m} |r_j| + K \sum_{i=1}^{n} c_i \tag{3.2}$$

評価関数 (3.2) の Kは，正の実数でモデルの幅に対するウェイトであり，誤差データのもつ可能性をどの程度考慮するかを決めるパラメータである．

ここで，通常の可能性回帰モデルの評価関数と同様に，評価関数はただ一つの形で表現されるものではなく，幾通りもの表現方法が存在しうることに注意すべきである．例えば，

$$\sum_{j=1}^{m} \mathbf{c}|\mathbf{x}_j| \ , \ \sum_{i=1}^{n} c_i \ , \ \sum_{i=1}^{n} c_i^2 \ , \cdots \tag{3.3}$$

評価関数の形は問題に応じて適切なものが採用されるべきである．Kを小さな値にすれば，モデルの幅よりもデータのモデルに対する誤差を最小にすることに力点が置かれているので通常のファジィ回帰モデルと同様になる．それとは逆に Kの値を大きくすれば，データのモデルに対する誤差よりもモデルの幅を最小にすることに力点が置かれているのでファジィロバスト回帰モデルとなる．このパラメータ Kによって，意思決定者あるいは分析を行う者のもつ経験的知見を反映させることができる．

得られたデータ全部のうち，どれが正常なデータであり，どれが誤差データであるかは一見，見分けがつかない．本論文で提案するロバスト性をもつ可能性回帰モデルでは，どのデータが特異データであり，モデルの外側に位置させるかを距離概念をもとに決定することを提案している．特異データの影響を取り除き，システムが本来もつ可能性を示す可能性回帰モデルを誤差データの可能な組合せの中から選び出す方法を提案しており [11]，本モデルは組合わせ最適型可能性回帰モデルと捉えることができる．本モデルでは，組合わせ最適問題を解くために単純遺伝的アルゴリズムを用いている．

4. 超楕円体によるデータの選別

先に述べたファジィロバスト回帰モデルを得るには，データを可能性概念，あるいは距離概念のうちどの概念で扱うべきかを決める必要がある．このため，m 個のデータについて 2 つの概念のうちどの概念で扱うか判断するために 2^mだけの計算を必要とする．例えば，$m = 10$ であれば 2^{10}，$m = 20$ なら 2^{20}の計算を要する．しかしながら，計算で得られたモデルのうち実現可能なモデルは数割にすぎない．

データを可能性概念で扱うファジィ回帰モデルはデータの可能性を表現するためにデータ全てを包含する．言い換えれば，ファジィ回帰モデルは周辺データによってモデルの形状が決まり，データ群中心付近に分布するデータはファジィ回帰モデルの構築には影響していないことがわかる．このことから，ファジィロバスト回帰モデルにおいても，モデル構築に大きな影響を与えている周辺データを特異データとしてモデルを構築すると効率よく求めることができることは明らかである．このため，本論文では周辺データを選別する方法として超楕円体を用いることを提案する．

4.1. 超楕円関数

データ数 m の p 変量データ $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m]$ とその第 i データ $\mathbf{u}_i = [u_{1i}, u_{2i}, \cdots, u_{pi}]'$ を考える．$\mathbf{U}$ の平均は$\bar{\mathbf{U}} = [\bar{\mathbf{u}}, \bar{\mathbf{u}}, \cdots, \bar{\mathbf{u}}]$，$\bar{\mathbf{u}} = [\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_p]'$である．

分散共分散行列

$$\Sigma = \frac{1}{m}(\mathbf{U}-\bar{\mathbf{U}})(\mathbf{U}-\bar{\mathbf{U}})' = [\sigma_{ij}] \quad (i,j=1,\cdots,p) \tag{4.1}$$

を用いると標準化した距離の自乗は,

$$\sum_{i=1}^{p}\sum_{j=1}^{p}\frac{(x_i-\bar{u}_i)(x_j-\bar{u}_j)}{\sigma_{ij}} \tag{4.2}$$

