多重共線性に対する適切化手法とその実証的比較研究
－ 地価の回帰分析を例として －

Regularization Methods for Multicollinearity and Their Empirical Comparison
: An Example of Application to Regression Analysis of Land Price

堤盛人*・清水英範**・井出裕史***
Morito TSUTSUMI, Eihan SHIMIZU and Hiroshi IDE

*正会員　工修　東京大学大学院助手　工学系研究科　社会基盤工学専攻
**正会員　工博　東京大学大学院教授　工学系研究科　社会基盤工学専攻
***学生会員　東京大学大学院修士課程　工学系研究科　社会基盤工学専攻
（〒113-8656　東京都文京区本郷7－3－1）

Regression analysis is the most classical but still a very useful back analysis in social sciences. Multicollinearity is one of the most serious ill-posedness and many conventional regularization methods have been provided. The paper focuses on the multicollinearity problem and attempts to compare the theoretical and practical characteristics among regularization methods. The paper discuses principal component regression and ridge regression which are the most well-known ones among them. It is shown that they are compatible with the traditional regularization methods in inverse analysis. The two methods are applied to regression analysis of land price in order to demonstrate their practical characteristics.

Key Words : multicollinearity, regularization, principal component regression, ridge regression

1. はじめに

社会資本整備プロジェクトの計画段階では、都市あるいは地域における社会・経済政策上の観点から、プロジェクトの実施による人口や地価の変化などそのプロジェクトが地域に及ぼす影響を客観的かつ定量的に評価する必要がある。そのため、従来から様々な計量モデルが開発され、実際に適用されている。

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」と定義すれば、人口や地価などの地域の現況を再現するモデル、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「逆問題」である。様々な逆解析手法が存在する中で、分析対象としての変数をそれと直接的に関連する変数によって説明する回帰分析は、現在でも最も基本的かつ重要な手法の一つである。

回帰モデルのパラメータ推定において説明変数間が高い相関をもつ多重共線性が存在している状況は、解の安定性が損なわれるという意味で非適切であり、従来から様々な適切化手法が提案されてきた[1),2)]。非実験科学の分野ではデータの受動性が顕著であり、この多重共線性に悩まされることが非常に多い[3)]。

ところでここ2～3年、プロジェクトの優先順位決定についての客観的基準の確立並びに決定過程における透明性の確保がこれまで以上に厳しく問われるようになり、いわゆる費用便益分析の重要性が増している[4)]。便益の計測手法の一つである資産価値アプローチと呼ばれる方法では、プロジェクトの実施による地価の上昇分を計測することにより便益を推定する。その際、地価が都心からの交通所要時間や周辺環境などの土地特性によって決まると考え、土地の市場における情報からこの地価の関数を推定することにより、交通プロジェクトによる地価上昇を計測するヘドニック・アプローチが用いられることが多い。ヘドニック・アプローチでは、地価の決定要因のうち交通プロジェクト以外の要因を取り去る必要があるため、多くの変数を導入して地価関数を推定する。このため、地価関数として線形式を用いると多重共線性が発生することが多い[5)]。

本研究は、地価の推定を目的とした回帰分析を例に、多重共線性が生じている状況に対して、これまで提案されてきたいくつかの適切化手法を適用し、パラメータの推定結果や予測精度がどの程度異なるかについて実証的な考察を行う。

以下、まず第2章では、多重共線性について極く簡単に説明しながら、本論文で用いる記号等を示す。

次に第3章において、多重共線性に対する適切化手法

について説明する。パラメータの値に関する確率分布を事前情報として持つ場合には、ベイズ的な解釈に基き、観測データの情報と併せて適切化を行う方法も考えられる[6),7)]。しかしながら、政策分析を目的として社会経済活動をモデル化する場合には、分析者がパラメータについてこれらの事前情報を持っていることはまれである[8),9)]。そこで本論文では、非ベイズ的な適切化手法のみを扱うこととし、ベイズ的な手法については若干触れるにとどめる。具体的には、社会経済分析において多重共線性に対する適切化手法として用いられることの多い主成分回帰とリッジ回帰を取り上げ、逆解析という枠組みの中でその位置づけを整理する。

非適切逆問題における適切化手法は、観測データに関する条件からなる問題に対し他の条件を加える等の変更を行い、元の条件と追加された条件とのバランスを取って解を見つけるという意味で「折衷」と呼ばれる[10)]。リッジ回帰についても、この折衷の役割を担うリッジ・パラメータ（適切化パラメータ）の同定方法に関して多くの既存研究が存在する。そこで第４章では、統計学の枠組みの中で比較的興味深い発展を経てきたリッジ・パラメータの同定問題について既存研究を整理する。

第５章では、公示地価データを用いた回帰モデルに主成分回帰とリッジ回帰を適用し、それらが推定結果や予測精度にどの程度影響を及ぼすかについて考察する。

2. 多重共線性

2.1 最小二乗法による回帰モデルのパラメータ推定と多重共線性

以下本論文では、y を被説明変数、x_j（$j = 1,2,\cdots, m$）を説明変数、β_j（$j = 0,1,2,\cdots, m$）をパラメータとする線形モデル(1)を扱う。

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m \tag{1}$$

データの数を n（$>m+1$）とし、誤差項を含んだ統計モデルを以下のように行列表記する。なお t はベクトル及び行列の転置を表すとする。

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

$$\boldsymbol{y} = (y_1, \cdots, y_i, \cdots, y_n)^t, \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_m)^t, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{m1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{mn} \end{pmatrix}$$
$$\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_i, \cdots, \varepsilon_n)^t,$$

$\boldsymbol{\varepsilon}$は誤差項であり、式(3)(4)を満たすものと仮定する。

$$E(\boldsymbol{\varepsilon}) = \boldsymbol{0} \tag{3}$$
$$Var(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I} \quad (\sigma^2 : \text{定数}) \tag{4}$$

