

尤度距離におけるリッジパラメータの 推定量の選定方法

竹内 秀一

A Procedure of Selecting the Estimator of the Ridge Parameter in Likelihood Distance

Hidekazu TAKEUCHI

Many influence measures have been proposed for the purpose of assessing the influence of observations in linear regression. Among these, the likelihood distance based on the log-likelihood is discussed and issues regarding its application to ridge regression are considered. One important issue is that an appropriate procedure to take account of influence measures in ridge regression has not been clearly laid out. Since assessment of the influence of individual observations differs according to the kind of estimators of the ridge parameter, it is necessary to predetermine an appropriate estimator of the ridge parameter. In this paper, the type of evaluation criterion on which to base the selection of the estimator of the ridge parameter is examined, so as to take account of the influence of observations using likelihood distance in ridge regression. New estimators of the ridge parameter are also proposed and a numerical comparison with a real data set to indicate the effectiveness of the estimator is performed.

1 はじめに

線形回帰分析における観測値の影響力評価を, Hoerl and Kennard[4] によって提案されたリッジ回帰 (ridge regression) へ応用した場合の課題について考える. 線形回帰分析における観測値の影響力評価を行なうための診断統計量は数多くあるが, 本論文では, 対数尤度規準に基づく診断統計量 (influence measure) である尤度距離 (likelihood distance) に対してリッジ回帰を適用する場合を取り上げる (竹内 [13][14] を参照). リッジ回帰における診断統計量による影響力評価において, 大きな課題の一つが, リッジパラメータの選定方法が明確になっていないという点である. どのようなリッジパラメータを適用するのかにより, 個々の観測値の影響力評価が異なるため, 適切な評価規準に基づくリッジパラメータの選定方法を定めておく必要がある.

これまでにも, リッジパラメータの代表的な推定量が, Hoerl, Kennard and Baldwin[5] や Lawless and Wang[7] によって提案されている. ただ, これらの推定量は, ある評価規準では過大な値をとりやすいという性質があるため, 近年, 縮小推定量がいくつか提案されている. たとえば,

Strawderman[10] や Casella[1] の提案する推定量, これを改良 (縮小) したものととして Shinozaki and Chang[8][9] の提案する推定量, さらに別の視点から改良を加えた Kubokawa and Srivastava[6] の推定量などがある.

他方, これとは別な視点から, 竹内 [14] において, リッジ回帰における観測値除去法に基づく診断統計量として尤度距離を導出し, その性質を調べるためにリッジパラメータに関して偏微分をしてその導関数を求め, リッジパラメータによる観測値の影響力変化を数値的に確認している.

本論文では, リッジ回帰における尤度距離を利用して観測値の影響力を調べる場合に, どのような評価規準を基にリッジパラメータを選定するのかについて検討し, その下でリッジパラメータの新たな推定量を提案する. また, リッジパラメータの新たな推定量の有効性を示すために, 実際のデータについて数値的な比較検討も行い, 代替的な推定量として適していることを示す.

本論文の構成は以下のとおりである. 第2節では基本的な各種の統計量およびリッジ回帰における尤度距離とその導関数などを定義する. 第3節において, リッジ回帰における尤度距離に対してある評価規準を示してリッジパラメータの選定方法を与え, リッジパラメータの新たな推定量を提案する. 第4節では, リッジパラメータの新たな推定量を実際のデータに基づく数値例に対して適用し, その有効性について検討する. 第5節は全体のまとめと今後の課題である.

2 定義

本節では, 線形回帰モデルにおけるリッジ推定量を導入し, リッジ回帰における診断統計量として尤度距離およびその導関数の定義をする. また, 尤度距離と関連する診断統計量である Cook の距離についても定義しておく.

2.1 線形回帰およびリッジ回帰

ここでは, 線形回帰モデルとして,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

を考える. このとき, \mathbf{y} は $n \times 1$ の目的変数ベクトル, \mathbf{X} は $n \times p$ のフルランクの説明変数行列, $\boldsymbol{\beta}$ は $p \times 1$ の回帰係数ベクトル, そして $\boldsymbol{\varepsilon}$ は $n \times 1$ の誤差ベクトルであり, 正規分布 $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ に従うものとする. ただし, \mathbf{I}_n は n 次の単位行列を表す. また, $\boldsymbol{\beta}$ の最小 2 乗推定量は $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ として得られ, σ^2 の不偏推定量は $\hat{\sigma}^2 = \mathbf{e}'\mathbf{e}/(n-p)$ となる. ただし, 「 \prime 」は行列あるいはベクトルの転置を表し, \mathbf{e} は残差ベクトルであり, $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$ である. このとき, \mathbf{H} は説明変数行列 \mathbf{X} から構成されるハット行列 (hat matrix) $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ であり, その第 i 対角成分 h_{ii} がてこ比である. このてこ比については, $1/n \leq h_{ii} < 1$ とする. さらに, 残差ベクトル \mathbf{e} の第 i 成分 e_i を標準化した $t_i = e_i/(\hat{\sigma}\sqrt{1-h_{ii}})$ を標準化残差 (内的スチューデント化残差) とし, t_i

を第 i 成分とする標準化残差ベクトルを

$$\mathbf{t} = \frac{1}{\hat{\sigma}} [\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-\frac{1}{2}} \mathbf{e}$$

とする。ただし、 $\text{diag}(\mathbf{A})$ は正方行列 \mathbf{A} の対角成分のみを取り出し、非対角成分をすべて 0 にした行列を表す。