と表現できる．このマハラノビスの距離 (4.2) から超楕円関数が求められるが，データが各軸方向に分布していることはほとんどなく，超楕円体をデータ分布にあった形に修正する必要がある．一般に，データ分布に合うように変換すると (4.2) 式は分散共分散行列Σを用いることにより,

$$(\mathbf{x}-\bar{\mathbf{u}})'\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{u}}) = \sum_{i=1}^{p}\sum_{j=1}^{p}\frac{(x_i-\bar{u}_i)(x_j-\bar{u}_j)}{\sigma_{ij}} \tag{4.3}$$

と一般化できる．

分散共分散行列Σの固有値λ_iとそれに対応する固有ベクトル$\mathbf{e}_i = [e_{1i}, e_{2i}, \cdots, e_{pi}]'$を用いると

$$\Sigma = \sum_{i=1}^{p}\lambda_i \mathbf{e}_i \mathbf{e}_i' \tag{4.4}$$

とスペクトル分解できる．(4.4) 式は,

$$\Sigma^{-1} = \sum_{i=1}^{p}\frac{1}{\lambda_i}\mathbf{e}_i \mathbf{e}_i' \tag{4.5}$$

と再表現できることから超楕円関数は，固有値問題により求めることができる．

以上のことから超楕円関数は,

$$\begin{aligned}(\mathbf{x}-\bar{\mathbf{u}})'\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{u}}) &= (\mathbf{x}-\bar{\mathbf{u}})'\sum_{i=1}^{p}\frac{1}{\lambda_i}\mathbf{e}_i\mathbf{e}_i'(\mathbf{x}-\bar{\mathbf{u}}) \\ &= \sum_{i=1}^{p}\frac{1}{\lambda_i}\{(\mathbf{x}-\bar{\mathbf{u}})'\mathbf{e}_i\}^2 \le l^2\end{aligned} \tag{4.6}$$

となる．この超楕円関数を用いることにより周辺データを識別し，効率的な計算ができる．

ここで，超楕円関数のパラメータl^2は抽出する特異データの候補数hによって定められる．本論文においては，パラメータl^2は以下のようにして決定している．

1. 特異データの候補数hを決める．
2. 最初は全データが超楕円関数の内部に入るようにパラメータを定める．
3. 予め定めておいた特異データの候補数だけのデータが超楕円体の外部に出るようにパラメータl^2を調整する．

上記において特異データの候補数hを決定する際には経験的に，あるいはヒューリスティックに決める必要がある．特異データの候補数が少ない場合には歪んだモデルになる可能性が大きく，候補数が多い場合は計算回数が多くなるが最適解が得られやすくなる．以上の理由から，特異データの候補数hには注意する必要がある．

表1　世帯数と人口

No.	1	2	3	4	5	6	7	8	9	10	11
X	39.9	22.4	27.2	25.8	37.1	30.0	23.3	25.6	36.6	80.7	32.3
Y	96.9	54.7	68.8	58.4	90.2	79.6	55.1	100.0	92.4	182.3	79.6
No.	12	13	14	15	16	17	18	19	20	21	22
X	57.1	43.1	62.8	41.5	68.1	57.6	72.6	69.6	74.3	37.9	24.8
Y	151.3	103.4	154.0	101.8	161.2	142.3	139.6	140.2	198.2	85.2	53.8

表2　超楕円体による計算効率の比較

	総組合せ数	必要計算回数
超楕円体を用いない場合	2^{22}	$2^{22} = 4194304$
提案モデル (本数値例)	2^{22}	$2^h = 2^5 = 32$

ただし，h は超楕円体で抽出した特異データの候補の数.

4.2. 数値例

数値例として表1のデータを用いる．本数値例で用いるデータは性質上，大きな誤差が混入していることは考えにくい．しかし，数個の特徴的なデータが影響するとファジィ回帰モデルでは歪んだ傾向を示すことになる．提案モデルにおいて数個の特徴的データの影響を少なくすることによって分析対象の傾向を把握できることを示す.