パラメータ推定のための通常最小二乗法（Ordinary Least Squares Method : OLS）は、二次形式の最適化問題である（式(5)）。

$$\min_{\boldsymbol{\beta}} \quad \Phi(\boldsymbol{\beta}) = \left\| \boldsymbol{y} - \boldsymbol{X\beta} \right\|^2 \tag{5}$$

ただし、$\|\|$ はベクトルのユークリッド・ノルムを表す。

式(5)の一階条件から、いわゆる正規方程式(6)が導かれる。

$$\boldsymbol{X}^t \boldsymbol{X\beta} = \boldsymbol{X}^t \boldsymbol{y} \tag{6}$$

ここで、$\boldsymbol{X'X}$が正則行列であればOLS推定量 $\boldsymbol{\beta}_O$ が得られる。

$$\boldsymbol{\beta}_O = \left(\boldsymbol{X}^t \boldsymbol{X} \right)^{-1} \boldsymbol{X}^t \boldsymbol{y} \tag{7}$$

OLS 推定量 $\boldsymbol{\beta}_O$ は不偏性・一致性・効率性などの統計学的に望ましい性質を持つことが知られている。

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

$$Var\left(\boldsymbol{\beta}_O\right) = \sigma^2 \left(\boldsymbol{X}^t \boldsymbol{X} \right)^{-1} \tag{9}$$

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

多重共線性とは、説明変数の間に高い相関があることを言い、式(10)で表される式(5)の Hesse 行列（の 1/2）である$\boldsymbol{X'X}$がランク落ちに近い状態となる。

$$\frac{\partial^2 \Phi}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} = 2\boldsymbol{X}^t \boldsymbol{X} \tag{10}$$

つまり、式(5)あるいは式(6)が解の安定性を欠く非適切問題となり、パラメータ推定値の分散が大きくなる。

この多重共線性 *multicollinearity* という用語は、1934年 Ragnar Frisch が彼の合流分析という本の中で最初に使用したと言われる[8)]。

多重共線性を診断する尺度としてはいくつかの指標があるが[9)]、自然科学系においてなじみの深いものは、次の条件数 *condition number*[11)]であろう。

行列$\boldsymbol{A}$の条件数とは線形システム

$$\boldsymbol{y} = \boldsymbol{AX} \tag{11}$$

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　$\delta \boldsymbol{y}$ ($= \boldsymbol{A}\, \delta \boldsymbol{x}$) 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。行列$\boldsymbol{A}$が正則行列で、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、$\boldsymbol{A}$の条件数 $\kappa(\boldsymbol{A})$ は次式に示すとおりである[12)]（$\lambda_{max}, \lambda_{min}$ はそれぞれ行列$\boldsymbol{A}$の最大、最小固有値）。

$$\kappa(\boldsymbol{A}) = \lambda_{max} / \lambda_{min} \tag{12}$$

正規方程式(6)の解$\boldsymbol{\beta}$への観測ベクトル $\boldsymbol{y}$ の誤差の影響は、$\boldsymbol{X'X}$ の条件数 $\kappa(\boldsymbol{X'X})$ で与えられる。多重共線性が生じると最小固有値が0に近くなり、条件数 $\kappa(\boldsymbol{X'X})$ 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）。

2.2 変数の基準化

第３章で述べるように、リッジ回帰を適用する際に、変数の値が単位に依存することに起因する問題が生じる。そのため、次のような変数の規準化が行われることが多い[9]。

$$u_{ji}=\frac{\left(x_{ji}-\overline{x_j}\right)}{s_j} \quad (i=1,2,\cdots,n) \tag{13}$$

ただし、

$$\overline{x_j}=\sum_{i=1}^{n}x_{ji} \tag{14}$$

$$s_j=\frac{1}{n}\sqrt{\sum_{i=1}^{n}(x_{ji}-\overline{x_j})^2} \tag{15}$$

このとき各 i に対し、u_{ji} は平均0、分散1となる。

$$E[u_{ji}]=\frac{1}{n}\sum_{i=1}^{n}u_{ji}=0 \tag{16}$$

$$Var[u_{ji}]=\frac{1}{n}\sum_{i=1}^{n}(u_{ji}-E[u_{ji}])^2=1 \tag{17}$$

第３章・第４章における重要な結果は、変数の基準化の有無に関わらず成立するが、以下本論文では、式(13)～(15)に示した手順に従い、説明変数が基準化されているものとする。そして、表記上の理由から、基準化された変数からなる行列を改めて $\boldsymbol{X}$ とおくこととする。

3. 多重共線性の適切化手法

多重共線性の問題に対しては、様々な統計学的適切化手法が提案されてきた[13]。ここではそれらのうち、主成分回帰とリッジ回帰について説明する。

3.1 主成分回帰

主成分回帰 *Principal Component Regression* では $\boldsymbol{X}$ に関して式(18)に示す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。

$$\boldsymbol{Z}=\boldsymbol{XP} \tag{18}$$

ここで $\boldsymbol{P}$ は $\boldsymbol{X}'\boldsymbol{X}$ の固有ベクトルから成る $(m+1)\times(m+1)$ の直交行列（$\boldsymbol{P}^{-1}=\boldsymbol{P}'$）である（$\boldsymbol{X}'\boldsymbol{X}$ が重複固有値を持つときは、直行化された固有ベクトルを用いて $\boldsymbol{P}$ とする）。0 に近い固有値に対応する主成分をいくつか除いたあとの主成分行列を $\boldsymbol{Z}_L$（$l\times(m+1)$）と表わし、y を $\boldsymbol{Z}_L$ に回帰する。

$$y=\boldsymbol{Z}_L\boldsymbol{\alpha}_L \tag{19}$$

これにOLSを用いると、式(20)が得られる。

$$\boldsymbol{\alpha}_L=\left(\boldsymbol{Z}_L'\boldsymbol{Z}_L\right)^{-1}\boldsymbol{Z}_L'y \tag{20}$$

