IV－66

空間相互作用モデル推定法の頑健性比較

東北大学　学生員　○河野秀俊
東北大学　正　員　内田　敬
東北大学　フェロー　宮本和明

１．はじめに

空間分布量モデルパラメータ推定の問題点として、観測の不確定性による外れ値の存在がある．これが誤差項の正規性を阻害する要因となり、通常最小二乗法（OLS : Ordinary Least Squares）による回帰分析の前提条件を満たさない．

一方、位置データを考慮した空間分布適合度指標：SFI（Spatial Fit Indicator）が提案され、SFI 規準パラメータ推定法（SFIE）が研究されている．[1]本研究では外れ値が存在する、すなわち誤差項の正規性が満たされない場合における SFIE の適用可能性についてOLS等との比較をもとに検討することを目的とする．

２．頑健性の定義

「頑健（ロバスト）である」とは、①データの一部に誤りがあるときにもパラメータの推定値にずれが生じにくい（抵抗力がある）こと、②誤差の分布が正規分布のときだけでなく、裾の広い非正規型の誤差分布に対してもパラメータの推定値の分散が十分小さくなることである．また、広義では「外れ値が存在しても誤った結論を導かないこと」ということができる．

３．比較対象となるパラメータ推定法

３．１　最小二乗推定量（OLSE）

最小二乗法は測定値に偏りがなく、その誤差が正規分布をし、モデルにも近似の誤差がないという前提を置いている．この前提が崩れると、たとえばわずか一個の外れ値のために全く異なった値が得られる危険性を有している．

３．２　Huber のM推定量

M推定量（一般化最尤推定量；generalized maximum likelihood estimator からの拡張）は残差を平均からのずれに応じてウエイトを調整する．ウエイト関数は次式のように残差に逆比例した形で表される．

$$w(u)=\frac{\rho'(u)}{u}=\frac{\psi(u)}{u} \quad \text{（ウエイト関数）} \qquad (1)$$

ここで、u：誤差、ρ (u)：損失関数、Ψ (u)：影響関数である．

本研究で用いた Huber のM推定量のウエイト関数は、次式で表される．（r：調整定数）

$$w(u)=\begin{cases} 1 & |u|\le r \qquad (2) \\ \dfrac{r}{|u|} & |u|\ge r \qquad (3) \end{cases}$$

OLS ではウエイトが常に１であり、M推定量の特殊例と解釈できる．

Huber のM推定量はくり返し再加重最小二乗法によって計算することができる．[2)]

３．３　SFI 規準パラメータ推定量（SFIE）

１）定義と定式化

SFI は空間分布量の予測値の適合度を残差の空間分布にもとづいて定量化する．過大推定ポイントから過小推定ポイントへ残差を移送してならす（観測値に一致させる）ことを考え，移送量と移送距離の積で与えられる移送コストの総和が小さい予測値分布ほど観測値分布への適合度が高いと評価する．予測値分布が線形モデルで与えられる場合，SFI を最小にする（空間分布適合度を最大にする）モデルパラメータは，SFI 算出のためのＬＰ問題に予測モデル式を組み込んだ下記の問題を解くことで得ることができる．

$$S=\sum_i\sum_j d_{ij}T_{ij} \rightarrow \min \quad \text{（目的関数）} \qquad (4)$$

$$\sum_j T_{ij} \le R_i^+ \quad \text{（供給条件）} \qquad (5)$$

$$\sum_j T_{ji} \ge R_i^- \quad \text{（需要条件）} \qquad (6)$$

$$R_i^+ - R_i^- = Y_i - A_i \quad \text{（残差の定義）} \qquad (7)$$

$$\sum_i Y_i - \sum_i A_i = 0 \quad \text{（総量一致条件）} \qquad (8)$$

$$Y_i=\sum_k \alpha_k X_{ki} + (\alpha_0 - m) \quad \text{（予測モデル）} \qquad (9)$$

$$T_{ij}\ge 0,\ R_i^+,\ R_i^- \ge 0,\ \alpha_0,\alpha_k \ge 0$$

d_{ij}：ゾーン i−j 間の残差移送距離

T_{ij}：ゾーン i−j 間の残差移送量

R_i^+, R_i^-：ゾーン i の過大予測量、過小予測量

Y_i：ゾーン i の予測値

A_i：ゾーン i の観測値

X_{ij}：予測モデルの第 k 説明変数

α：予測モデルのパラメータ　　m：調整項

２）頑健推定としての SFIE

SFIE のM推定量としての特徴を見るために、目

的関数を損失関数ρ・ウエイト関数ωに関連づける．

$$
\begin{aligned}
S &= \sum_i \sum_j (d_{ij} T_{ij}) \\
&= \frac{1}{2} \sum_i \sum_j (d_{ij} T_{ij} + d_{ji} T_{ji}) \\
&= \frac{1}{2} \sum_i \sum_j \left(R_i^+ \frac{d_{ij} T_{ij}}{R_i^+} + R_i^- \frac{d_{ji} T_{ji}}{R_i^-}\right) \\
&= \frac{1}{2} \sum_i \sum_j (R_i^+ \bar{d}_i^+ + R_i^- \bar{d}_i^-) \\
&= \frac{1}{2} \sum_i R_i \bar{d}_i \qquad (10)
\end{aligned}
$$