線形回帰の一つの代替的方法としてリッジ回帰がある (Hoerl and Kennard[4] や Groβ[3] などを参照)。リッジ回帰における回帰係数ベクトル β の推定量 (以下、リッジ推定量) を、リッジパラメータ $k (\geq 0)$ により、

$$\hat{\beta}^R \equiv (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}$$

と定義する。すると、最小 2 乗推定量の場合と同じく、残差ベクトル \mathbf{e}^R は、 $\mathbf{e}^R = \mathbf{y} - \mathbf{X}\hat{\beta}^R = (\mathbf{I}_n - \mathbf{H}^R)\mathbf{y}$ となる。ただし、リッジ回帰におけるハット行列 \mathbf{H}^R は $\mathbf{H}^R = \mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'$ であり、その第 i 対角成分 h_{ii}^R がリッジ回帰における t 比である。このとき、 $0 < h_{ii}^R < 1$ である。

また、第 i 番目の観測値を除去したときの回帰係数ベクトル β の最小 2 乗推定量は

$$\hat{\beta}_{(i)} = (\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)}$$

と定義される。ただし、添字の (\cdot) は n 個の観測値の中から除去される観測値の番号を表す。通常の最小 2 乗推定量の場合と同様に、第 i 番目の観測値を除去したときのリッジ推定量は

$$\hat{\beta}_{(i)}^R = (\mathbf{X}'_{(i)}\mathbf{X}_{(i)} + k\mathbf{I}_p)^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)}$$

となる。このとき、観測値除去に関してリッジパラメータ k は一定 (独立) であると仮定するが、 k についての挙動を調べる場合は観測値を固定した上である種の変数として扱う。 σ^2 の不偏推定量 $\hat{\sigma}^2$ の場合についても、第 i 番目の観測値を除去したときの推定量 $\hat{\sigma}_{(i)}^2$ を

$$\hat{\sigma}_{(i)}^2 = \frac{\mathbf{y}'_{(i)}(\mathbf{I}_n - \mathbf{H}_{(i)})\mathbf{y}_{(i)}}{n - p - 1} = \frac{\mathbf{y}'_{(i)}\mathbf{y}_{(i)} - \mathbf{y}'_{(i)}\mathbf{X}_{(i)}(\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1}\mathbf{X}'_{(i)}\mathbf{y}_{(i)}}{n - p - 1} = \frac{n - p - t_i^2}{n - p - 1} \hat{\sigma}^2$$

と定義する。

特に、 $k = 0$ とすれば、 $\hat{\beta}^R = \hat{\beta}$ 、 $\hat{\beta}_{(i)}^R = \hat{\beta}_{(i)}$ 、あるいは $\mathbf{H}^R = \mathbf{H}$ など、リッジ回帰の統計量が通常の線形回帰の統計量と一致することがわかる。

2.2 リッジ回帰における診断統計量

リッジ回帰における代表的な診断統計量として、Cook の距離および尤度距離を与えておく (Cook and Weisberg[2] を参照)。また、リッジ回帰における尤度距離については、リッジパラメータ k に関して偏微分した結果である導関数を与え、 $0 < k \leq 1$ の領域について検討する。

リッジ回帰における Cook の距離 CD_i^R は、Takeuchi[11] によって提案された通常の線形回帰に

尤度距離におけるリッジパラメータの推定量の選定方法

における Cook の距離を拡張することにより導入され、

$$CD_i^R \equiv \frac{(\hat{\beta}^R - \hat{\beta}_{(i)}^R)'(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)(\hat{\beta}^R - \hat{\beta}_{(i)}^R)}{p\hat{\sigma}^2} = \{s_i(c_i + d_i)\}^2 \quad (2.1)$$

と与えられる (Takeuchi[12] を参照). ただし, c_i は

$$\mathbf{c} = \frac{1}{\sqrt{p}}\mathbf{\Pi}^{\frac{1}{2}}\mathbf{t}$$

の第 i 成分であり, このとき,

$$\mathbf{\Pi} = \text{diag}(\mathbf{H})[\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-1}$$

であり, この第 i 対角成分が, $\pi_i = h_{ii}/(1 - h_{ii})$ である. つぎに, d_i は

$$\mathbf{d} = \frac{1}{\sqrt{p}}\mathbf{\Pi}^{\frac{1}{2}}\mathbf{u}$$

の第 i 成分であり, このとき,

$$\mathbf{u} = \frac{1}{\hat{\sigma}}[\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-\frac{1}{2}}(\mathbf{e}^R - \mathbf{e})$$

であり, この第 i 成分が u_i である. 最後に, s_i は

$$\mathbf{S} = \text{diag}(\mathbf{I}_n - \mathbf{H})[\text{diag}(\mathbf{I}_n - \mathbf{H}^R)]^{-1}$$

の第 i 対角成分であり, $s_i = (1 - h_{ii})/(1 - h_{ii}^R)$ である. 特に, $k = 0$ のとき, 竹内 [14] などから

$$CD_i^R \Big|_{k=0} = CD_i = c_i^2$$

とする.

(2.1) 式で与えられるリッジ回帰における Cook の距離 CD_i^R においては, $\hat{\beta}^R - \hat{\beta}_{(i)}^R$ の中に挟まれる行列の選び方によっていくつかの定義式が考えられるが, ここでは, $\hat{\beta}^R$ の分散共分散行列 $\text{Var}(\hat{\beta}^R) = \sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}$ の逆行列を選ぶものとする (たとえば, Walker and Birch[15] を参照).

