

関数モデル法による発生集中量推定式の回帰係数の検討

早稲田大学 正員 中川 義 英
早稲田大学 正員 大塚 全 一
東京都 正員 迎見 隆 士

1. はじめに

本研究は多変量解析手法をもとに構築したパーソントリップの発生集中量推定式を事例にとり、回帰係数の比較検討をおこなっている。その中で特に多重共線性の問題、説明変数の選択の問題、及び推定式の安定性・妥当性の検討を主なねらいとしている。説明変数の選択基準の問題⁽¹⁾⁽²⁾、モデルの予測誤差、データ誤差と予測精度の問題、予測誤差基準によるモデルの評価、等の研究がこれまでになされてきている。これらの研究成果を踏まえつつ、社会現象を反映している説明変数相互間には、どうしても多重共線性が存在するとの立場をとり、発生集中量の回帰推定式を構築するには、どの程度までの多重共線性なら容認でき、又、多重共線性からみた説明変数の選択基準を探り出そうとする研究へつなげることが本稿の目的である。

回帰推定式を求める手法としては、重回帰、ステップワイズ回帰、主成分回帰、リッジ回帰の4手法を用いている。事例に用いた目的種別別のパーソントリップ発生集中量は、東京都市圏パーソントリップ調査（昭和43年及び53年）から得られた、東京都の市区町村別の結果である。推定式の説明変数は常驻地居住人口、工業出荷額、道路延長、小学校数など資料の得られた18変数である。説明変数相互には高い相関が存在しているものもあるが、各手法の相違を検討するため、18変数すべてを操作を加えずに入力資料として用いている。

2. 推定式の問題点

発生集中量の推定式を求める手法としては、原単位法、伸び率法、関数モデル法などが用いられているが、本稿で取り上げる手法は関数モデル法のうち回帰モデルによる方法である。同じ手法で構築した回帰推定式であっても地域や年度による差がみられるのが通例である。一方、説明変数間に高い相関がある場合には多重共線の問題が生じ、説明変数の係数値に影響を及ぼし、推定式に大きな相違が表われてしまうことも多い。

(1) 推定式の安定性

推定式が安定しているということは何で判断するかには様々な方法（統計的な評価）、意見があるが、ここでは仮に調査年度及び調査地域が異なっても、得られた推定式の説明変数の係数値が似かよっていれば安定性があるという判断で検討する。パーソントリップ調査が実施された都市圏ごとに調査の背景、調査方法、抽出率、集計方法などに

表-1 発生量の予測式

都市圏名	地域区分	目的種別		
		通 勤	通 学	業 務
東京 (昭43)		$T=0.62PE_{23}+2566$	$T=0.24P+189$	$T=0.59EE_{23}-3125$
東京 (昭53)	1	$T=0.65PE_{23}-2229$	$T=0.20P-186$	$T=0.59EE_{23}-3030$
	2	$T=0.66PE_{23}+656$	$T=0.22P-474$	$T=0.43EE_{23}+243$
	3	$T=0.75PE_{23}-189$	$T=0.25P-727$	$T=0.38EE_{23}+1055$
	4	$T=0.74PE_{23}+476$	$T=0.23P-239$	$T=0.43EE_{23}+340$
京阪神 (昭45)	1	$T=0.30P+98$	$T=0.19P+1053$	$T=0.29EE_2+0.71EE_3-8444$
	2	$T=0.35P-10207$	$T=0.20P+291$	$T=0.26EE_2+0.34EE_3+2265$
	3	$T=0.31P+2383$	$T=0.21P-145$	$T=0.20EE_2+0.61EE_3+3348$
	4	$T=0.33P+2525$	$T=0.20P-1825$	$T=0.36EE_2+0.64EE_3-1790$
	5	$T=0.33P-1936$	$T=0.20P+455$	$T=0.30EE_2+0.62EE_3+2603$
仙台 (昭47)		$T=0.77PE_{23}-78$	$T=0.93PS+75$	$T=0.84EE_{23}+400$

