【土木計画学研究・講演集 No.20(1) 1997年11月】

# 線形回帰モデルにおける多重共線性とその改善手法に関する逆解析的考察*

*Ill-Posed Inverse Problem and Regularization Methods in Linear Regression*
*— Multicollinearity and Its Countermeasures —*

堤 盛人**・清水 英範***・松葉 保孝****
by Morito TSUTSUMI**, Eihan SHIMIZU***, Yasutaka MATSUBA****

## 1. はじめに

ここ数年、工学や理学の分野を中心として逆問題 *inverse problem* あるいは逆解析 *back analysis* という言葉がよく使われるようになってきている（例えば、久保[1]）。計量モデルを用いて人口や地価などを予測する問題を「順問題」と呼べば、人口や地価などの地域の現況を再現するモデルを推定する問題、若しくはモデルが定式化された上でのパラメータを推定する問題は、一種の逆問題である。

ある問題の解に対して、① 解の存在性 *existence* ② 解の一意性 *uniqueness* ③ 解の連続性 *continuity* あるいは安定性 *stability* の３つの要件がすべて満足されているとき、その問題は適切 *well-posed* であるという[2]。逆問題ではこれらの適切性に関する要件を満足しない不適切 *ill-posed* な場面に直面することが多い。

プロジェクトの計画段階において、分析モデルとして用いられることの多い線形回帰モデルにおけるパラメータ推定では、③の解の安定性が失われる多重共線性 multicollinearity に悩まされることが少なくない。多重共線性が存在する場合におけるパラメータ推定の改善方法に関しては、統計学・計量経済学を始めとして多くの学問分野で研究がなされ、様々な手法が提案されている。本稿ではそれらの既存研究の相互関係を逆問題という視点からまとめて示す。

## 2. 多重共線性

本稿では、つぎのような線形回帰モデルについて考える。なお $t$ はベクトル・行列の転置を表す。

$$y=\mathbf{x}\boldsymbol{\beta} \tag{1}$$

$\mathbf{x}=(X_1,X_2,\cdots,X_k),\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_k)^t$

$y$：内生変数　$\mathbf{x}$：外生変数ベクトル

$\boldsymbol{\beta}$：パラメータベクトル

### (1) 最小二乗法によるパラメータの推定

各変数に関し $m$ 個のデータセットが得られたものとし、式(1)を次のように書き換える（$i=1,\cdots,m>k$）。

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u} \tag{2}$$

$\mathbf{y}=(y_1,y_2,\cdots,y_i,\cdots,y_m)^t$

$$\mathbf{X}=\begin{pmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1m} & X_{2m} & \cdots & X_{km} \end{pmatrix}$$

$\mathbf{u}=(u_1,u_2,\cdots,u_i,\cdots,u_m)^t$　：残差ベクトル

ここで残差は次式を満たす確率変数と仮定する。

$$E(\mathbf{u})=\mathbf{0},\quad Var(\mathbf{u})=\sigma^2\mathbf{I} \tag{3}$$

以下の記述を容易にするため、外生変数はそれぞれ平均 0、分散 1 に規準化されているものとする。

通常の最小二乗法（OLS）によるパラメータ推定は、

$$\min_{\boldsymbol{\beta}}\quad \Phi(\boldsymbol{\beta})=\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2 \tag{4}$$

ただし、$\|\mathbf{x}\|$ は $\mathbf{x}$ のユークリッドノルムを表す。式(4)の一階条件から、次の正規方程式が導かれる。

$$\mathbf{X}^t\mathbf{X}\boldsymbol{\beta}=\mathbf{X}^t\mathbf{y} \tag{5}$$

$\mathbf{X}^t\mathbf{X}$ が正則行列であれば式(4)の解 $\boldsymbol{\beta}_O$ が得られる。

$$\boldsymbol{\beta}_O=(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} \tag{6}$$

ところで、式(2)に観測データを代入した $\boldsymbol{\beta}$ に関する方程式(7)は、残差の自由度が正であれば、特別な

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta} \tag{7}$$

場合を除き一般に解を持たない不適切問題である。最小二乗法(4)は不適切問題(7)の $\boldsymbol{\beta}$ の近似解を求めようとするものであり、近似解 $\boldsymbol{\beta}_O$ を与える $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ は $\mathbf{X}$ の最小二乗型一般逆行列と呼ばれる[3]。