つぎに, 竹内 [13] によって導出されたリッジ回帰における尤度距離 LD_i^R を以下のように与える.

$$LD_i^R \equiv 2[L(\hat{\beta}^R) - L(\hat{\beta}_{(i)}^R)] \quad (2.2)$$

ただし, 通常の線形回帰の場合と同じく,

$$L(\hat{\beta}^R) : \beta \text{ が } \hat{\beta}^R \text{ のときの対数尤度 および } L(\hat{\beta}_{(i)}^R) : \beta \text{ が } \hat{\beta}_{(i)}^R \text{ のときの対数尤度}$$

とする (具体的な対数尤度については竹内 [13] を参照). (2.2) 式の定義に基づいて, リッジ回帰における尤度距離 LD_i^R を (2.1) 式の Cook の距離 CD_i^R を利用して表現すると,

$$\begin{aligned} \text{LD}_i^R = n \log T_i + \frac{1}{T_i} \cdot \frac{n}{n-p} \left[\frac{h_{ii}^{R2}}{h_{ii}} p \text{CD}_i^R + 2s_i(t_i + u_i)u_i^* \right] \\ + n \left(\frac{1}{T_i} - 1 \right) \left[1 + \frac{\mathbf{u}' \text{diag}(\mathbf{I}_n - \mathbf{H})\mathbf{u}}{n-p} \right] \end{aligned} \quad (2.3)$$

となる。ただし、

$$T_i = \frac{n}{n-1} \cdot \frac{n-p-t_i^2}{n-p}$$

であり、 h_{ii}^{R2} は $\mathbf{H}^{R2} = (\mathbf{H}^R)^2$ の第 i 対角成分であり、それに u_i^* は

$$\mathbf{u}^* = \frac{1}{\hat{\sigma}} [\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-\frac{1}{2}} (\mathbf{e}^{R2} - \mathbf{e}^R)$$

の第 i 成分である。ここで、 $\mathbf{e}^{R2} = (\mathbf{I}_n - \mathbf{H}^{R2})\mathbf{y}$ である。

さらに、(2.3) 式を変形すると

$$\text{LD}_i^R = n \log T_i + \frac{1}{T_i} \cdot \frac{np}{n-p} \text{CD}_i^{R*} + n \left(\frac{1}{T_i} - 1 \right) \left[1 + \frac{\mathbf{u}' \text{diag}(\mathbf{I}_n - \mathbf{H})\mathbf{u}}{n-p} \right] \quad (2.4)$$

と表現することもできる。ただし、(2.4) 式において

$$\text{CD}_i^{R*} \equiv \frac{\pi_i^*}{\pi_i} \left[\left\{ s_i(c_i + d_i) + \frac{d_i^*}{\pi_i^*} \right\}^2 - \left(\frac{d_i^*}{\pi_i^*} \right)^2 \right] \quad (2.5)$$

とする。このとき、(2.5) 式において、 d_i^* は

$$\mathbf{d}^* = \frac{1}{\sqrt{p}} \mathbf{\Pi}^{\frac{1}{2}} \mathbf{u}^*$$

の第 i 成分であり、また π_i^* は

$$\mathbf{\Pi}^* = \text{diag}\{(\mathbf{H}^R)^2\} [\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-1}$$

の第 i 対角成分であり、 $\pi_i^* = h_{ii}^{R2}/(1 - h_{ii})$ である。

特に、 $k=0$ のとき、リッジ回帰における尤度距離 LD_i^R は、竹内 [13] から

$$\text{LD}_i^R \Big|_{k=0} = \text{LD}_i = n \log T_i + \frac{1}{T_i} \cdot \frac{np}{n-p} \text{CD}_i + n \left(\frac{1}{T_i} - 1 \right)$$

となり、通常の線形回帰における尤度距離 LD_i に一致する。

最後に、リッジ回帰における尤度距離 LD_i^R のリッジパラメータ k に関する導関数を与える。導関数 DL_i は、竹内 [14] から $k > 0$ の領域について

$$\begin{aligned} \text{DL}_i \equiv \frac{\partial}{\partial k} \text{LD}_i^R = \frac{2}{k} \cdot \frac{n}{n-p} \left[\frac{p}{T_i} \cdot \frac{s_i}{\pi_i} \left\{ \left(\frac{1}{s_i s_i^*} - \frac{1}{s_i s_i^{**}} + \pi_i^* \left(\frac{1}{s_i} - \frac{1}{s_i^*} \right) \right) \text{CD}_i^R \right. \right. \\ \left. \left. + (c_i + d_i) \left(s_i \left(\pi_i^* - \frac{1}{s_i^*} \right) d_i^* + 2d_i^{**} \right) + (d_i^*)^2 \right\} + \left(\frac{1}{T_i} - 1 \right) \mathbf{u}' \text{diag}(\mathbf{I}_n - \mathbf{H})\mathbf{u}^* \right] \end{aligned} \quad (2.6)$$

と得られている。ただし、 s_i^* は

尤度距離におけるリッジパラメータの推定量の選定方法

$$\mathbf{S}^* = \text{diag}(\mathbf{I}_n - \mathbf{H})[\text{diag}(\mathbf{I}_n - \mathbf{H}^{R2})]^{-1}$$

の第 i 対角成分であり, $s_i^* = (1 - h_{ii})/(1 - h_{ii}^{R2})$ である. また, s_i^{**} は

$$\mathbf{S}^{**} = \text{diag}(\mathbf{I}_n - \mathbf{H})[\text{diag}(\mathbf{I}_n - \mathbf{H}^{R3})]^{-1}$$

の第 i 対角成分であり, $s_i^{**} = (1 - h_{ii})/(1 - h_{ii}^{R3})$ である. このとき, h_{ii}^{R3} は $\mathbf{H}^{R3} = (\mathbf{H}^R)^3$ の第 i 対角成分である. さらに, d_i^{**} は

$$\mathbf{d}^{**} = \frac{1}{\sqrt{p}} \mathbf{\Pi}^{\frac{1}{2}} \mathbf{u}^{**}$$

の第 i 成分であり, このとき,

$$\mathbf{u}^{**} = \frac{1}{\hat{\sigma}} [\text{diag}(\mathbf{I}_n - \mathbf{H})]^{-\frac{1}{2}} (\mathbf{e}^{R3} - \mathbf{e}^{R2})$$

である. ここで, $\mathbf{e}^{R3} = (\mathbf{I}_n - \mathbf{H}^{R3})\mathbf{y}$ である.