ここで，説明変量 X を世帯数，目的変量 Y を人口とし，モデルは以下の単回帰モデルとする.

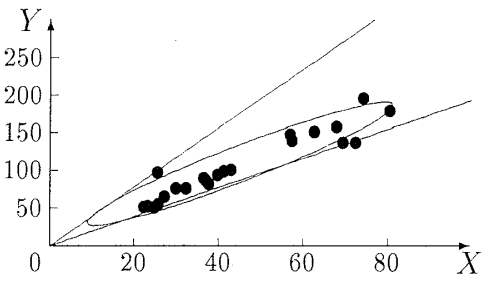
$$Y = A_1 X$$

表1に示すデータにより得られた超楕円体は,

$$0.006(Y - 108.591)^2 + 0.029(X - 45.014)^2 \\ -0.025(Y - 108.591)(X - 45.014) = 3.667$$

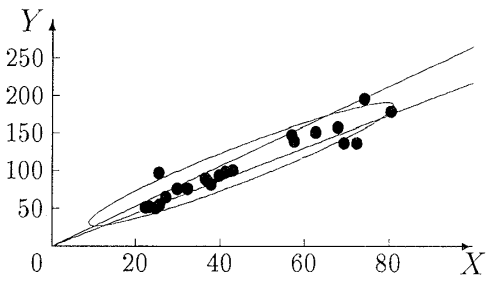
超楕円体は，特異データの候補を5とし，5つのデータが超楕円体の外側に位置するようにして求めた．従って，データ数22の本数値例では 2^{22} もの計算を必要とするが，本数値例では超楕円体を用いることにより5つのデータについて議論するだけでよく，2^5 の計算で解が得られることになる.

ここで，超楕円体を用いる効用を表2に示す．表2では，超楕円体を用いずに最適解を求める場合，および超楕円体を用いて最適解を求めた場合の計算回数を示している．また，提案モデルは本数値例での計算回数である．本数値例では，特異データの候補を5として超楕円体の形状を定めているため，特異データの候補に対して距離概念で扱うべきか，可能性概念で扱うべきかを決めるための計算は32回である．これに対し，超楕円体を用いない場合には全てのデータに対してどの概念で扱うか決める必要がある．本数値例ではデータ数が22であることから $2^{22} = 4194304$ 回の計算が必要である．超楕円体を用いた場合，用いない場合の 32/4194304 だけの計算が必要となる．超楕円体を用いない場合には，データ群中央付近のデータをも可能性概念，あるいは距離概念で扱うべきか分析しているために計算回数が大きくなる．超楕円体等を用い合理的な分析を行うことにより計算回数が飛躍的に減少する.



$Y = (2.915, 0.992)X$

図 4　K=1 の時の最適解



$Y = (2.409546, 0.240191)X$

図 5　K=500 の時の最適解

モデルのパラメータ Kはヒューリスティックに決めるため，幾通りかの値を用いて分析を行い，適当な解を得る Kを採用する．本数値例では，K=1 と K=500 とした場合の解を示す．

上式の超楕円体を用いることにより得られた回帰モデルを図 4(K=1)，図 5(K=500) に示す．

K=1 としたときの最適解は，全データを可能性概念で扱い，通常の可能性回帰モデルになっていることが理解できる．K=1 では，モデルのあいまいさが大きく，システムの特徴を把握しにくい．

K=500 としたときの最適解は，可能性概念で扱うデータと距離概念で扱うデータとに分離できており，ファジィロバスト回帰モデルになっている．K=1 とし，全データを可能性概念で扱う場合と比べて，特異データによる影響が小さくなっており，モデルにロバスト性が得られていることが理解できる．