*キーワード：調査論、情報処理

**正員、工修、東京大学大学院工学系研究科社会基盤工学専攻
（〒113 文京区本郷７－３－１
TEL：03-3812-2111 ext.6128 FAX：:03-5689-7290 ）
***正員、工博、東京大学大学院工学系研究科社会基盤工学専攻
（ 同上 ）
****正員、工修、大成建設株式会社
（〒169 新宿区百人町３－２５－１ サンケン）

$\boldsymbol{\beta}_O$ は不偏性・一致性・効率性などの統計的に望ましい性質を持つ。

$$E(\boldsymbol{\beta}_O) = \boldsymbol{\beta}^* \tag{8}$$

$$Var(\boldsymbol{\beta}_O) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{9}$$

ただし、$\boldsymbol{\beta}^*$はパラメータの真の値である。

### (2) 多重共線性

多重共線性とは、説明変数の間に高い相関があり $\mathbf{X}'\mathbf{X}$ がランク落ちに近い状態をいう。このことは、方程式(5)が解の安定性を欠く不適切問題であることを意味する。このとき、式(9)からも分かるように多重共線性を生じている変数に関るパラメータの推定値の分散が大きくなる。

多重共線性を診断する尺度には、分散拡大要因 *variance inflation factor* 、条件数（状態数） *condition number* などと呼ばれるものがある[4]。ここでは条件数について簡単に説明する。

行列 **A** の条件数とは線形システム

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} \tag{10}$$

においてインプット $\boldsymbol{x}$ の変化 $\delta\boldsymbol{x}$ がアウトプット $\boldsymbol{y}$ に及ぼす影響 $\delta\boldsymbol{y}\ (= \mathbf{A}\delta\boldsymbol{x})$ の相対的な大きさを、ノルムの比で測ったときの最大値である。とくにノルムとして行列のスペクトルノルムを用いた場合には、**A** の条件数 $\kappa(\mathbf{A})$は次のようになる[3]。

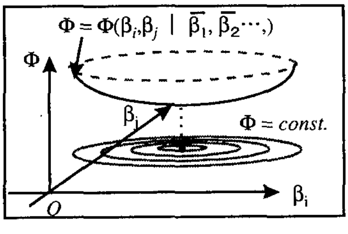
$$\kappa(\mathbf{A}) = \lambda_1/\lambda_k \tag{11}$$

ただし、$\lambda$は正則行列 **A** の固有値であり、

$$\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_k\ (\geqq 0) \tag{12}$$

正規方程式(5)の解 $\boldsymbol{\beta}$ への観測ベクトル $\mathbf{y}$ の誤差の影響は、$\mathbf{X}'\mathbf{X}$ の条件数 $\kappa(\mathbf{X}'\mathbf{X})$で与えられる。多重共線性が生じると最小固有値が 0 に近くなるため、条件数 $\kappa(\mathbf{X}'\mathbf{X})$が大きくなる。すなわち、わずかな観測誤差がパラメータの推定結果に大きな影響を与える（解の不安定性）。

$\mathbf{X}'\mathbf{X}$ は 2 次形式の最適化問題(4)の Hesse 行列（の1/2）でもある（図—1）。




図—1 多重共線性下の解の不安定性

$$\frac{\partial \Phi}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\ , \frac{\partial^2 \Phi}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = 2\mathbf{X}'\mathbf{X} \tag{13}$$

## 3. 多重共線性の改善手法

多重共線性の問題に対しては、従来から様々な解決法が提案されてきた。ここではそれらのうち、主成分回帰と Ridge 回帰について簡単に説明する。

### (1) 主成分回帰

主成分回帰では **X** に関して式(14)のような主成分分析を行い、互いに無相関な主成分を導出する。

$$\mathbf{Z}=\mathbf{XP} \tag{14}$$

ここで **P** は $\mathbf{X}'\mathbf{X}$ の固有ベクトルから成る $k\times k$ の直交行列（$\mathbf{P}^{-1}=\mathbf{P}'$）である。*0* に近い主成分をいくつか除いたあとの主成分行列を $\mathbf{Z}_L$（$l\times k$）と表わし、**y** を $\mathbf{Z}_L$ に回帰する。

$$\mathbf{y}=\mathbf{Z}_L\boldsymbol{\alpha} \tag{15}$$

$\mathbf{Z}_L$ に対応した **P** を $\mathbf{P}_L$（$l\times k$）、$\boldsymbol{\Lambda}$ を $\boldsymbol{\Lambda}_L$（$l\times l$）と書くと、主成分回帰によるパラメータ推定値 $\boldsymbol{\beta}_S$ は

$$\boldsymbol{\beta}_S = \mathbf{P}_L\boldsymbol{\alpha}_L \tag{16}$$

$$= \mathbf{P}_L(\mathbf{P}_L'\mathbf{X}'\mathbf{X}\,\mathbf{P}_L)^{-1}\mathbf{P}_L'\mathbf{X}'\mathbf{y} \tag{17}$$