3 リッジパラメータの推定量および選定方法

リッジ推定量に対してよく利用される代表的なリッジパラメータ k は以下の2つである. 1つは, Hoerl, Kennard and Baldwin[5] によって提案されている $k_1 = p\hat{\sigma}^2 / \hat{\beta}'\hat{\beta}$ であり, もう1つは Lawless and Wang[7] によって提案されている $k_2 = p\hat{\sigma}^2 / \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}$ である. この他にも Shinozaki and Chang[8][9] や Kubokawa and Srivastava[6] によりリッジパラメータに関する経験ベイズ推定量がいくつかの評価規準に基づき提案されている. けれども, 第4節の数値例で示されるように, 尤度距離 LD_i^R を利用した場合には, これまで提案されているリッジパラメータの推定値が小さく評価される傾向にある. そこで, リッジパラメータの推定値が比較的大きくなる k_1 を改良することにより, リッジパラメータの新たな推定量を提案する.

新たな推定量を検討するための素材として, (2.6) 式における尤度距離の導関数 DL_i を, リッジパラメータ k による陽な形式に直してみる. 付録から (2.6) 式は

$$\begin{aligned} \text{DL}_i = & \frac{2}{k} \cdot \frac{n}{n-p} \left[\frac{p}{T_i \pi_i} \cdot \frac{1 - h_{ii}}{1 - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij}^2} \left\{ \left(\sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^2 \ell_{ij}^2 \left(\sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^2 \ell_{ij}^2 - 1 \right) \right. \right. \right. \\ & \left. \left. \left. + \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^3 \ell_{ij}^2 \left(1 - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij}^2 \right) \right\} \left\{ \frac{c_i + \frac{\sqrt{\pi_i}}{\hat{\sigma}\sqrt{p}\sqrt{1-h_{ii}}}(y_i - e_i - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij} m_j)}{1 - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij}^2} \right\}^2 \right. \\ & \left. + \left\{ c_i + \frac{\sqrt{\pi_i}}{\hat{\sigma}\sqrt{p}\sqrt{1-h_{ii}}}(y_i - e_i - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij} m_j) \right\} \frac{k\sqrt{\pi_i}}{\hat{\sigma}\sqrt{p}\sqrt{1-h_{ii}}} \times \right. \end{aligned}$$

$$\begin{aligned} & \times \left(\frac{2 \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^2 \ell_{ij}^2 - 1}{1 - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij}^2} \sum_{j=1}^p \frac{\delta_j}{(\delta_j + k)^2} \ell_{ij} m_j + 2 \sum_{j=1}^p \frac{\delta_j^2}{(\delta_j + k)^3} \ell_{ij} m_j \right) \\ & + \frac{k^2 \pi_i}{\delta^2 p (1 - h_{ii})} \left(\sum_{j=1}^p \frac{\delta_j}{(\delta_j + k)^2} \ell_{ij} m_j \right)^2 \left. \right\} + \left(\frac{1}{T_i} - 1 \right) \frac{k^2}{\delta^2} \sum_{j=1}^p \frac{\delta_j}{(\delta_j + k)^3} m_j^2 \quad (3.1) \end{aligned}$$

と変形することができる。ただし、 $\delta_j (> 0)$ は、説明変数行列 \mathbf{X} を

$$\mathbf{X} = \mathbf{L} \mathbf{G}^{\frac{1}{2}} \mathbf{U}'$$

と特異値分解することにより得られる $\mathbf{X}'\mathbf{X}$ の第 j 固有値であり、 ℓ_{ij} は \mathbf{L} の第 (i, j) 成分であり、 m_j は $\mathbf{m} = \mathbf{L}'\mathbf{y}$ の第 j 成分を表す (詳細な定義や条件については付録を参照)。本来は、「(3.1) 式 = 0」として、 k についての方程式を解くことができれば問題ないが、この方程式は容易に解くことができない。しかしながら、(3.1) 式に基づいて最適な k を得るためには、 δ_j や m_j との関連性が強いことを示唆していることは容易にわかる。

この関連性を基に、 k の新たな推定量を考案する。竹内 [14] の数値例から、 k の値としては k_1 よりも大きな値になるものが望ましいと考えられる。そこで、 k_1 を改良する推定量を以下のように提案する。

k_1 の定義式において、分子の p および δ^2 を定数とみなし、分母の $\hat{\beta}'\hat{\beta}$ について δ_j および m_j を使って表現すると

$$\begin{aligned} \hat{\beta}'\hat{\beta} &= \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-2}\mathbf{X}'\mathbf{y} \\ &= \mathbf{y}'\mathbf{L}\mathbf{G}^{\frac{1}{2}}\mathbf{U}'(\mathbf{U}\mathbf{G}\mathbf{U}')^{-2}\mathbf{U}\mathbf{G}^{\frac{1}{2}}\mathbf{L}'\mathbf{y} = \mathbf{y}'\mathbf{L}\mathbf{G}^{-1}\mathbf{L}'\mathbf{y} = \mathbf{m}'\mathbf{G}^{-1}\mathbf{m} = \sum_{j=1}^p \frac{m_j^2}{\delta_j} \end{aligned}$$

となるので、 k_1 は、

$$k_1 = \frac{p\delta^2}{\sum_{j=1}^p \frac{m_j^2}{\delta_j}}$$

と表される。このとき、 k_1 の値を大きくするためには、分母の分母である δ_j を大きいものと置き換えればよいことがわかる。そこで、 k の新たな推定量として、 p 個の固有値 δ_j をすべて同じ固有値 δ_{max} に置き換えると、

$$k^* = \frac{p\delta^2}{\sum_{j=1}^p \frac{m_j^2}{\delta_{max}}} = \delta_{max} \frac{p\delta^2}{\sum_{j=1}^p m_j^2} \quad (3.2)$$