本数値例では，データに大きな誤差が混入しているわけではないが，システムのあいまいさを最小にするために本モデルが適用できることを示した．

5. 経済パラメータの環境要因への影響

現在アジア地域では，経済活動が活発に行われており，エネルギー消費に伴う環境変化が問題となっている．

アジア地域の経済状態と環境問題の分析に本モデルを用いる．経済状態に関しては，人口，国内総生産 (GDP)，1 次エネルギー消費量，環境に関しては NO_x，SO_x，CO_2の観測量を用いる．

1 次エネルギーは，石炭，石油，ガス，1 次電力の商用エネルギーであるが，発展途上国は非商用エネルギーである植物性燃料に依存していることが多いため，植物性燃料も含めて 1 次エネルギーとして取り扱った．

NO_xと SO_xは地球の酸性化に関わる物質であり，NO_xは主成分をオゾンとする光化学オキシダントの生成に関係しており，CO_2は地球温暖化に関わる物質である．このため，環境を分析するために NO_x，SO_x，CO_2をそれぞれ目的変数として分析を行った．

本応用例で用いるデータは性質上，大きな誤差が混入していることは考えにくい．しかし，数個の特徴的なデータが影響するとファジィ回帰モデルでは歪んだ傾向を示すことになる．提案モデルにおいて数個の特徴的データの影響を少なくすることによって分析対象の傾向を把握できることを示す．

自然対数をとった 1975 年のデータを用いる (表 3)．説明変数は人口を X_1，GDP を X_2，

表3　分析データ

国名	人口	GDP	1次エネルギー消費量	NO_x	SO_x	CO_2
1	6.84	5.22	5.87	8.22	9.23	12.67
2	4.71	6.72	5.79	7.75	7.85	12.51
3	6.40	5.00	4.98	7.23	7.41	11.16
4	4.87	3.99	3.63	5.80	5.30	9.33
5	3.56	3.76	3.33	5.39	7.06	10.04
⋮	⋮	⋮	⋮	⋮	⋮	⋮
19	0.35	-0.19	0.43	3.43	3.66	7.02
20	-1.83	1.10	-1.56	0.69	-0.92	4.90

人口：百万人, GDP：'80P10億US$
1次エネルギー消費量：Mtoe, NO_x排出量：1,000 ton
SO_x排出量：1,000 ton, CO_2排出量：1,000 T-C
(注) toe：ton oil equivalent

1次エネルギー消費量をX_3とし，NO_xの推定値をY_{NO_x}，SO_xの推定値をY_{SO_x}，CO_2の推定値をY_{CO_2}とする．

本応用例では，提案モデルによる対象システムの可能性を解析することを目的としている．このため，本来であれば共線性等を考慮する必要があるが本論文では厳密に分析を行わない[2, 9]．また，分析に用いるモデルには以下に示すファジィ重回帰モデルを用いる．

$$Y = A_0 + A_1X_1 + A_2X_2 + A_3X_3$$

超楕円体により8つの誤差データを識別した後，分析を行った．NO_xモデルの超楕円体は

$$23.389(Y_{NO_x} - 4.589)^2 + 7.591(Y_{NO_x} - 4.589)(X_1 - 3.171)$$
$$+ \cdots + 44.296(X_3 - 2.362)^2 = 3.545$$

SO_xモデルの超楕円体は

$$1.322(Y_{SO_x} - 4.712)^2 + 1.322(Y_{SO_x} - 4.712)(X_1 - 3.171)$$
$$+ \cdots + 10.193(X_3 - 2.362)^2 = 3.226$$

CO_2モデルの超楕円体は

$$3.183(Y_{CO_2} - 8.451)^2 + 2.969(Y_{CO_2} - 8.451)(X_1 - 3.171)$$
$$+ \cdots + 12.234(X_3 - 2.362)^2 = 3.203$$