$\boldsymbol{\Lambda}$ を $\boldsymbol{X}'\boldsymbol{X}$ の固有値 λ_k（$\lambda_1\geqq\lambda_2\geqq\cdots\geqq\lambda_{m+1}\geqq 0$ とする）を対角要素とする対角行列とし、$\boldsymbol{Z}_L$ に対応した $\boldsymbol{\Lambda}$ を $\boldsymbol{\Lambda}_L$（$l\times l$）、$\boldsymbol{P}$ を $\boldsymbol{P}_L$（$l\times(m+1)$）と書くと、主成分回帰によるパラメータ推定値 $\boldsymbol{\beta}_S$ は次式により求められる。

$$\boldsymbol{\beta}_S=\boldsymbol{P}_L\boldsymbol{\alpha}_L=\boldsymbol{P}_L\left(\boldsymbol{P}_L'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{P}_L\right)^{-1}\boldsymbol{P}_L'\boldsymbol{X}'y \tag{21}$$

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(23)の対角要素)は必ず小さくなる[9]。

$$E(\boldsymbol{\beta}_S)=\boldsymbol{P}_L\left(\boldsymbol{P}_L'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{P}_L\right)^{-1}\boldsymbol{P}_L'\boldsymbol{\beta}^*\neq\boldsymbol{\beta}^* \tag{22}$$

$$Var(\boldsymbol{\beta}_S)=\sigma^2\boldsymbol{P}_L\boldsymbol{\Lambda}_L^{-1}\boldsymbol{P}_L' \tag{23}$$

主成分分析は、1901年 K. Pearson によって創始され、1933年 H. Hotelling によってより一般的な形で再発見されたと言われる[3]。最初に主成分分析を回帰分析へ用いた主成分回帰法が、いつ誰によって初めて用いられたのかは定かではないが、1957年 Kendall による例は最も古いものの一つのようである[12]。

3.2 リッジ回帰

リッジ回帰 *Ridge Regression* は、Hoerl and Kennard[13),14)]によって提唱された方法であり、リッジ・パラメータと呼ばれる適当なスカラー c と単位行列 $\boldsymbol{I}$ を用いて、式(7)の $\boldsymbol{X}'\boldsymbol{X}$ を $\boldsymbol{X}'\boldsymbol{X}+c\boldsymbol{I}$ に置き換えるものである。（リッジ回帰のアイデア自体は、これより10年程前に遡る[15),16)]が、統計学的な手法というよりは、むしろ数値解析的手法としての意味合いが強く、そこでは推定量の性質等について議論されていない。）

$$\boldsymbol{\beta}_R=\left(\boldsymbol{X}'\boldsymbol{X}+c\boldsymbol{I}\right)^{-1}\boldsymbol{X}'y \tag{24}$$

なお、単位行列の定数倍 $c\boldsymbol{I}$ に代えてより一般の正値の対角行列を用いる一般化リッジ回帰も提案されているが[14]、本論文ではもっとも単純な式(24)のケースのみを扱うこととする。

式(25)から明らかなように、説明変数をそのまま用いると、その測定単位によってリッジ回帰による推定結果が異なる。そこで、単位系に依存せず同一の結果を導く方法として、しばしば、2.2で述べたような変数の基準化が行われる。ただし、この変数の基準化は、Bayes 的な意味での事前情報を捨てることになる。

式(25),(26)に示すようにリッジ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(26)の対角要素)はOLSの場合に比べ必ず小さくなる[9]。

$$E(\boldsymbol{\beta}_R)=\left(\boldsymbol{X}'\boldsymbol{X}+c\boldsymbol{I}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}^*\neq\boldsymbol{\beta}^* \tag{25}$$

$$Var(\boldsymbol{\beta}_R)=\sigma^2\left(\boldsymbol{X}'\boldsymbol{X}+c\boldsymbol{I}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}+c\boldsymbol{I}\right)^{-1} \tag{26}$$

3.3 主成分回帰並びにリッジ回帰と典型的な逆解析手法との関係

非適切問題に対する適切化の手法については、多く研究がなされているが、それらは大きく次の二つ方法に分けて考えることができる[17]。

(i) 逆変換の作用素を変えて平滑化するもの

(ii) 解空間を狭めるもの

本節では、(i)に属する打切り特異値分解並びに(ii)に属するダンプ付き最小二乗法及び Tikhonov の適切化との関連において主成分回帰とリッジ回帰の位置づけを簡単に整理する。

(1) 打切り特異値分解

まず、$\boldsymbol{X}'\boldsymbol{X}$ の固有値分解を次式に示す。

$$\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'=\sum_{i=1}^{m+1}\lambda_i \boldsymbol{p}_i \boldsymbol{p}_i{}' \tag{27}$$

ここで、$\boldsymbol{p}_i$は固有値 λ_iに対応する固有ベクトルである。

式(27)における直交行列 $\boldsymbol{P}$ 及び $\boldsymbol{P}'$ 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 $\{x:\|x\|^2=1\}$ にひきおこす歪みの度合いを示す[18]。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。

$\boldsymbol{X}'\boldsymbol{X}$が正則であればその逆行列は式(28)に示すように分解・展開することができる。

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}=\boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}'=\sum_{i=1}^{m+1}\frac{1}{\lambda_i}\boldsymbol{p}_i \boldsymbol{p}_i{}' \tag{28}$$