となる。ただし、 δ_{max} は k^* が $0 < k^* \leq 1$ を満たすものの中で最大の固有値 δ_j とする。

(3.2) 式の新たな推定量を算出する場合に注意すべき点は、通常のリッジパラメータにおいては、変数の個数 p から定数項分を除き、かつ $\hat{\beta}$ から定数項を除くことにより推定値を求めるが、 k^*

尤度距離におけるリッジパラメータの推定量の選定方法

においては、定数項の分離が容易ではないので、定数項も含めていることである。

もう一つの推定量として、 k_1 を単純に拡張したものである

$$k^{**} = \frac{1}{n} \sum_{i=1}^n k_{1(i)} = \frac{1}{n} \sum_{i=1}^n \frac{p\hat{\sigma}_{(i)}^2}{\hat{\beta}'_{(i)}\hat{\beta}_{(i)}} \quad (3.3)$$

も利用する。(3.3)式についても(3.2)式と同じく定数項を含めて算出する。

つぎに、リッジ回帰における尤度距離 LD_i^R に基づいて、最適な k の値を見つけるための評価規準を示す。ここでは、自然な評価規準として、 LD_i^R の平均である

$$LD^R(k) = \frac{1}{n} \sum_{i=1}^n LD_i^R \quad (3.4)$$

を利用する。 LD_i^R の値は大きいと影響力が大きいと判断され、逆に小さいと影響力は小さいと判断されるので、(3.4)式の $LD^R(k)$ の値をできるだけ小さくする k を最適な k とみなすことにする。

また、(3.4)式と対比するために、第 j 番目の観測値を除去することによる交差検証(クロスバリデーション)評価規準として

$$LD_j^R(k) = \frac{1}{n-1} \sum_{\substack{i=1 \\ i \neq j}}^n LD_i^R \quad (3.5)$$

も算出する。さらに、 LD_i^R の導関数 DL_i の挙動についても、(3.4)式の $LD^R(k)$ と同様の規準として

$$DL^R(k) = \frac{1}{n} \sum_{i=1}^n DL_i \quad (3.6)$$

を算出する。(3.6)式と対比するために、交差検証評価規準として $DL_j^R(k)$ の計算をすることもできるが、観測値ごとの挙動の変化は(3.5)式と同様なパターンになるので割愛する。

4 数値例

提案するリッジパラメータ k の新たな推定量を評価するために、実際のデータに対して数値計算を行う。数値計算をするためのデータとしては、Takeuchi[12]において例示されている「Artificial Data」(データ数は $n = 20$ であり、説明変数の数は定数項を含め $p = 4 + 1 = 5$) を利用する。これは、竹内[13][14]で利用した数値例と同じものである。

(3.4)式の $LD^R(k)$ および(3.5)式の $LD_j^R(k)$ の計算結果を数値的に対比すると煩雑になるので図4.1のグラフにまとめた。図4.1において、「全観測値」が(3.4)式の $LD^R(k)$ の計算結果をグラフ化したものであり、「No.2, No.10, No.14, No.17」はそれぞれの観測値を除いたとき(つまり、 $j = 2, 10, 14, 17$ それぞれ)の(3.5)式の $LD_j^R(k)$ の計算結果をグラフ化したものである。この他の16個の観測値については、「全観測値」と概ね重なるので省略した。なお、竹内[13]における尤度距離 LD_i^R の数値計算において、コンピュータプログラムのミスによる誤りがあったため今回の計算結果において修正をしている(グラフの形状に大きな違いはないが、縦軸のスケールが1桁程度

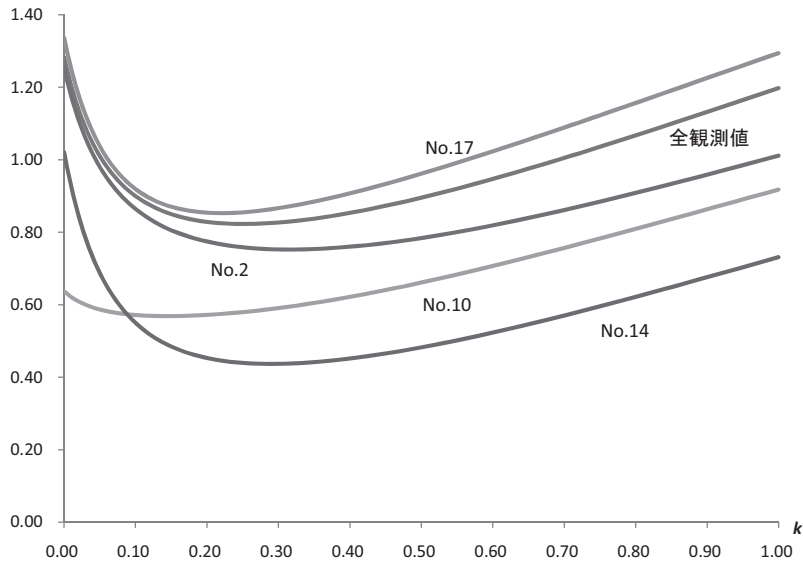


図 4.1 k の変化に基づく $LD^R(k)$ および $LD_j^R(k)$ のグラフ (一部抜粋)

異なる).

図 4.1 のグラフから $LD^R(k)$ の最小値 (最適解) は $k = k_{opt} = 0.251$ (より正確には $k_{opt} = 0.2513$) のときに 0.823 となる (詳しくは表 4.1 を参照). この k_{opt} のとき, 表 4.2 に示されているように $DL^R(k)$ の値はほぼ 0 になる. けれども, 影響力が大きいと判断された 3 個の観測値 No.2, No.10 それに No.14 と, 竹内 [14] で発見された DL_i の挙動として特徴的な観測値 No.17 については, その絶対値 $|DL_i|$ の値が 2 を超えており, 他の観測値と大きく異なる変化を示している. また, 観測値 No.2 と No.10 の組および観測値 No.14 と No.17 の組の変化量が, それぞれの組の中で相殺される関係が見られる点も特徴的である (表 4.1 の計算結果からは, このような変化を読み取ることはできない).