（ただし $\boldsymbol{\alpha}_L=(\mathbf{Z}_L'\mathbf{Z}_L)^{-1}\mathbf{Z}_L'\mathbf{y}$）

$$E(\boldsymbol{\beta}_S) = \mathbf{P}_L(\mathbf{P}_L'\mathbf{X}'\mathbf{X}\,\mathbf{P}_L)^{-1}\mathbf{P}_L'\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta}^* \neq \boldsymbol{\beta}^* \tag{18}$$

$$Var(\boldsymbol{\beta}_S) = \sigma^2\,\mathbf{P}_L\boldsymbol{\Lambda}_L^{-1}\mathbf{P}_L' \tag{19}$$

主成分回帰によって得られるパラメータは不偏性を持たないが、OLS の場合に比べ分散（式(19)の対角要素）は必ず小さくなる。

### (2) Ridge 回帰

Ridge 回帰は、Hoerl and Kennard[5)6)]によって提唱された方法であり、式(5)の $\mathbf{X}'\mathbf{X}$ を適当なスカラー（Ridge パラメータと呼ばれる）$c$ と単位行列 **I** を用いて $\mathbf{X}'\mathbf{X} + c\,\mathbf{I}$ に置き換えるものである。

$$(\mathbf{X}'\mathbf{X} + \mathrm{c}\,\mathbf{I})\,\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} \tag{20}$$

これより、Ridge 推定量が式(21)のように得られる。

$$\boldsymbol{\beta}_R = (\mathbf{X}'\mathbf{X} + \mathrm{c}\,\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \tag{21}$$

式(22)に示すように Ridge 推定量も主成分回帰による推定量同様バイアス推定量であるが、その分散（式

(23)の対角要素）は OLS の場合に比べ必ず小さくなる[4]。

$$E(\boldsymbol{\beta}_R) = (\mathbf{X}'\mathbf{X} + c\,\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^* \neq \boldsymbol{\beta}^* \quad (22)$$

$$Var(\boldsymbol{\beta}_R) = \sigma^2 (\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-1} (\mathbf{X}'\mathbf{X}) (\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-1} \quad (23)$$

## 4. 主成分回帰と Ridge 回帰に関する逆解析的考察

解の不安定性のような不適切性を改善して問題を解くことは、解析の適切化（正則化） *regularization* と呼ばれる。適切化の方法としては、次の二つがある[2]。

(i) 逆変換の作要素を変えて平滑化するもの

(ii) 解空間を狭めるもの

3.で説明した主成分回帰と Ridge 回帰は、次の(1)で述べるようにいずれも(i)に分類される適切化手法であるが、Ridge 回帰については(2)･(3)で述べるように(ii)の適切化手法として解釈することも可能である。

### （1）主成分回帰と Ridge 回帰の相互関係

まず、主成分回帰と Ridge 回帰の相互関係について説明するために、$\mathbf{X}'\mathbf{X}$ の固有値分解（特異値分解）を行う。

$$\mathbf{X}'\mathbf{X} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' = \sum_{i=1}^{k} \lambda_i \mathbf{p}_i \mathbf{p}_i' \quad (24)$$

ここで、$\mathbf{p}_i$ は固有値 $\lambda_i$ に対応する固有ベクトル、$\boldsymbol{\Lambda}$ は $\mathbf{X}'\mathbf{X}$ の固有値（$\lambda_1 \geqq \lambda_2 \geqq \cdots \lambda_k \geqq 0$ とする）を対角要素とする対角行列である。

式(24)における直交行列 $\mathbf{P}$ 及び $\mathbf{P}'$ が表す一次変換はベクトルの回転・鏡映を表し、ベクトルの長さを変えない。一方、$\boldsymbol{\Lambda}$ の表す一次変換は単位球面 $\{\mathbf{x}: \|\mathbf{x}\|^2 = 1\}$ にひきおこす歪みの度合いを示す[7]。多重共線性が存在すると、$\boldsymbol{\Lambda}$ の表す一次変換は単位球面を極端に歪んだ楕円体面に移す（図―２）。