ここで、λ_{m+1}が他の固有値に比べて非常に小さく 0 に近いとき、これを除去する主成分回帰は次のような置き換えに等しい[20]。

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\rightarrow \boldsymbol{P}\boldsymbol{\Lambda}_{(s)}{}^{-1}\boldsymbol{P}'=\sum_{i=1}^{m}\frac{1}{\lambda_i}\boldsymbol{p}_i \boldsymbol{p}_i{}' \tag{29}$$

ただし、$\boldsymbol{\Lambda}_{(s)}$は固有値 λ_k (k=1,…, m) 及び 0 を対角要素とする対角行列である。

固有値分解という名称は正方行列の分解に対するものである。正方行列以外の行列も含めた任意の行列についての同様な分解は特異値分解と呼ばれ、固有値に対応した対角行列の成分は特異値と呼ばれる[11]。小さい特異値を 0 と置き換えることにより、式(29)のように展開を途中で打切る方法は、非適切逆問題に対する適切化手法として最もオーソドックスなものの一つである打切り特異値分解 *Truncated Singular Value Decomposition* である[19]。

固有値分解を用いると、リッジ回帰はつぎのような置き換えに等しいことが分かり、主成分回帰とリッジ回帰の相互の関係が一層明らかになる[20]。

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\rightarrow \boldsymbol{P}\boldsymbol{\Lambda}_{(R)}\boldsymbol{P}'=\sum_{i=1}^{m+1}\frac{1}{\lambda_i+c}\boldsymbol{p}_i \boldsymbol{p}_i{}' \tag{30}$$

ここで、$\boldsymbol{\Lambda}_{(R)}$は、固有値$\lambda_k+c$ (k=1,…, m+1) を対角要素とする対角行列である。

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

(2) ダンプ付き最小二乗法

リッジ回帰に関しては、その手法の提案以降、様々な解釈が加えられてきた[1),8)]。その一つは、次に示すように、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。

$$\min_{\beta}\ \|y-X\beta\|^2 \qquad b:\text{定数} \tag{31}$$
$$s.t.\ \ \|\beta\|^2=b^2$$

式(31)に Lagrange 乗数法を用いて作成した制約無し最適化問題の一階条件は式(32)となり、これから式(24)が導出される (cはここでは、Lagrange 乗数) 。

$$\left(\boldsymbol{X}'\boldsymbol{X}+c\boldsymbol{I}\right)\beta=\boldsymbol{X}'y \tag{32}$$

制限付き最小二乗法は、ダンプ付き最小二乗法 *Damped Least Squares Method* とも呼ばれる[21),22)]。リッジ回帰を式(31)に示す制限付き最小二乗法と解釈することにより、(ii)の解空間を制限する適切化手法とみることが可能である。

(3) Tikhonov の適切化

Tikhonov の適切化 (近似) も、(ii)に分類される適切化手法の一つである[19]。

多重共線性下における非適切問題である式(5)に Tikhonov の適切化法を適用する。ここで、安定化項には、βの関数として様々なものが考えられるが、ここでは単純パラメータベクトルのユークリッド・ノルムの二乗を採用する。

$$\min_{\beta}\ \ F(\beta)=\|y-X\beta\|^2+c\|\beta\|^2 \tag{33}$$

このとき一階条件は式(32)となる。

なお、式(33)は式(31)の制限最小二乗法において、ベクトルのノルム b を 0 としたものと等価であり、3. 3. (2)で述べたダンプ付き最小二乗法も Tikhonov の適切化法の特別な場合と考えることが可能である。

$\boldsymbol{X}'\boldsymbol{X}$が完全にランク落ちした$\lambda_{m+1}=0$のとき、$\boldsymbol{X}'\boldsymbol{X}$の一般逆行列 $(\boldsymbol{X}'\boldsymbol{X})^{+}$は式(34)のように表される。

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

ところで、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[23)]。

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

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、十分小さいδに対して $(\boldsymbol{X}'\boldsymbol{X}+\delta\boldsymbol{I})^{-1}\boldsymbol{X}'$が$\boldsymbol{X}$の Moore-Penrose 型一般逆行列の近似を与えることを意味する。右辺は、リッジ回帰における式(24)の係数と同じ形式をしており、$\boldsymbol{X}'\boldsymbol{X}$が完全にランク落ちしている場合には、リッジ回帰は Moore-Penrose 型一般逆行列の近似を用いた推定とも解釈可能である。このことは、Marquardt （非線形最適化問題における Tikhonov の適切化の例である Leveberg[22]-Marquardt[24] 法の提唱者）によって詳しく論じられている[20]。

4. 適切化パラメータの同定

4.1 リッジ・パラメータ

非適切逆問題における適切化手法は、観測データに関する条件からなる問題に対し他の条件を加える等の変更を行い、元の条件と追加された条件とのバランスを取って解を見つけるという意味で「折衷」と呼ばれる[19]。リッジ回帰では、リッジ・パラメータがこの折衷の中心的役割を担う。ダンプ付き最小二乗の解釈に従えば、パラメータベクトルのノルムについての情報があれば、リッジ・パラメータが Lagrange 乗数として求まる。あるいは、本論文では扱わない Bayes 的な解釈に基づけば、パラメータの値に関する確率分布を事前情報として持っていれば、観測データの情報を利用することでリッジ・パラメータを求めることができる。例えばパラメータの事前情報を$\boldsymbol{\beta}\sim N(0,\sigma_\beta^2\boldsymbol{I})$と仮定すれば、リッジパラメータは $c=\sigma^2/\sigma_\beta^2$ と与えられる[2),7)]。