ここでは、次の各式をを導入した．

$$\bar{d}_i^+ = \sum_j \frac{d_{ij} T_{ij}}{R_i^+} \qquad \text{(平均移出距離)} \quad (11)$$

$$\bar{d}_i^- = \sum_j \frac{d_{ji} T_{ji}}{R_i^-} \qquad \text{(平均移入距離)} \quad (12)$$

$$R_i = R_i^+ + R_i^- \qquad \text{(絶対残差)} \quad (13)$$

$$\bar{d}_i = \bar{d}_i^+ + \bar{d}_i^- \qquad \text{(平均移送距離)} \quad (14)$$

したがって、損失関数、ウエイト関数は

$$\rho(u) = \frac{1}{2} R_i \bar{d}_i \qquad (15)$$

$$w(u) = \frac{\rho'}{R_i} = \frac{1}{2} \frac{\bar{d}_i}{R_i} \qquad (16)$$

と書け、ウエイトが残差に逆比例することが分かる．

4．　数値実験の方法

空間相互作用（交通量）モデルとして、片側制約重力モデルを取り上げる．これは対数変換によって線形モデルとなる．

$$Y_i = \alpha_0 X_{1i}^{\alpha_1} X_{2i}^{\alpha_2} \exp(-\alpha_3 D_i) \qquad (17)$$

$$\Leftrightarrow \ln Y_i = \ln \alpha_0 + \alpha_1 \ln X_{1i} + \alpha_2 \ln X_{2i} - \alpha_3 D_i \qquad (18)$$

Y_i ：（ゾーンｉから）中心部への交通量【観測値】

X_{1I}：（ゾーンiの）非就業者数

X_{2I}：（ゾーンiの）３次従業者数

D_i ：（ゾーンｉから）中心部までの距離

仙台ＰＴデータおよびゾーンを用いて、数値実験を行う．計１７３ゾーンからランダムに５０ゾーン抜いてきたものを用意し、外れ値をノイズ付加（あるゾーンの観測値を意図的に２倍する）によって与える．

これらのデータを用いて上述の３つの推定法（OLSE、SFIE、Huber のM推定量）を用いてパラメータ推定を行い、各パラメータの変動を比較する．

5．適用結果・比較

パラメータ推定の結果を図１・表１に示す．図には代表して３次従業者パラメータを示すが、他の２つも同じ傾向を示す．OLSE は元の観測値が大きいほど変動が大きい．これは元の観測値が大きいほど、それを二倍することにより残差が大きくなり、外れ値の影響が大きいことを示している．一方、SFIE は位置データを考慮した推定法であるために、観測点の密度が相対的に小さい（中心地からの距離が大きい）ゾーンでは、移送距離が大きくなりパラメータが不安定になる．図２にゾーン配置を示す．

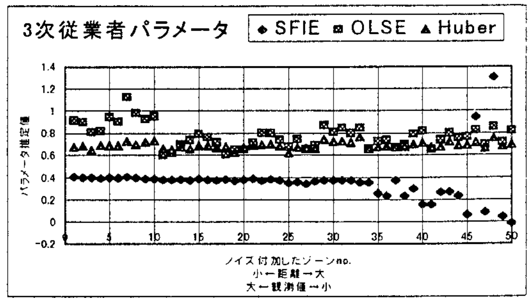

図１　３次従業者パラメータの変動

表１・各パラメータの変動係数

	変動係数（下段はno.35まで）			
	非就業者	3次従業者	距離	定数項
SFIE	−0.24498	0.5503616	0.452428	0.31776
	−0.13861	0.1469377	0.193755	0.065548
OLSE	3.154208	0.1364147	0.130667	−0.94643
	3.268777	0.1478298	0.07621	−1.02477
Huber	−0.31353	0.0435441	0.065973	0.623161
	−0.34129	0.04436	0.05902	0.64177

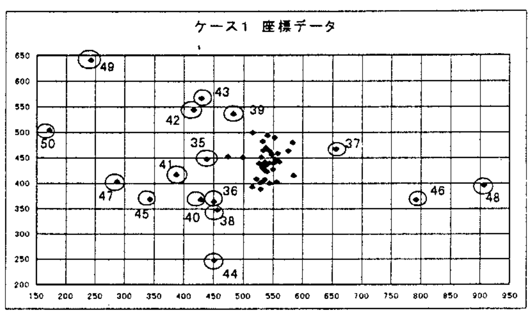

図２・ゾーン配置

6．おわりに

SFIE は、観測点が集中している場合においては Huber のM推定量と同程度に頑健であると言える．しかし、サンプル点の相対密度に影響を受け、ゾーンの配置パターンによって頑健性が損なわれる場合がある．従って、ゾーン密度が一様でない場合の残差移送距離の与え方を検討することが今後の課題と考えられる．

<参考文献>

1）　宮本和明、内田敬、明石和之：SFI 規準による空間分布モデルのためのパラメータ推定法、土木計画学研究・講演集、No. 20、pp. 63-66、1997

2）　蓑谷千凰彦：計量経済学における頑健推定、多賀出版、1992.