(3.4) 式に基づく評価規準によれば, 最適な $k = k_{opt}$ は k_1 や k_2 よりもかなり大きな値になることがわかる. そこで新たに提案する k^* や k^{**} を適用した場合との比較を試みる. 固有値は大きいものから,

$$\delta_1 = 19.004, \delta_2 = 3.082, \delta_3 = 0.441, \delta_4 = 0.221, \delta_5 = 0.185$$

となるが, $\delta_{max} = \delta_1$ とすると, $k^* = 1.603$ となり, k の値が 1 を超えてしまう. そこで, 次善の策として $\delta_{max} = \delta_2$ とすると, $k^* = 0.252$ となり数値計算により得られた最適な k の値 (k_{opt}) に近いものが得られた. 表 4.1 の $LD^R(k)$ の計算結果からわかるように, 提案する k^* や k^{**} は従来から利用されている k_1 や k_2 よりも最適な k_{opt} に近いことがわかる. また, 表 4.2 の $DL^R(k)$ の計算結果からも, 提案する k^* や k^{**} の方が従来から利用されている k_1 や k_2 よりも最適な k_{opt} にか

表 4.1 $LD^R(k)$ の計算結果

No.	LD_i^R					
	$k = 0$ (LD_i)	$k = k_2$ = 0.073	$k = k_1$ = 0.109	$k = k^{**}$ = 0.248	$k = k^*$ = 0.252	$k = k_{opt}$ = 0.251
1	0.940	-0.006	-0.173	-0.299	-0.297	-0.298
2	1.842	1.554	1.597	2.043	2.059	2.054
3	0.220	0.012	-0.040	-0.127	-0.129	-0.128
4	0.223	-0.004	-0.081	-0.244	-0.247	-0.246
5	0.069	0.163	0.197	0.319	0.322	0.321
6	0.619	0.630	0.616	0.571	0.570	0.570
7	0.433	-0.085	-0.155	-0.115	-0.110	-0.112
8	0.032	0.081	0.080	0.009	0.007	0.007
9	0.029	0.023	0.035	0.071	0.071	0.071
10	13.537	8.003	6.912	5.459	5.443	5.448
11	0.247	0.034	0.017	0.052	0.054	0.054
12	0.047	-0.010	-0.032	-0.100	-0.102	-0.102
13	0.079	0.046	0.014	-0.098	-0.101	-0.100
14	6.251	7.356	7.608	8.108	8.119	8.116
15	0.028	0.048	0.043	-0.007	-0.009	-0.009
16	0.273	0.418	0.460	0.569	0.572	0.571
17	0.246	0.533	0.531	0.219	0.207	0.211
18	0.281	0.214	0.187	0.125	0.124	0.125
19	0.092	0.089	0.081	0.050	0.050	0.050
20	0.156	-0.120	-0.148	-0.141	-0.140	-0.141
$LD^R(k)$	1.282	0.949	0.887	0.823	0.823	0.823
小数第 6 位まで	1.282258	0.948899	0.887489	0.823194	0.823180	0.823177

なり近いことがわかる。

さらに、 k^* および k^{**} を比較すると、 k^* の方が k^{**} よりも最適な k_{opt} に近いので優れているといえる。しかしながら、固有値の算出など数値計算上の手間を含めて考えると、 k^{**} が k^* に比べて大きく劣るとまではいえないであろう。この数値例においても、リッジパラメータの有効数字を小数第 2 位までとすれば、 k^* および k^{**} 、それに k_{opt} の数値はすべて $k = 0.25$ となる。観測値の影響力を厳密に評価する場合は、リッジパラメータの有効数字が小数第 2 位まででは問題だが、おおよその目安となる計算結果を得られればよいという程度であれば、 k^* および k^{**} の間に大きな相違はないといえよう。また、表 4.1 および表 4.2 の評価規準の数値を見ても、観測値の影響力は k^* と k^{**} について大差はなく、どちらも最適な k_{opt} の数値にかなり近いものといえる。

表 4.2 $DL^R(k)$ の計算結果

No.	DL_i				
	$k = k_2$ = 0.073	$k = k_1$ = 0.109	$k = k^{**}$ = 0.248	$k = k^*$ = 0.252	$k = k_{opt}$ = 0.251
1	-6.211	-3.192	0.308	0.343	0.333
2	0.213	1.979	3.788	3.800	3.797
3	-1.736	-1.148	-0.324	-0.314	-0.317
4	-2.356	-1.844	-0.641	-0.616	-0.623
5	0.994	0.897	0.892	0.894	0.893
6	-0.306	-0.405	-0.195	-0.186	-0.189
7	-2.882	-1.082	1.042	1.065	1.058
8	0.161	-0.186	-0.680	-0.684	-0.683
9	0.300	0.320	0.197	0.194	0.195
10	-38.517	-22.599	-3.804	-3.600	-3.659
11	-0.828	-0.145	0.431	0.436	0.434
12	-0.624	-0.559	-0.441	-0.438	-0.439
13	-0.812	-0.879	-0.707	-0.701	-0.703
14	8.452	5.583	2.681	2.662	2.667
15	-0.052	-0.223	-0.435	-0.437	-0.437
16	1.286	0.983	0.708	0.706	0.707
17	0.922	-0.895	-2.776	-2.776	-2.776
18	-0.792	-0.652	-0.280	-0.273	-0.275
19	-0.193	-0.232	-0.193	-0.191	-0.191
20	-1.265	-0.403	0.204	0.205	0.205
$DL^R(k)$	-2.212	-1.234	-0.011	0.004	0.000
小数第6位まで	-2.212230	-1.234134	-0.011275	0.004407	-0.000130