しかしながら第1章でも述べたように、政策分析を目的として社会経済活動をモデル化する場合には、分析者がパラメータについてこのような事前情報を持っていることは非常にまれであるため、リッジ・パラメータの同定に際して新たに別の基準を導入する必要となる。何故なら、分散を小さくするだけが目的であれば、リッジ・パラメータをいくらでも大きくすることにより可能であるが、それにつれてバイアスも大きくなってしまうからである。

このような場合にしばしば導入される基準の一つに、次式に示す平均二乗誤差 *mean square error* がある。

$$MSE(\boldsymbol{\beta}) \equiv E\left[\left(\boldsymbol{\beta}-\boldsymbol{\beta}^*\right)'\left(\boldsymbol{\beta}-\boldsymbol{\beta}^*\right)\right]$$

$$= tr\left[Var(\boldsymbol{\beta})\right]+E\left[\boldsymbol{\beta}-\boldsymbol{\beta}^*\right]'E\left[\boldsymbol{\beta}-\boldsymbol{\beta}^*\right]$$

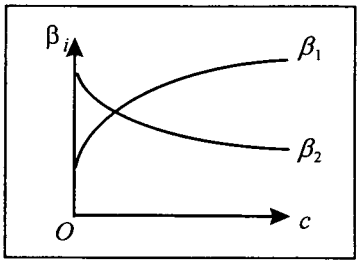
$$= Variance+(Bias)^2 \qquad (36)$$

式(36)における分散はリッジ・パラメータの減少関数であり、偏りはその増加関数であるため、MSE を最小にするようなリッジ・パラメータが必ず存在する[13]（モデルが線形であるので、このことは、説明変数の予測値 y に関する平均二乗誤差を考えても同じである）。しかしながら、式(36)の定義から明らかなように、パラメータの真の値$\boldsymbol{\beta}^*$を知らない限り MSE を最小にする意味で最適なリッジ・パラメータを選択することはやはり不可能である。そこで、次節に示すように、MSE の最小化を別な方法で間接的に達成しようとする同定方法がいろいろと提案されてきた。

4.2 リッジ・パラメータの同定方法

(1) リッジ・トレース

Hoerl and Kennard [14]は、リッジ・トレースというリッジ・パラメータ c とパラメータの値 β_i をプロットした図を用い、パラメータ β_i の値が安定するような c を採用することを提唱した(図－1)。しかし、この基準は極めて曖昧で分析者の恣意性が介入する余地が大きく、しかも、選んだリッジ・パラメータが、パラメータあるいは予測値の平均二乗誤差を OLS の場合に比べて小さくしているかどうか不明である。



図－1 リッジ・トレース

(2) Allen's PRESS

Allen[25]は、最も単純な交差確認法 *Cross Validation* を利用した方法を用いたモデル選択法を提唱した。n 個のデータのうち、i 番目のデータを除いた n-1 個のデータに OLS を用いたパラメータの推定値を$\boldsymbol{\beta}^{(i)}$とする。次に、このパラメータと i 番目のデータの説明変数を用いて、i 番目のデータの予測値を求め、実測値 y_i との差を誤差とみなす。n 個のデーター一つ一つがちょうど一度ずつ除かれるように n 回この手順を繰り返す。

Allen[26]によるリッジ・パラメータの選択方法は、このような手順によって求められた予測値の誤差の期待値 PRESS (*Prediction Sum of Squares*) を最小にする c を求めるものである。

$$PRESS(c) \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\left[\boldsymbol{X}\boldsymbol{\beta}^{(i)}(c)\right]_i-y_i\right)^2 \qquad (37)$$

ただし、$[\cdot]_i$ はベクトルの i 番目の要素を表す演算子とする。

紙面の都合で詳細は省くが、Allen の PRESS は説明変数の単位系（座標系）に依存し、次式により定義される行列 $\boldsymbol{A}(c)$ が対角行列に近づくと解が不安定になるという問題が生じる[27]。

$$A(c) \equiv X\left(X^t X + cI\right)^{-1} X^t \tag{38}$$

(3) Mallows' C_p

PRESS を誤差分散で基準化したものを J_p とおく。

$$J_p(c) \equiv \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\left[X\beta(c)\right]_i - y_i\right)^2 \tag{39}$$

Mallows は、J_p の期待値として定義される C_p を最小にするリッジ・パラメータを採用することを提唱した[27]。

$$C_p(c) \equiv E\left[J_p(c)\right]$$

$$\approx \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\left[X\beta(c)\right]_i - y_i\right)^2 + tr\left\{A(c)\right\}^2 - tr\left\{I - A(c)\right\}^2 \tag{40}$$

(4) GCV

Cross Validation GCV は適切化パラメータの推定法として最も有名なものの一つであり、Wahba[28]・Golub et al.[6] は Allen の PRESS を拡張し、座標依存性の問題点を克服した、次式に示す $V(c)$ を最小化する方法を提唱し、GCV (*Generalized Cross-Validation*)と呼んだ。

$$V(c) \equiv \frac{1}{n}\left\|(I - A(c)y)\right\|^2 \Big/ \left[\frac{1}{n}tr(I - A(c))\right]^2 \tag{41}$$

GCV では、Mallows による C_p で必要な誤差分散 σ^2 の推定が不要であるという利点もある。

GCV は、広く一般に、適切化パラメータの推定法として有名なものの一つである。

このように、適切化パラメータの選択方法について様々な方法が提案されているが、結局のところどれを採用するのが一番良いのかということについては、結論に至っていない。同様に、主成分回帰における折衷においても、どの値より小さい固有値を *0* にするのかを決める明確な基準は存在しない。

リッジ・パラメータ c の同定については、(1)は前述のような恣意性の問題が存在すること、また、(2)～(4)については(4)が包括的にこれらを包含すること、さらに、(3) (4)は現実問題としてほぼ同じくらいの c を与えること[9]から、次章の解析においては、(4)の Wahba・Golub らによる GCV 基準を用いる。なお、(3), (4)についての詳細な式展開は蓑谷[9]によって示されているので、必要に応じて参照されたい。