ただし、T：発生トリップエンド数 P：常驻地居住人口 PE_{23} ：常驻地2次3次産業就業人口

PS：通学地学生生徒数 EE_{23} ：従業地2次3次産業就業人口

EE_1 ：従業地1次産業就業人口 EE_2 ：従業地2次産業就業人口 EE_3 ：従業地3次産業就業人口

ある程度の違いがあるのでは単純な比較はできないが、例えば、表-1のような結果が各報告書に記載されている。東京都市圏の43年と53年の結果を取

り上げてみると、通勤の発生量が常任地2次3次産業就業人口に対する係数が53年では高くなり、業務では従業員比2次3次産業就業人口の係数が53年では低くなっている。このように両年度の調査における発生集中量の予別式には時間的変化が表われている。

調査地域の違いとしては、①同一都市圏内での地域の違い、②都市圏の違い、がある。前者は同一都市圏であっても地域の特性による差が無視できないため正確を期するため採用している予別式であり、将来の地域区分を推定した上で式のあてはめをおこなうこととなる。説明変数が都市圏によって異なるが、前者も後者もその係数値に差がみられる。

以上のように調査年度、地域の違いにより予別式も異なっている。これは現在の現象をより良く説明し、将来にわたっても社会的状況が変わらない前提に立って推定式を構築するを得ないためである。

(2) 説明変数の制約

発生集中量の推定にあたっては説明変数自体が予別可能であり、その精度が確保できる必要があるが、本稿では説明変数そのものについては言及していない。

一方、説明変数どうしに多重共線性の問題が存在すると、回帰係数の推定値は不安定となり、係数の符号に合理性が欠ける場合も多くなる。又、回帰係数の意味が解釈できない事がある。さらに説明変数の選択基準の問題も重要となる。原単位に比べ回帰モデルでは説明変数の回帰係数の意味付け、定数項の値などに検討すべき事柄が存在している。前述したように本研究では、多重共線性の少ない説明変数を選択し統計的にみて単に発生集中量を十分に良く説明する変数を選択するというのではなく、多重共線性を前提としその中でいかに説明変数を選択すべきかに主眼を置き検討することとする。

3. 回帰係数を求める手法

重回帰は共線性のあるデータに用いる場合、その影響を強く受けるが、共線性のない、しかも便宜やすい変数を選択した際には有効な手法である。本稿では他の手法による結果との比較の基礎にするとともに、年度差による回帰係数の安定性、符号の合理性の検討に用いる。

ステップワイズ回帰は一定規則のもとで選択された変数の種類ならびにその係数値を重回帰と比較するとともに、年度によって選択される変数の違いの検討をおこなう。

主成分回帰は議論のある手法ではあるが、その考えから多重共線性の影響を除去するという意味をもっている。そこで主成分ではなく、もとの説明変数に対する回帰係数を用い、重回帰の結果と比較するとともに、年度差による回帰係数値の^(7.8)違い率について検討をおこなう。

リッジ回帰は多重共線性の問題を積極的に解決しようとする一つの手法であり比較的新しい手法であるため、その実用性の検討は充分にはなされていない。リッジ回帰と重回帰との違いは下記の式で、 $k_i = 0$ の場合が重回帰、 k_i の値を変化させ β_0 が安定した時の β_i の値をリッジ回帰係数という。

$$\beta = (X'X + KI)^{-1} \cdot X'Y$$

ただし

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_n \end{bmatrix} \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} \quad K = \begin{bmatrix} k_1 & & 0 \\ & k_2 & \\ 0 & & \ddots & \\ & & & k_i & \\ & & 0 & & k_n \end{bmatrix}$$

リッジ回帰は平均二乗誤差が小さくなる推定量をつくる試みであり、最小二乗法に相対する方法であるが、 k_i の値を大きくとってしまうと残差の二乗和も大きくなってしまいますので、できるだけ小さい値で決めた方がよい。この k_i を決める判断基準には定まったものがないが一般に後述するリッジ軌跡を調べることで選んでいる。

4. 説明変数

事例に用いる被説明変数は昭和43年及び昭和53年ともに、表-2の6変数である。

説明変数は表-3に示す18変数であるが、これらの説明変数群が最適であるとは限らない。本篇では手法の相互比較をおこなうこともねらっているため、18変数の相関関係に特別な配慮を払わずに一括して回帰分析に用いている。

説明変数の相関関係は図-1に示しているが、昭和43年と昭和53年とで大きな違いはない。違いは登録自動車台数(X13)と他の変数との相関が昭和53年ほど昭和43年では高くないといった点であった。