$\mathbf{X}'\mathbf{X}$ が正則であればその逆行列は

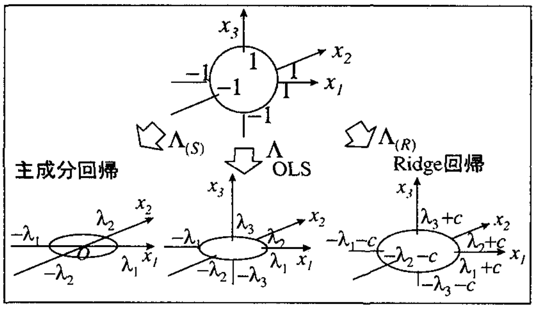
$$(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' = \sum_{i=1}^{k} \frac{1}{\lambda_i} \mathbf{p}_i \mathbf{p}_i' \quad (25)$$

固有値分解を用いると、主成分回帰や Ridge 回帰の相互の関係が明確に理解できる[8]。たとえば $\lambda_k$ が他の固有値に比べて非常に $0$ に近いとき、これを除去する主成分回帰と Ridge 回帰は、それぞれつぎのような置き換えに等しい（図２）。

$$(\mathbf{X}'\mathbf{X})^{-1} \rightarrow \mathbf{P}\boldsymbol{\Lambda}_{(S)}\mathbf{P}' = \sum_{i=1}^{k-1} \frac{1}{\lambda_i} \mathbf{p}_i \mathbf{p}_i' \quad (26)$$

$$(\mathbf{X}'\mathbf{X})^{-1} \rightarrow \mathbf{P}\boldsymbol{\Lambda}_{(R)}\mathbf{P}' = \sum_{i=1}^{k} \frac{1}{\lambda_i + c} \mathbf{p}_i \mathbf{p}_i' \quad (27)$$

このように、主成分回帰と Ridge 回帰は、いずれも逆変換作用素$(\mathbf{X}'\mathbf{X})^{-1}$を修正する適切化手法である。



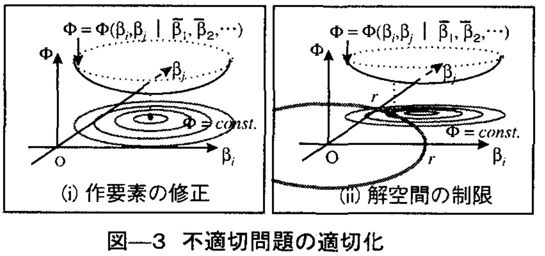

図―２ 対角行列の表す一次変換

### （２） Ridge 回帰に関するその他の解釈

Ridge 回帰に関しては、これまで様々な解釈が加えられた（例えば Maddala[9]）。その一つは、次のようにパラメータのノルムに関して制約条件のついた制限最小二乗法として解釈するものである。

$$\min_{\boldsymbol{\beta}} \ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \qquad r\text{：定数} \quad (28)$$
$$s.t. \ \ \|\boldsymbol{\beta}\|^2 = r^2$$

式(28)に Lagrange 乗数法を用いて作成した制約無しの最適化問題の一階条件から、Ridge 推定量を与える式(20)が容易に導出される。




図―３ 不適切問題の適切化

Ridge 回帰を式(28)に示す制限付き最小二乗法と解釈することにより、解空間を制限する適切化手法とみることも可能である（図―３）。

### （３） Tikhonov の適切化

さらに、Ridge 回帰は(ii)に分類される適切化手法の

一つである Tikhonov の適切化（近似）としても解釈可能である。

Tikhonov の適切化は Hilbert 空間上の線形作用素を対象とする抽象化された理論である。工学で扱う線形の連立方程式、微分方程式等の多くは、Hilbert 空間上の作用素方程式の問題に帰着させることが可能であるから [10]、不適切性を改善する方法としての Tikhonov の適切化は応用範囲が広い。

$f,g$ を $f \in F, g \in G$ （ $F,G$：Hilbert 空間 ）とし、$T$ を $F$ から $G$ へのコンパクト線形作用素とする。いま、線形作用素方程式(29)の近似解を与える汎関数(30)

$$Tf=g \tag{29}$$

$$\prod_{f \in F} = \|Tf - g\|^2 \tag{30}$$

が不適切であるとする。Tikhonov の適切化法とは式(30)の汎関数のかわりにこれに安定化項をつけた

$$\prod_{f \in F} = \|Tf - g\|^2 + \alpha\|f\|^2 \tag{31}$$

$\alpha$：パラメータ

を用いるものである [11]。

多重共線性下における不適切問題(4)に Tikhonov の適切化法を適用し、安定化項 を加えると

$$\min_{\boldsymbol{\beta}} \quad F(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + c\|\boldsymbol{\beta}\|^2 \tag{32}$$