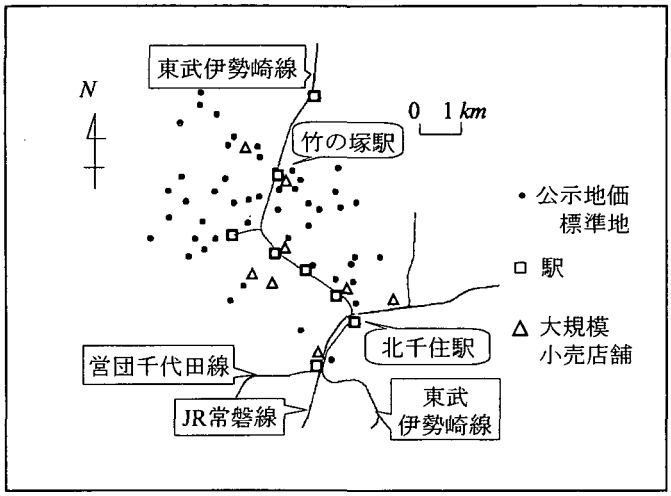
5. 地価データを用いた解析結果

本論文では、ある時点の地価が、その時点における都心からの交通所要時間や周辺環境などの土地特性によって決まると考え（クロス・セクション分析）、線形の回帰式によってその構造を推定する問題を取り上げる。

具体的には、東京都足立区内の平成８年公示地価の標準地のうち東武伊勢崎線北千住駅（ターミナル駅）から竹の塚駅までの各駅を最寄り駅とする住宅地合計 52 点を対象として、つぎのような回帰モデルを用いる。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_5 x_{5i} + \varepsilon_i \tag{42}$$
$$(i = 1, 2, \cdots, 52)$$

y： 公示地価（円／m²）
x_1： 標準地の地積（m²）
x_2： 最寄り駅までの距離（m）
x_3： 最寄り駅から北千住駅までの所要時間（分）
x_4： 法定容積率（%）
x_5： 最寄りの大規模小売店舗までの距離（m）



図－２　モデル適用対象地域の概略

x_3 は Windows 版「駅すぱあと」（(株)ヴァル研究所）を用いて算出し、x_5 は、大規模小売店舗名簿（東京都労働経済局、平成 8 年）をもとに GIS を用いて算出した。それ以外のデータは、「地価マップ東京都　平成８年（(財)土地情報センター編集）」による。

対象地域の概略は図－２に示すとおりである。通常最小二乗法によるパラメータの推定結果を表－１に示す。自由度修正済みの相関係数は 0.90 であった。

表－１　OLS によるパラメータの推定結果

	推定値	t 値	標準偏差
β_0	3.11×10^5	17.5	1.77×10^4
β_1	2.04×10^2	2.75	74.0
β_2	-25.0	-3.50	6.94
β_3	2.49×10^3	-2.03	1.23×10^3
β_4	2.33×10^2	4.92	47.3
β_5	-1.22	-0.13	9.26

表－1において、β_5 の標準偏差が推定値と比較して大きく、t 値から推定値の信頼性が低いと判断される。原因の一つとして多重共線性の存在が考えられるため、各説明変数間の相関を計算したところ、x_2 と x_5 の相関係数は0.85 であった。

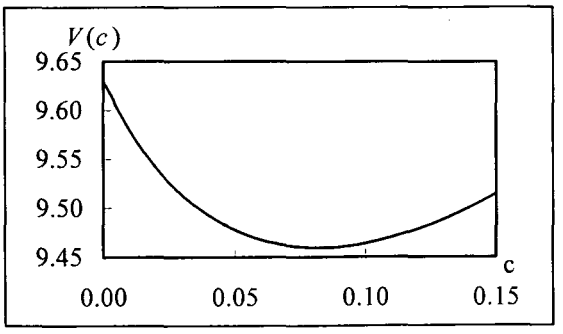
そこで、$X'X$ の固有値分解を行った（表－2）。条件数は20.7 であり、説明変数間の相関がパラメータの安定性を損なっていると考えられる。

表－2　固有値分解の結果

固有値 λ_j	$K_j=\lambda_j/\lambda_1$	寄与率
$\lambda_1=2.25$	$K_1=1.0$	47.0%
$\lambda_2=1.29$	$K_2=1.8$	48.7%
$\lambda_3=0.879$	$K_3=2.7$	23.7%
$\lambda_4=0.368$	$K_4=6.4$	8.9%
$\lambda_5=0.113$	$K_5=20.7$	2.4%

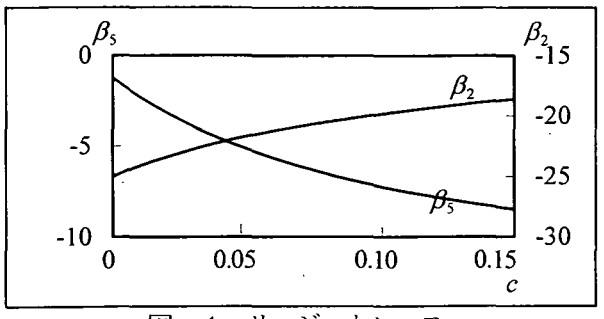
単に地価を推定する回帰モデルを作成するだけであれば、このような変数を取り除く方法が考えられる。しかしながら、社会資本整備プロジェクトによる影響分析では、政策によって直接影響を受けるこれらの変数を残したまま分析モデルを構築する必要が生じる。