回帰推定式の説明変数として、図-1の相関関係からすれば、常住地居住人口(X1)、常住地就業人口(X2)、地方財政歳出額(X8)、登録自動車台数(X13)、小学校数(X14)、中学校数(X15)などの変数を同時に用いるべきでないとの意見もあるが、本篇では18変数すべてを用いて検討している。

表-2 被説明変数

記号	トリップ	目的の種類	略称
T ₁	発生 トリップ数	自宅 - 勤務先	通勤発生
T ₂		自宅 - 学校	通学発生
T ₃		勤務先 - 勤務先	業務発生
T ₄	集中 トリップ数	自宅 - 勤務先	通勤集中
T ₅		自宅 - 学校	通学集中
T ₆		勤務先 - 勤務先	業務集中

表-3 説明変数

記号	内容	記号	内容
X 1	常住地居住人口	X 10	郵便貯金残高
X 2	常住地就業人口	X 11	道路延長
X 3	従業地就業人口	X 12	道路面積
X 4	工業出荷額	X 13	登録自動車台数
X 5	卸売業年間販売額	X 14	小学校数
X 6	飲食店年間販売額	X 15	中学校数
X 7	小売業年間販売額	X 16	商業系土地面積
X 8	地方財政歳出額	X 17	工業系土地面積
X 9	銀行預金残高	X 18	住居系土地面積

5. 各手法の回帰係数の比較検討

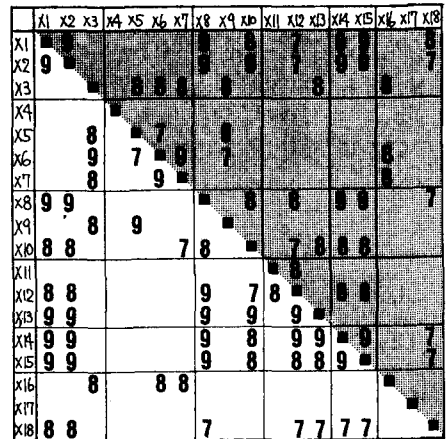
東京都の市区町村に各手法を適用したが、説明変数の値が入りできなかった所があるため、昭和43年では35市区町、53年では49市区町を対象に手法を適用することとした。なお東京都全域では55市区町村となる。

重回帰による結果は、18変数をすべて採用した場合と、逆行列を計算する時に既に掃き出し計算された変数との重相関係数の平方が1~0.01に入っていない変数を除いた場合との2種類が5求めた。

ステップワイズ回帰では説明変数の個数が2~6に減少することとなった。

主成分回帰では固有値、累積寄与率等を検討した結果、第1主成分から第3主成分までを使って回帰分析をおこなうこととし、その数、もとの説明変数に対する係数値を求め比較することとした。主成分の固有値を検討することで多重共線性の存在を確認できるが、第3主成分までを取り上げることで多重共線性の影響を少しでも除去するという意味合いを持つ。

リッジ回帰において、パラメータを変化させた時の回帰係数 β^* の変化を示すリッジ軌跡は図-2のようになる。この図に於て β_0 から β_9 のリッジ軌跡は省略している。回帰係数 β^* は標準化したものであり、他の手法の結果との比較では、標準化していない回帰係数を計算し用いている。本篇では、リッジ軌跡において β^* の値がほぼ安定してきている $k=0.2$ の時の値をリッジ回帰係数として採用した。被説明変数が変わること採用すべき k の値も変わるが、本篇では一定($k=0.2$)としたまま比較することとした。またリッジ回帰における説明変数の選択基準の一つのルールに従うならば、 β^* の値が安定しない説明変数や β^* の値が0.0に近いような説明変数は取り除くことができる。



9 : 相関係数0.9以上
8 : 相関係数0.8~0.9
7 : 相関係数0.7~0.8

注：右上は昭和43年、左下は昭和53年である

図-1 説明変数間の相関関係

(1) 43年と53年の回帰係数の比較

本節では通勤発生を別に取上げ、各手法ごとの回帰係数の比較をおこなう。表-4に得られた回帰係数の値をまとめて示す。

1) 重回帰

回帰係数の値が年によって違っているとともに符号も異なる結果が得られた。符号が43年と53年で同じなのは18変数のうち9変数であるが、同符号であっても係数の値が大きく違っている。又、常驻地居住人口(X1)と常驻地就業人口(X2)とは、高い正の相関を有しているのに、符号が逆になるなど多重共線性が顕著に表われている。さらに回帰係数の違いとともに定数の違いにも着目する必要があるが、これは今後の検討課題である。