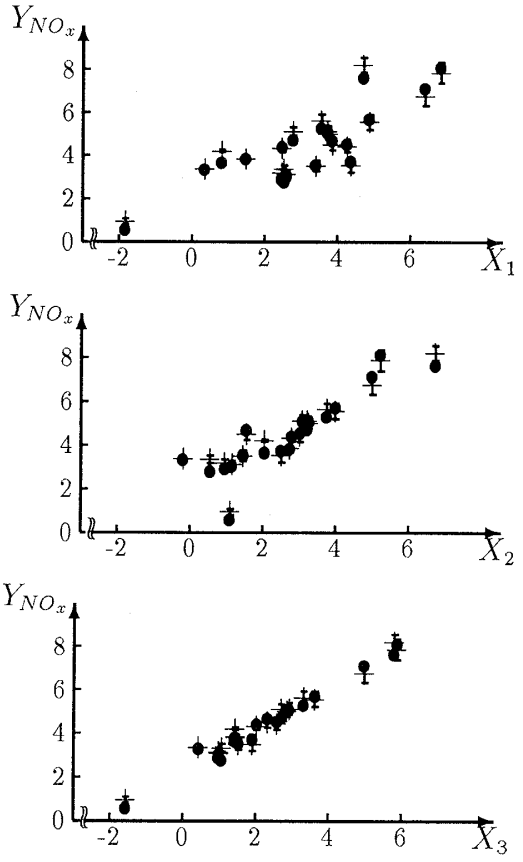
誤差データとモデルの間には2^8の組合せが存在し，組合せの中から最適なモデルを探索するために遺伝的アルゴリズムを用いた．集団サイズを200，交叉率を70%，突然変異率を1%，淘汰率を90%として試行を行った．表4の結果では，$SO_x(K=1)$と$CO_2(K=1)$の最適解到達率がそれぞれ36%，46%と低くなっている．この時の収束状況を表5に示す．また，遺伝的アルゴリズムは確率探索の手法であるため，同条件で50回の試行を行った．$SO_x(K=1)$と$CO_2(K=1)$は最適解到達率が他と比べると低いが準最適解，あるいは準々最適解では他と同様に高い確率で収束している．SO_x，CO_2共に最適解，準最適解，準々最適解を比較すると評価値の差は非常に小さく，これらの解の間には相違が小さいことが分か

表4　遺伝的アルゴリズムによる探索結果

	K	平均終端世代	最適解の平均	平均発生個体数	最適解到達率
NO_x	1	18	0.228	838	92
	100	18	8.298	854	100
SO_x	1	19	0.717	928	36
	100	25	26.934	903	94
CO_2	1	22	0.739	955	46
	100	19	12.581	954	90

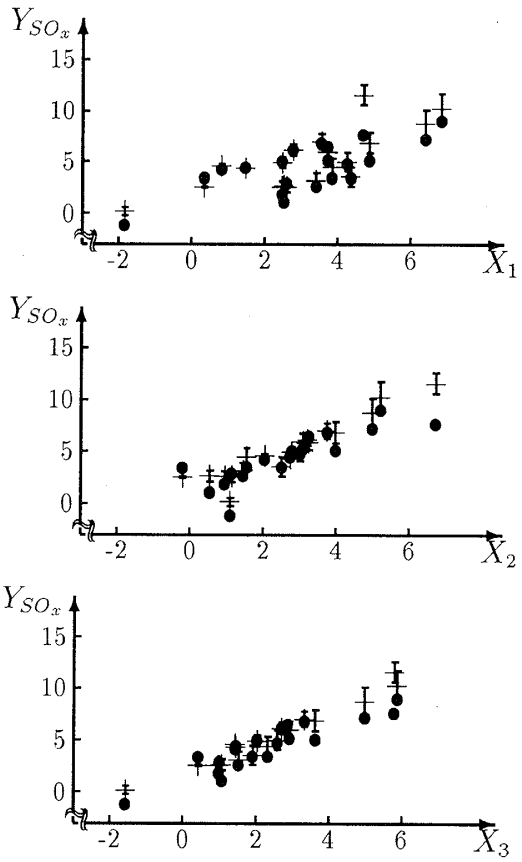
表5　SO_x, CO_2(K=1) について

解の順位	SO_x		CO_2	
	評価値	到達率	評価値	到達率
1	0.717176	36	0.738750	46
2	0.717177	56	0.738752	90
3	0.717178	98	0.738753	100