5 まとめと今後の課題

本論文では、リッジ回帰における尤度距離 LD_i^R を基にある一つの評価規準を提示し、新たなリッジパラメータ k の選定方法を提案した。加えて、リッジパラメータの新たな推定量を提案し、その有効性を実際のデータに基づき数値的に検証もした。

今後の課題としては、リッジ回帰を一般化リッジ回帰(説明変数ごとに k の値が異なる場合)に拡張し、複数個のリッジパラメータについても同様の検討をすることが考えられる。また、提案したリッジパラメータの新たな推定量を他のデータに対しても適用し、どのようなデータ構造の場合に有効であるのかを調べる必要もある。

付録: (3.1) 式の導出

説明変数行列 \mathbf{X} を以下のように特異値分解する.

$$\mathbf{X} = \mathbf{L}\mathbf{G}^{\frac{1}{2}}\mathbf{U}'$$

ただし, \mathbf{G} は $p \times p$ の対角行列であり, その第 j 対角成分が $\mathbf{X}'\mathbf{X}$ の第 j 固有値 $\delta_j (> 0)$ である (固有値は大きいものから付番する, つまり $\delta_1 \geq \delta_2 \geq \dots \geq \delta_p$ とする). その第 j 固有値 δ_j に対応する固有ベクトルを第 j 列にもつ行列が \mathbf{U} であり, $p \times p$ の正方行列になり $\mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathbf{I}_p$ を満たす. また, \mathbf{L} は $n \times p$ の行列であり $\mathbf{L}'\mathbf{L} = \mathbf{I}_p$ を満たす. 加えて, 目的変数ベクトル \mathbf{y} については, $\mathbf{L}'\mathbf{y} = \mathbf{m}$ と変換し, この第 j 成分を m_j と表す.

この特異値分解の成分を利用して, リッジ回帰におけるハット行列の第 i 対角成分を以下のように表す. h_{ii}^R は $\mathbf{H}^R = \mathbf{L}(\mathbf{I}_p + k\mathbf{G}^{-1})^{-1}\mathbf{L}'$ の第 i 対角成分であるので,

$$h_{ii}^R = \mathbf{x}_i(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{x}_i' = \ell_i(\mathbf{I}_p + k\mathbf{G}^{-1})^{-1}\ell_i' = \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij}^2$$

となる. ただし, ℓ_i は \mathbf{L} の第 i 行 (ベクトル) であり, ℓ_{ij} は \mathbf{L} の第 (i, j) 成分, つまり ℓ_i の第 j 成分を表す. 同様に, $\mathbf{H}^{R2} = (\mathbf{H}^R)^2 = \mathbf{L}(\mathbf{I}_p + k\mathbf{G}^{-1})^{-2}\mathbf{L}'$ であるので, この第 i 対角成分は

$$h_{ii}^{R2} = \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^2 \ell_{ij}^2$$

と表すことができ, さらに, $\mathbf{H}^{R3} = (\mathbf{H}^R)^3 = \mathbf{L}(\mathbf{I}_p + k\mathbf{G}^{-1})^{-3}\mathbf{L}'$ であるので, この第 i 対角成分は

$$h_{ii}^{R3} = \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^3 \ell_{ij}^2$$

と表すことができる.

よって, リッジパラメータ k の関数である $s_i, s_i^*, s_i^{**}, \pi_i^*, d_i, d_i^*, d_i^{**}$, そして $\mathbf{u}'\text{diag}(\mathbf{I}_n - \mathbf{H})\mathbf{u}^*$ は以下のように表現できる.

ハット行列に関わる統計量は,

$$s_i = \frac{1 - h_{ii}}{1 - h_{ii}^R} = \frac{1 - h_{ii}}{1 - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij}^2}, \quad s_i^* = \frac{1 - h_{ii}}{1 - h_{ii}^{R2}} = \frac{1 - h_{ii}}{1 - \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^2 \ell_{ij}^2},$$

$$s_i^{**} = \frac{1 - h_{ii}}{1 - h_{ii}^{R3}} = \frac{1 - h_{ii}}{1 - \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^3 \ell_{ij}^2}, \quad \pi_i^* = \frac{h_{ii}^{R2}}{1 - h_{ii}} = \frac{\sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^2 \ell_{ij}^2}{1 - h_{ii}}$$

となる.

つぎに、残差ベクトルに関わる統計量は、

$$e^R = (\mathbf{I}_n - \mathbf{H}^R)\mathbf{y} = [\mathbf{I}_n - \mathbf{L}(\mathbf{I}_p + k\mathbf{G}^{-1})^{-1}\mathbf{L}']\mathbf{y} = \mathbf{y} - \mathbf{L}(\mathbf{I}_p + k\mathbf{G}^{-1})^{-1}\mathbf{m}$$

から

$$e_i^R = y_i - \ell_i(\mathbf{I}_p + k\mathbf{G}^{-1})^{-1}\mathbf{m} = y_i - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij} m_j$$

となり、

$$d_i = \frac{1}{\sqrt{p}} \sqrt{\pi_i} \frac{1}{\hat{\sigma} \sqrt{1 - h_{ii}}} (e_i^R - e_i) = \frac{\sqrt{\pi_i}}{\hat{\sigma} \sqrt{p} \sqrt{1 - h_{ii}}} (y_i - e_i - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij} m_j)$$

が得られる。同様に、

$$e_i^{R2} = y_i - \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^2 \ell_{ij} m_j$$

から

$$\begin{aligned} d_i^* &= \frac{1}{\sqrt{p}} \sqrt{\pi_i} \frac{1}{\hat{\sigma} \sqrt{1 - h_{ii}}} (e_i^{R2} - e_i^R) \\ &= \frac{\sqrt{\pi_i}}{\hat{\sigma} \sqrt{p} \sqrt{1 - h_{ii}}} \left\{ (y_i - \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^2 \ell_{ij} m_j) - (y_i - \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \ell_{ij} m_j) \right\} \\ &= \frac{k \sqrt{\pi_i}}{\hat{\sigma} \sqrt{p} \sqrt{1 - h_{ii}}} \sum_{j=1}^p \frac{\delta_j}{(\delta_j + k)^2} \ell_{ij} m_j \end{aligned}$$