表-4 通勤発生における回帰係数の比較

説明変数	重回帰		※制限付重回帰		ステップワイズ回帰		主成分回帰		リッジ回帰	
	43年	53年	43年	53年	43年	53年	43年	53年	43年	53年
X1 常驻地居住人口	-0.56	0.08	-0.69	0.43	-0.67	0.03	0.03	0.04	0.06	0.06
X2 常驻地就業人口	-0.77	0.73	-0.86		-0.86	0.41	0.07	0.08	0.10	0.10
X3 従業地就業人口	-0.42	-0.08					-0.01	-0.01	-0.02	-0.01
X4 工業出荷額	0.87	0.09		0.15		0.09	0.05	0.06	-0.24	0.08
X5 卸売業年間販売額	0.11	0.22	-0.01	0.02			-0.02	-0.01	-0.02	-0.01
X6 飲食店年間販売額	24.95	0.98	13.25	-1.29			-0.80	-0.26	0.55	-0.51
X7 小売業年間販売額	-2.26	0.09	-2.07	0.35			0.04	0.05	0.37	0.18
X8 地方財政支出額	-8.88	1.53	-2.07				20.81	3.53	31.96	2.13
X9 銀行預金残高	0.18	-0.02	0.03	-0.04			-0.04	-0.01	-0.00	-0.01
X10 郵便貯金残高	30.77	0.15	9.30				7.32	0.84	5.74	1.69
X11 道路延長	-0.46	0.55	-0.13	0.11			0.18	0.15	-0.01	-0.01
X12 道路面積	0.66	0.09		-0.27			0.43	0.33	0.13	0.06
X13 登録自動車台数	0.00	-0.00	-0.00	-0.00	-0.00		0.00	0.00	0.00	0.00
X14 小学校数	6.33	-3.74		-4.28			3.88	3.58	5.08	2.00
X15 中学校数	3.67	-6.07		-2.34	-13.53		7.95	8.04	11.40	5.61
X16 商業系土地面積	-0.88	-0.51	-0.79	-0.33			-0.09	-0.08	-0.49	-0.10
X17 工業系土地面積	0.01	0.08	0.16	0.08	0.13		-0.02	0.09	0.05	0.12
X18 住居系土地面積	-0.17	0.01	-0.17	-0.04	-0.21		0.11	0.09	0.10	0.12
定数	144.36	13.74	99.91	9.46	39.31	0.03	-21.12	-57.92	-170.6	-17.99
決定係数 R ²	0.9731	0.997	0.967	0.997	0.953	0.995	0.888	0.951		

※制限付とは逆行列を計算する時、既に掃き出し計算された変数との重相関係数の平方が 1~0.01 とならない変数を除いた場合をさす

計が充てできることを示している。

43年の回帰係数を見ると、概念的には正の要因として定まるはずの説明変数が負の符号になるなど合理性に欠ける点があったり、43年、53年で選択される変数が異なるなどの問題点が生じている。

2) 主成分回帰

重回帰、ステップワイズ回帰に比べ、主成分回帰では回帰係数が安定している。回帰係数の値は飲食店年間販売額(X6)、地方財政支出額(X8)、郵便貯金残高(X10)の3変数を除いて両年とも似かよった値を示している。又、符号では工業系土地面積(X17)を除き、すべて一致しているとともに説明変数の符号条件が感覚とほぼ合致する結果が得られた。しかしながら、説明変数の選択問題はこのままでは解決できない。相関関係の強い説明変数群が互いに関与し合った形で回帰係数が定まっているのであり、説明変数の個数を減らした場合、その回帰係数の絶対値が増加する傾向にある。たとえば表-5に示すように、18変数から3変数にすることで常驻地居住人口(X1)の回帰係数の値が、