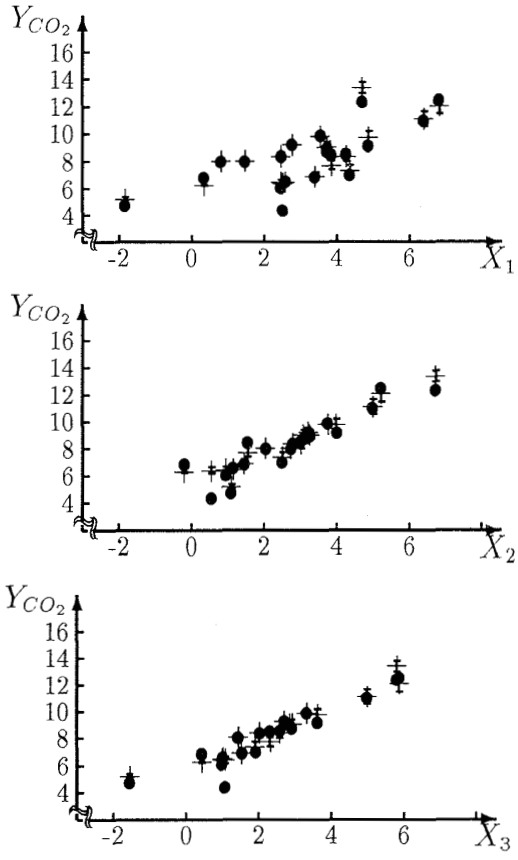
$$Y_{NO_x} = (2.830, 0) + (-0.363, 0.071)X_1 + (-0.214, 0)X_2 + (1.475, 0)X_3$$

図6　NO_xの最適モデル ($K = 100$)



$$Y_{SO_x} = (1.979, 0) + (-0.578, 0.209)X_1 + (0.164, 0.000)X_2 + (1.936, 0)X_3$$

図7　SO_xの最適モデル ($K = 100$)



$$Y_{CO_2} = (5.960, 0) + (-0.403, 0.087)X_1 + (0.349, 0)X_2 + (1.202, 0)X_3$$

図 8　CO_2の最適モデル $(K = 100)$

る．このことから，$SO_x(K = 1)$ と $CO_2(K = 1)$ の最適解到達率は低いが，高い確率で最適解と同様のモデルを探索しており，最適解到達率が低いことによる問題はないであろう．最適解を図 6~8 に示す．データは 4 次元空間上に存在しているが，説明のため各説明変数と目的変数による 2 次元平面で表現している．モデルは 4 次元平面により可能性の幅を表現しており，モデル及びデータは共に線形性を示している．

各図において“•”は実測値を示している．区間は可能性の幅を表現し，その上限は可能性の上限，中心は可能性の中心，下限は可能性の下限に対応している．

$$Y_{NO_x} = (2.717, 0.107) + (-0.382, 0.082)X_1 + (-0.290, 0.040)X_2 + (1.543, 0)X_3 \tag{5.1}$$

NO_xモデルは，$K = 1$ とした (5.1) 式においてモデルの幅が小さくなっている．X_1及び X_2 の可能性の幅は，それぞれ 0.082，0.040 と小さい．

$K = 100$(図 6) の場合においてもモデルの形状が大きく変化することもなかった．$K = 1$ と $K = 100$ では係数の幅が多少異なるが，モデルの可能性の幅が小さくならなかった．

従って，NO_xモデルは他国と傾向が大きく異なるデータが混入していないことがわかる．

$$Y_{SO_x} = (2.537, 0) + (-1.470, 0.269)X_1 + (-0.583, 0.448)X_2 + (3.530, 0)X_3 \tag{5.2}$$