が得られ、さらに同様にして、

$$e_i^{R3} = y_i - \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^3 \ell_{ij} m_j$$

から

$$\begin{aligned} d_i^{**} &= \frac{1}{\sqrt{p}} \sqrt{\pi_i} \frac{1}{\hat{\sigma} \sqrt{1 - h_{ii}}} (e_i^{R3} - e_i^{R2}) \\ &= \frac{\sqrt{\pi_i}}{\hat{\sigma} \sqrt{p} \sqrt{1 - h_{ii}}} \left\{ (y_i - \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^3 \ell_{ij} m_j) - (y_i - \sum_{j=1}^p \left(\frac{\delta_j}{\delta_j + k} \right)^2 \ell_{ij} m_j) \right\} \\ &= \frac{k \sqrt{\pi_i}}{\hat{\sigma} \sqrt{p} \sqrt{1 - h_{ii}}} \sum_{j=1}^p \frac{\delta_j^2}{(\delta_j + k)^3} \ell_{ij} m_j \end{aligned}$$

が得られる。

最後に、

$$\mathbf{u}' \text{diag}(\mathbf{I}_n - \mathbf{H}) \mathbf{u}^* = \frac{1}{\hat{\sigma}^2} (\mathbf{e}^R - \mathbf{e})' (\mathbf{e}^{R2} - \mathbf{e}^R)$$

となるので、

$$\begin{aligned} (\mathbf{e}^R - \mathbf{e})' (\mathbf{e}^{R2} - \mathbf{e}^R) &= \mathbf{y}' (\mathbf{H} - \mathbf{H}^R) (\mathbf{H}^R - \mathbf{H}^{R2}) \mathbf{y} \\ &= \mathbf{y}' \mathbf{L} [\mathbf{I}_p - (\mathbf{I}_p + k\mathbf{G}^{-1})^{-1}] \mathbf{L}' \mathbf{L} [(\mathbf{I}_p + k\mathbf{G}^{-1})^{-1} - (\mathbf{I}_p + k\mathbf{G}^{-1})^{-2}] \mathbf{L}' \mathbf{y} \end{aligned}$$

$$\begin{aligned}
 &= \mathbf{m}'[(\mathbf{I}_p + k\mathbf{G}^{-1})^{-1} - 2(\mathbf{I}_p + k\mathbf{G}^{-1})^{-2} + (\mathbf{I}_p + k\mathbf{G}^{-1})^{-3}]\mathbf{m} \\
 &= \sum_{j=1}^p \left\{ \frac{\delta_j}{\delta_j + k} - 2\left(\frac{\delta_j}{\delta_j + k}\right)^2 + \left(\frac{\delta_j}{\delta_j + k}\right)^3 \right\} m_j^2 \\
 &= \sum_{j=1}^p \frac{\delta_j}{\delta_j + k} \left(1 - \frac{\delta_j}{\delta_j + k}\right)^2 m_j^2 \\
 &= k^2 \sum_{j=1}^p \frac{\delta_j}{(\delta_j + k)^3} m_j^2
 \end{aligned}$$

と表すことができる。

したがって、以上の結果を (2.6) 式に代入して整理すると、(3.1) 式を得ることができる。

参考文献

- [1] Casella, G. (1980), Minimax ridge regression estimation, *The Annals of Statistics*, **8**, 1036–1056.
- [2] Cook, R.D. and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman and Hall.
- [3] Groß, J. (2003), *Linear Regression*, Berlin: Springer.
- [4] Hoerl, A. E. and Kennard, R. W. (1970), Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55–67.
- [5] Hoerl, A. E., Kennard, R. W. and Baldwin, K. F. (1975), Ridge regression: some simulations, *Communications in Statistics — Theory and Methods*, **4**, 105–123.
- [6] Kubokawa, T. and Srivastava, M. S. (2004), Improved empirical Bayes ridge regression estimators under multicollinearity, *Communications in Statistics — Theory and Methods*, **33**, 1943–1973.
- [7] Lawless, J. F. and Wang, P. (1976), A simulation study of ridge and other regression estimators, *Communications in Statistics — Theory and Methods*, **5**, 307–323.
- [8] Shinozaki, N. and Chang, Y. -T. (1993), Minimavity of empirical Bayes estimators of the means of independent normal variables with unequal variances, *Communications in Statistics — Theory and Methods*, **22**, 2147–2169.
- [9] Shinozaki, N. and Chang, Y. -T. (1996), Minimavity of empirical Bayes estimators shrinking toward the grand mean when variances are unequal, *Communications in Statistics — Theory and Methods*, **25**, 183–199.
- [10] Strawderman, W. E. (1978), Minimax adaptive generalized ridge regression estimators, *Journal of the American Statistical Association*, **73**, 623–627.
- [11] Takeuchi, H. (1991), Detecting influential observations by using a new expression of Cook's distance, *Communications in Statistics — Theory and Methods*, **20**, 261–274.
- [12] Takeuchi, H. (1994), Sensitivity analysis with an extension of Cook's distance in ridge regression, *Journal of the Japan Statistical Society*, **24**, 221–236.
- [13] 竹内秀一 (2007), リッジ回帰における尤度距離による影響力評価, 人文自然科学論集, **123** 号, 3–16.

- [14] 竹内秀一 (2008), 尤度距離におけるリッジパラメータに関する影響力, 人文自然科学論集, **125**号, 57-71.
- [15] Walker, E. and Birch, J. B. (1988), Influence measures in ridge regression, *Technometrics*, **30**, 221-227 (Correction **30**, 469-470).