表5 3変数の時の通勤発生の回帰係数

説明変数	重回帰		主成分回帰		リッジ回帰	
	43年	53年	43年	53年	43年	53年
X1 常驻地居住人口	0.43	0.01	0.12	0.16	0.12	0.15
X2 常驻地就業人口	-0.45	0.77	0.31	0.39	0.22	0.39
X3 従業地就業人口	-0.05	-0.02	-0.08	0.00	-0.07	-0.01
定数	259.04	-141.87	308.13	-95.73	427.76	-24.07
決定係数 R ²	0.921	0.990	0.867	0.988		

43年、53年ともに 0.03 又は 0.04 から 0.12 又は 0.16 へ増加している。

④ リッジ回帰

主成分回帰と似よった結果となっているが、回帰係数の値は工業出荷額 (X4)、飲食店年間販売額 (X6)、小売業年間販売額 (X7)、地方財政支出額 (X8)、郵便貯金残高 (X10) など 10 変数で異なり、符号では飲食店年間販売額が 43 年と 53 年とで異なっている。リッジ回帰も説明変数の個数を 18 から 3 へ減らすことにより、それぞれの回帰係数の絶対値が増加している。

(2) 手法の違いによる回帰係数の比較

重回帰とステップワイズ回帰の手法、特に重回帰の手法を共線性のある説明変数群に直接用いることには問題がある。相関係数などをもとに説明変数の選択をおこなわずに手法を適用すると、説明変数の回帰係数が不安定または合理性のないものになることが多い。この点が主成分回帰またはリッジ回帰と異なっている。重回帰ならびにステップワイズ回帰で得られた回帰係数は説明変数の個数に違いがあるものの似よった値となっている。しかしながら重回帰における多重共線性の問題はそのままステップワイズ回帰に引きつがれている。

重回帰およびステップワイズ回帰は主成分回帰、リッジ回帰に比べ決定係数が高くなっている。(リッジ回帰はモデルの持つ意味からも当然であるし、決定式の観測式の違いによる) このことは少なくとも現状再現という観点では重回帰又はステップワイズ回帰がすぐれているといえる。しかし回帰係数の安定性という観点においては重回帰、ステップワイズ回帰に問題がある。

主成分回帰とリッジ回帰とは異なる計算過程をとったものの似よった結果となっている。回帰係数の値が比較的安定し、符号も合理的であるという共通性はあるが、リッジ回帰では結果として説明変数の選択をおこなうことができるという違いがある。

(3) 43 年推定式の比較検討

43 年推定式に 53 年の説明変数の値を代入することによって、53 年推定値を各手法ごとに求め、推定式の時間的安定性を検討する。

43 年から 53 年の消費者物価指数の上昇は 43 年を 100 とすれば 53 年で 250 位となる。そこで説明変数のうち金額を伴う経済指標 (X4~X10) については、53 年の説明変数の実績値を 43 年推定式に代入する際に 2.5 割った上で用いている。43 年推定式に説明変数の 43 年実績値を代入した結果ならびに上記方法による 53 年推定値を図-3 に明示する。通勤発生について両年の実績値と各手法による推定値とを比較した結果、以下のような事が判明した。

- ① 43 年の現状を説明するには重回帰又はステップワイズ回帰が優れていること。
- ② 53 年の推定値には、ばらつきがあること。
- ③ 主成分回帰の 53 年推定値は実績値に比べ大きめであること。
- ④ 例えば図-3 に示した世田谷区、武蔵野市などでは 43 年から 53 年にかけて実績値が増加しているにもかかわらず、重回帰およびステップワイズ回帰では 53 年推定値が減少していること。
- ⑤ リッジ回帰では増加するか減少するかの傾向については、ある程度度示していること。

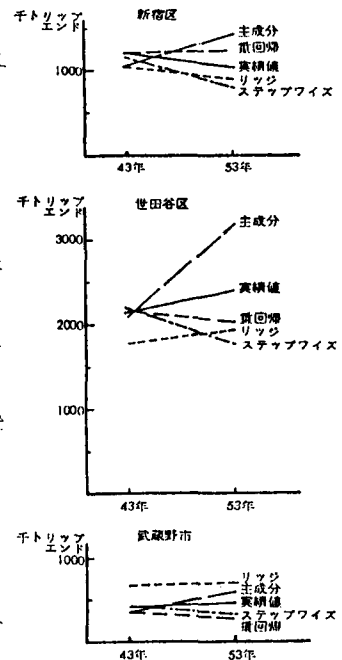


図-3 通勤発生における実績値と推定値

6. 各手法の問題点と多重共線性

(1) 重回帰

重回帰では多重共線性が大きな問題となる。しかし共線性のない説明変数を選択するには困難がある。パーソントリップ調査等における発生集中量推定式の場合、既存の解析によりある種の目安が得られており、その労力と試行錯誤は少なくなっている。しかし現在のところ最適な説明変数は交通の目的や地域、年度により異なる可能性がある。したがって最適な説明変数の選択が難しいとともに普遍性の問題も残されている。