そこで、多重共線性の適切化を行うために、主成分回帰（PCR）及びリッジ回帰（RR）の適用した。なお、2.2 において述べたように、リッジ回帰の適用に際しては、各説明変数 x_i は平均 *0* 分散 *1* となるように基準化を行ったが、結果については他との比較のため、式(13)～(15)の基準化とは逆の変換（詳細は蓑谷 [9]を参照されたい）を行ったものを記載している。また、リッジ・パラメータの同定に関しては、4.2.(4)で説明した GCV による基準を用いて $c=0.082$ とした（図－3）。



図－3　GCV における $V(c)$の推移

図－4にはリッジ・トレースを示した。GCV による最適なパラメータ値では、パラメータ値の変化は未だ収束しておらず、リッジ・トレースによってリッジ・パラメータを同定する方法では、過大にバイアスを与える可能性を示唆している。



図－4　リッジ・トレース

ところで、GCV は適切化パラメータの推定法として最も有名なものの一つであるが、GCV が適切化パラメータに対して鈍感な場合もあり、その有効性への疑問が指摘されることも少なくない [29]。しかし、本論文で対象としたリッジパラメータの選定においては、図－3を見る限りそのような問題は生じていないことが確認される。

パラメータ推定値とその推定精度は表－3,4に示すとおりである。変数間に強い相関のあった x_2 と x_5 に関わるパラメータ推定値が、適切化によって大きく変化し、同時に推定値の標準偏差が小さくなっていることが分かる。また、たとえば政策に直接関わるパラメータ β_2（最寄り駅までの距離）については、主成分回帰による推定結果はリッジ回帰によるそれに比べて絶対値で約3割も値が小さくなっている。このような結果の相異は、政策分析の結論に重大な影響を及ぼす可能性がある。

表－3　各手法によるパラメータの推定値

	パラメータ推定値		
	OLS	PCR	RR
β_0	3.11×10^5	2.99×10^5	3.09×10^5
β_1	2.04×10^2	2.41×10^2	2.04×10^2
β_2	**-25.0**	**-14.9**	**-20.4**
β_3	-2.49×10^3	-2.60×10^3	-2.36×10^3
β_4	2.32×10^2	2.59×10^2	2.29×10^2
β_5	**-1.22**	**-13.75**	**-6.61**

表－4　推定パラメータの標準誤差と t 値

	左：標準偏差　右：t 値					
	OLS		PCR		RR	
β_0	1.77×10^4	17.6	1.58×10^4	18.9	1.59×10^4	20.0
β_1	74.1	2.8	70.5	3.4	62.3	3.3
β_2	**7.14**	**-3.5**	**1.74**	**-8.6**	**4.35**	**-4.7**
β_3	1.22×10^2	-2.0	1.24×10^2	-2.1	1.05×10^2	-2.3
β_4	47.3	**4.9**	44.3	**5.8**	41.4	**5.5**
β_5	**9.25**	**-0.1**	**3.59**	**-3.8**	**5.90**	**-1.1**

そこで、パラメータの推定方法による推定値の違いが、実際の政策分析にどの程度の影響を及ぼすかについて考察を行うため、対象地域の北西に新たな駅ができたこと

を想定した場合の土地（住宅地）資産価値上昇額を試算した。OLS、PCR、RR の各手法によるパラメータを用いた試算価値上昇額（及びその標準偏差）は、それぞれ、800 億（±450 億）、550 億（±400 億）、650 億（±400 億）円程度であり、予測値に関しては２～３割程度の違いが生じる。ただしここでは、各標準地が代表するエリアの面積は大雑把な方法で求めたものであるため、評価値自体にはそれ程意味はない。また、このような計算を便益計測手法として用いる際には、満足すべき細かな仮定について別途吟味を要することは言うまでもない。しかし、説明変数の変化等については同一の条件下で計算を行っているため、各手法の概算値の相対的大小関係やその比にはある程度意味があると考えられる。

これらことから、実際の政策分析において解くべき問題の非適切性を検討すること、さらに非適切が存在する場合にどのような適切化手法を用いるかということは、政策判断に重要な影響を及ぼし得ることが示唆される。

6. おわりに

本論文では、社会資本整備プロジェクトの影響分析において非常に重要な要因である地価について、現況データからその決定構造を探るという逆問題を取り扱った。

最も基本的な逆解析手法である回帰分析を例に、しばしば生じる多重共線性という非適切性について、第２章で簡単に説明した。次に、第３章では、多重共線性の適切化手法のうち主成分回帰とリッジ回帰について、数理科学の分野における典型的な適切化手法との対応関係を整理して示した。次に、第４章では、リッジ回帰における適切化パラメータの同定方法に関する既存研究を、理論の歴史的発展経緯に即して整理し示した。

無論、第３章の内容は、各手法の本質まで十分理解している者にとっては、おおよそ自明な内容とも思われる。しかし、学問分野の細分化・専門化が進むとともに、それらのことを見落としがちになるのも事実である。実際、本論文で扱うリッジ回帰は、非適切逆問題に対する典型的な適切化手法と解釈可能であるにもかかわらず、統計学以外の分野ではその名前はあまり知られていない。一方、計量経済学の分野では、リッジ回帰をアドホックな手法と見る向きもある[30]。

逆解析の意義の一つは、様々な分野で用いられ一見違って見える手法を、逆解析という視点から再度眺めることにより、実は本質的に同じであるということを理解することにあると考える。そのような意味において、多重共線性とその適切化手法は、行解析の意義を理解する格好の題材であると思われる。

最後に第５章では、多重共線性が存在する地価の回帰分析において、各適切化手法の適用がモデルの推定あるいは実際の政策判断にどのような影響を与え得るかについて実証を行った。その結果、これらの手法の選択が、社会資本整備プロジェクトの影響分析上無視しえない影響を及ぼす可能性を確認した。