このとき一階条件（Euler 方程式）は

$$\mathbf{X}'\mathbf{X}\,\boldsymbol{\beta} + c\,\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} \tag{33}$$

となって式(20)が導出され、Ridge 推定量が得られる。なお、式(32)は式(28)の制限最小二乗法において、ベクトルのノルム $r$ を $0$ としたものと等価である。

### (4) Tikhonov の適切化のその他の応用

$\mathbf{X}'\mathbf{X}$ が完全にランク落ちした $\lambda_k=0$ のとき、$\mathbf{X}'\mathbf{X}$ の一般逆行列( $\mathbf{X}'\mathbf{X}$ )$^+$は式(34)のように表される。

$$(\mathbf{X}'\mathbf{X})^+ = \sum_{i=1}^{k-1} \frac{1}{\lambda_i} \mathbf{p}_i \mathbf{p}_i' \tag{34}$$

一般逆行列のうち Moore-Penrose 型一般逆行列 $\mathbf{X}_{MP}^-$ には、次の公式が知られている [12]。

$$\mathbf{X}_{MP}^- = \lim_{\delta \to 0} (\mathbf{X}'\mathbf{X} + \delta\mathbf{I})^{-1}\mathbf{X}' \tag{35}$$

式(35)は十分小さな $\delta$ に対し、( $\mathbf{X}'\mathbf{X} + \delta\mathbf{I}$ )$^{-1}\mathbf{X}'$ が $\mathbf{X}$ の Moore-Penrose 型一般逆行列の近似を与えることを意味するが、式(35)の右辺は Ridge 回帰における式(21)の係数と同じ形式をしている。式(35)は方程式の解が存在しない場合における Tikhonov 近似から容易に導かれる。

$\mathbf{X}'\mathbf{X}$ が完全にランク落ちしている場合には、Ridge 回帰は M-P 一般逆行列の近似を用いた推定とも解釈できる。式(34)と Ridge 回帰との関係については、Golub et al. [13]に詳しく示されている。

## 5.おわりに

本稿では、線形回帰モデルにおいてしばしば分析者を悩ませる多重共線性について、その改善方法を逆解析という観点から整理した。尚、実際の適用等については、講演会発表時に示したい。

**謝辞**　Ridge 回帰が Tikhonov の適切化と本質的に同じであることは、東京大学地震研究所堀宗朗助教授からご指摘いただいた。記して感謝の意を表したい。

## 参考文献

1)久保司郎：逆問題の考え方と枠組　工学的側面を中心として，数理科学，No.403，pp.28-34，サイエンス社，1997.
2)久保司郎：逆問題，培風館，1992.
3)中川徹・小柳義夫：最小二乗法による実験データ解析，東京大学出版会，1982.
4)蓑谷千凰彦：計量経済学の新しい展開，多賀出版，1992.
5)Hoerl, A. E. and Kennard, R. W. ：Ridge Regression : Biased Estimation for Nonorthogonal Problems , Technometrics, 12, No.1, pp.55-67, 1970.
6)Hoerl, A. E. and Kennard, R. W. ：Ridge Regression : Applications to Nonrothogonal Problems, Technometrics, 12, No.1, pp.69-82, 1970.
7)田辺国士：数値的方法における特異値，数理科学，No.212，pp.46-50，1981.
8)Marquardt, D. W. ：Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation, Technometrics, 12, No.3, pp.591-612, 1970.
9)Maddala, G.S.：Introduction to Econometrics Second Edition, Premtice-Hall, 1988［和合肇訳：計量経済分析の方法，シーエーピー出版，1996］.
10)篠崎寿夫他共著：ヒルベルト空間上の線形作要素入門，現代工学社.，1996.
11)Groetsch, C. W.：Inverse Problems in the Mathematical Sciences, Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, 1993［金子晃他共訳：数理科学における逆問題，サイエンス社，1996］
12)田島稔・小牧和雄：最小二乗法の理論とその応用，東洋書店，pp.349-350，1986.
13)Golub, G. H., Heath, M. and Wahba, A, G.：Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter, Technometrics, 21, No.2, pp.215-223, 1979.