これに併し、主成分回帰又はリッジ回帰は、前者で説明変数の選択をおこなわず考えられるすべての変数を取り込んだまま、多重共線性の問題に共処しながら普遍性をみつけ出す一つの手番りとなる可能性がある。

(2) ステップワイズ回帰

重回帰における問題点の多重共線性はステップワイズ回帰で解決することはできないし変数選択の手法としてもいくつかの留意点がある。例えば今回の検討において43年通勤発生トリップエンドを求めるに際し常住地居住人口を最初に取り入れているのは良いが、次に負の要因として常住地就業人口、さらに負の要因として住居系土地面積が各ステップで選ばれているのは不合理である。少なくとも説明変数相互の共線性が強い場合にステップワイズ回帰をそのまま用いるには問題があり、むしろ様々な変数を組み合わせながら試行錯誤で最適な説明変数の組み合わせを求めた方が良い場合もある。

(3) 主成分回帰

重回帰、ステップワイズ回帰に比べ主成分回帰では説明変数の係数値が安定し、かつ符号の合理性も得られた。しかし各主成分が何を意味するかという問題や計算過程が複雑というような問題が残る。すなわち主成分としてどのような情報を取り込み、どのような情報を切り捨てたのかの解釈が困難である。

(4) リッジ回帰

リッジ回帰は主成分回帰に似かより係数の安定性、合理性が高くなっている。しかし、リッジ軌跡を調べる際の尤の決め方には確固とした方法がなく仮説明変数が異なるごとに尤の値が異なることもある。その一方で多重共線性を考慮した変数選択をおこなうことができるので数多くの事例を通して実用性の検討を加えるべき一つの手番である。

7. おわりに

推定式の問題点としては多重共線性、説明変数の選択の他に定数項等の検討も残されている。定数項の存在は回帰係数の値との関係もあるが、発生集中量を説明するのにあまり大きな値をもつ定数であったり、定数の符号が一概に解釈できない場合には、その推定式の信頼性がなくなる場合がある。回帰係数は一般に、他の説明変数を固定した上で当該説明変数が単位増えたときにもたらされる仮説明変数の変化の大きさである、とされている。しかし定数項の意味合いを十分に把握しないと、この解釈も無意味なものになることがあると同時に説明変数間に強い相関関係がある場合にも無意味となってしまうなどの問題がある。回帰推定式の検討にあたっては回帰係数の検討にだけ意味があるのではなく、それとともに定数の検討も必要となろう。

参考文献

- 1) 佐藤 肇一、五十嵐 力、山形 耕一、実験的回帰分析法による発生交通量推定モデルの構築、土木学会第34回年報、1979。
- 2) 佐藤 肇一、桐城 信、高立 忠夫、発生重回帰モデルにおける説明変数選択法に関する研究、第3回土木学会交通計画委員会、1981。
- 3) 太田 勝敏、交通需要予測の誤差分析、土木学会第28回年報、1973。
- 4) 太田 勝敏、交通モデルと予測誤差について一虚数度分析の概念、土木学会第30回年報、1975。
- 5) 山形 耕一、交通需要予測におけるデータ誤差と予測精度への影響に関する研究、地域学研究(2巻(昭和56年度)、1982。
- 6) 桐城 信、五十嵐 忠夫、山形 耕一、発生重回帰モデルの評価方法に関する研究、交通工学、Vol.16, No.6, 1981。
- 7) S.チャビー、B.アライズ、佐和 隆光、加納 恒 訳、回帰分析の実際、新曜社、1981。
- 8) 中村 正一、マイコンプログラムつきリッジ回帰分析、BASIC数値、8月号、1982。
- 9) Hoel, A.E. and R.W. Kennard, Ridge regression: An Application to nonorthogonal problems, Technometrics, 12, 1970。