なお、本論文では扱わなかったが、Baye and Parker によって式(43)に示す *r-k* 推定量と呼ばれるものが提案されている（原論文中では、c の代りに k が用いられている）[30]。

$$\beta_c(c) \equiv P_r(P_r'X'XP_r + cI)^{-1}P_r'X'y \qquad (43)$$

この *r-k* 推定量は、通常最小二乗、主成分回帰及びリッジ回帰による各推定量を、その特殊ケースとして包含する推定量である[31]。

$$\beta_{m+1}(0) = (X'X)^{-1}X'y = \beta_O \qquad (44)$$

$$\beta_l(0) = P_L(P_L'X'XP_L)^{-1}P_L'X'y = \beta_S \qquad (45)$$

$$\beta_{m+1}(c) = (X'X + cI)^{-1}X'y = \beta_R \qquad (46)$$

r-k 推定量が実際の適切化手法としてどれ程の意義を持つものかについては、必ずしも明確ではないが、逆解析手法の中でどのように位置づけられるかについては興味深い問題であり、今後取り組んでみたい。

謝辞

リッジ回帰が Tikhonov の適切化と解釈可能であることに関して、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) Hocking, R. R. : Developments in Linear Regression Methodology : 1959-1982, Technometrics, Vol. 25, No.3, pp.219-230, 1983.
2) Draper, N. R. and Nostrand, R. C. V. : Ridge Regression and James-Stein Estimation : Review and Comments, Technometrics, Vol. 21, No.4, pp.451-465, 1979.
3) 竹内啓編集代表：統計学辞典，東洋経済新報社，1989.
4) 森杉壽芳編著：社会資本整備の便益評価　一般均衡理論によるアプローチ，勁草書房，1997.
5) 金本良嗣：ヘドニック・アプローチによる便益評価と理論的基礎，土木学会論文集，No.449／IV-17，pp.47-56，土木学会，1992.
6) Golub, G. H., Heath, M. and Wahba, A, G. : Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter, Technometrics, Vol. 21, No.2, pp.215-223, 1979.
7) 本城勇介：逆解析における事前情報とモデルの選択，「講座　地盤工学における逆解析　第５章」，土と基礎, Vol. 43, No. 7, pp.63-68, Vol. 43, No. 8, pp.51-56, 1995.
8) Maddala, G. S. : Introduction to Econometrics (Second Edition), Prentice-Hall, 1988［和合肇訳：計量経済分析の方法，シーエーピー出版，1996］.

9) 蓑谷千凰彦：計量経済学の新しい展開，多賀出版，1992.
10) 逆解析の地盤工学への適用に関する研究委員会：委員会報告　第1章　基礎理論ワーキンググループ(WG1)報告，地盤工学における逆解析の適用と施工管理に関するシンポジウム　発表論文集，地盤工学会，1997.
11) 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.
12) 内山敏典・杉野元亮：消費構造の変容とその統計的分析，晃洋書房，1995.
13) Hoerl, A. E. and Kennard, R. W. : Ridge Regression : Biased Estimation for Nonorthogonal Problems , Technometrics, Vol. 12, No.1, pp.55-67, 1970.
14) Hoerl, A. E. and Kennard, R. W. : Ridge Regression : Applications to Nonorthogonal Problems, Technometrics, Vol..12, No.1, pp.69-82, 1970.
15) Hoerl, A. E. : Application of ridge analysis to regression problems, Chemical Engineering Progress, Vol. 58, No.3, pp. 54-59, 1962.
16) Hoerl, A. E. : Optimum solution of many variables equations, Chemical Engineering Progress, Vol. 55, No.11, pp. 69-78, 1959.
17) 久保司郎：逆問題，培風館，1992.
18) 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.
19) 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］.
20) Marquardt, D. W. :Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation, Technometrics, Vol. 12, No.3, pp.591-612, 1970.
21) Menke, M. : Geophysical Data Analysis : Discrete Inverse Theory Revised Edition, Academic Press, 1989. ［柳谷俊・塚田和彦　訳：離散インバース理論　逆問題とデータ解析，古今書院，1997］
22) Levenberg, K. : Method for the Solution of Certain Non-linear Problems in Least Squares, Quarterly of Applied Mathematics, Vol. 2, pp.164-168, 1944.
23) 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.
24) Marquardt, D. W. : An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics, Vol. 11, No. 2, pp.431-441, 1963.
25) Allen, D. M. : Mean Square Error of Prediction as a Criterion for Selection Variables , Technometrics, Vol. 13, pp.469-475, 1971.
26) Allen, D. M. : The Relationship Between Variable Selection and Data Agumentation and a Method for Prediction, Technometrics, Vol. 16, pp.215-223, 1974.
27) Mallows, C. L. : Some Comments on C_p , Technometrics, Vol. 15, No.4, pp.661-675, 1973.
28) Wahba, G. : A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for Solving Them, Applications of Statistics, ed. by Krishnaiah, P. R., pp.507-523, 1977.
29) 細田陽介・北川高嗣：不適切問題に対する MAICE-DP 法による最適正則化法について，日本応用数理学会論文誌，pp.47-58，Vol.3，No.2，1993.
30) Judge, G. G., Griffiths, W. E,, Hill, R. C., Lutkepohl, H. and Lee, T. C. : The Thepry and Practice of Econometrics, John Wiley and Sons., 1985.
31) Baye, M. R. and Parker, D. F. : Combining Ridge and Principal Component Regression : A Money Demand Illustration, Communications in Statistics A - Theory and Methods, Vol. 13, pp. 197-205, 1973.
32) Sarkar, N. : Comparisons among Some Estimators in Misspecified Linear Models with Multicollinearity, Annals of the Institute of Statistical Mathematics, Vol. 41, pp. 717-724, 1989.

（1998年4月24日 受付）