表6　アジア全体と比べて傾向の違う国々

国名	特徴
2	公害対策として排煙処理をしている唯一の国である．NO_x，SO_x，CO_2共にアジア全体と比較して観測量は下限よりも少なくなっている．
4	農業の就業者数は労働人口の半数でありGDPの1/4を占める．原油，天然ガス，石油製品が総輸出の3/4である．観測量はアジア全体と比べるとSO_xが比較的少ない．
6	電子部品等の工業製品を製造する工業経済である．観測量はアジア全体と比べてNO_xが少ない．
12	データ収集時に戦争が終結し，農業も自給できる状態ではない．化石燃料の使用量が多く，CO_2の観測量がアジア全体より比較的多い．
14	輸入原油の精製が主な産業である．観測量はアジア全体と比べてNO_xが低く，CO_2が多い．
15	農業国であり，労働力の90%が農業に従事している．観測量はアジア全体と比べてNO_x，SO_x，CO_2共に少ない．
19	放牧による農業経済である．工業の発展がめざましく，観測量はアジア全体と比べてSO_xが少なく，NO_x，CO_2が多くなっている．
20	原油，天然ガスが輸出の全てである．観測量はアジア全体と比べてNO_x，CO_xは少ない．

SO_xモデルは，1次エネルギー消費量(X_3)は$K=1$とする(5.2)式，$K=100$(図7)共に可能性の幅がなく，X_3以外は負の係数を示している．しかし，$K=100$においてX_2は正の係数をもち，可能性の幅は0である．このことから，SO_xモデルでは，X_2にアジアの傾向とは異なるデータが存在していることにより，X_2の可能性の幅が大きく歪んでいたことがわかる．

$$\begin{aligned} Y_{CO_2}=&(6.042,0.394)+(-1.101,0.264)X_1 \\ &+(0.196,0.081)X_2+(2.162,0)X_3 \end{aligned} \tag{5.3}$$

CO_2モデルは，$K=1$とする(5.3)式と$K=100$(図8)ではモデルの係数が大きく変わっている．X_1及びX_2の可能性の幅は，$K=100$とすることによって小さくなっている．CO_2モデルの可能性の幅が大きく，あるいはアジアとしての傾向と異なるデータを距離概念で扱うことによりアジアの傾向が得られた．

分析の結果，アジア全体の傾向とは異なる国は8ヶ国(表6)であった．

以下，$K=100$とした場合において，特徴的なものだけ考察する．NO_xでは，国2, 6, 14, 15, 20がアジア全体と比べると少なく，国19が多くなっている．SO_xでは，国2, 4, 15, 19がアジア全体と比べて少なくなっている．CO_2では，国2, 15, 20がアジア全体と比べて少なく，国12, 14 19が多くなっている．国2は公害対策を行っており，国15は農業国であるため公害物質(NO_x，SO_x，CO_2)を生成するような産業が活発ではないためにこれらの観測量が少なくなっている．国19は家畜を規範とする農業経済，工業の発展段階の途中であることから以上のような結果となった．国14は輸入原油の精製が主な産業である．このため，NO_xの観測量がアジア全体と比較して多くなっている．

また，アジアの傾向としては，GDP及び1次エネルギー消費量が大きな国ではCO_2の観測量が多く，人口が多い国では観測量が少なくなっている．NO_x，SO_x，CO_2共に1次エネ

ルギー消費量の係数は他の変数と比べると大きなウェイトをもっており，1次エネルギー消費量に観測量が左右されている．人口に関してはいずれも人口が大きければ観測量が少ない傾向にある．

また，モデルの可能性の幅が小さく見えても対数をとったデータを用いており，実際には可能性の幅は更に大きくなる．

本適応例では，アジア各国の環境状態を把握するために人口，GDP，1次エネルギー消費量を用いて分析を行った．

アジア全体の可能性をそのまま表現するモデルはファジィ回帰モデル ($K = 1$) によって得られるが，ファジィロバスト回帰モデル ($K = 100$) ではアジア全体を把握するための情報が得られることを示した．

つまり，ファジィロバスト回帰モデルでは全体と比べて説明変数と目的変数間の関係が異なる国を距離概念で扱うことにより，アジア全体としての傾向を把握することができた．

6. おわりに

本論文では，次のことを明らかにした．

(1) 超楕円体を用いて，誤差データを絞り込むことにより効率的にモデルを求める方法を示した．

(2) 数値例では簡単なデータを用いて本モデルの特徴を示した．

(3) ファジィロバスト回帰モデル [11] の構築を検討し，超楕円体を用いたファジィロバスト回帰モデルの解法を用いてアジア地域の環境要因への経済パラメータの影響を検討した．

参考文献

[1] 石渕 久生, 田中 英夫：混合 0-1 整数計画問題による区間回帰分析, 日本経営工学会誌, Vol.40, No.5 (1988), 312–319

[2] 岩田 暁一：経済分析のための統計的方法, 東洋経済新報社, 1983

[3] 科学技術庁科学技術政策研究所編：アジアのエネルギー利用と地球環境, 科学技術庁科学技術政策研究所編, 1992

[4] 小林 重信：遺伝的アルゴリズムの基礎と応用 [I], [II], オペレーションズ・リサーチ学会誌, Vol.38, No.5 (1993), 256–261, Vol.38, No.6 (1993), 311–318

[5] D. E. Goldberg : *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, 1989

[6] 坂和 正敏：ファジィ理論の基礎と応用, 森北出版, 1989

[7] H. Tanaka and J. Watada : Possibilistic Linear Systems and Their Application to The Linear Regression Model, *Fuzzy Sets and Systems*, Vol.27 (1988), 275–289

[8] 田中 英夫：ファジィモデリングとその応用, 朝倉書店, 1990

[9] 早川 毅：回帰分析の基礎, 朝倉書店, 1986

[10] Peter J. Roussecuw and Annick M. Leroy : *Robust Regression and Outlier Detection*, John Wiley, 1987, 1–18

[11] 藪内 賢之, 和多田 淳三, 辰巳 憲一：誤差データのファジィ回帰分析, 日本ファジィ学会誌, Vol.6, No.6 (1994), 1161–1170

[12] J. Watada : Fuzzy Time-series Analysis and Its Forecasting of Sales Volume, *Fuzzy Regression Analysis*, edited by J. Kacprzyk & M. Fedrizzi, 1992, 211–227

[13] J. Watada and Y. Yabuuchi : Fuzzy Robust Regression Analysis, *FUZZ/IEEE'94: Third IEEE international Conference on Fuzzy Systems*, in Florida, U.S.A., June 26 – July 2, 1994, 1370–1376

和多田 淳三
大阪工業大学 経営工学科
〒 535 大阪府大阪市旭区大宮 5–16–1
TEL : 06–954–4327
FAX : 06–952–6197
E-mail : watada@dim.oit.ac.jp

ABSTRACT

FUZZY ROBUST REGRESSION ANALYSIS BASED ON A HYPERELLIPTIC FUNCTION

Yoshiyuki Yabuuchi Junzo Watada
Osaka Institute of Technology

A fuzzy regression model illustrates the possibility of a considered system so that all given data are included in fuzzy intervals obtained by the model. So the fuzzy regression model is easily warped and bended by data with large error, when the error data are mixed in the possibilistic data. Especially the marginal data have the influence on configurating he shape of the model. Hyperelliptic function is employed to select marginal data by means of recognizing data positions by adjusting its parameters. The fuzzy robust regression model is proposed in this paper so that the data with large error can be removed out of the marginal data using distance.

As an application of the fuzzy robust regression model, the relation between economy in Asian region and environmental problem is analyzed. This application shows the meaning and effectiveness of the fuzzy robust regression model.