【土木計画学研究・講演集　No.21 (1)　1998年11月】

都市活動量空間分布モデルの頑健推定法*

—OD 交通量重力モデルの SFI 規準推定—

A Robust Estimation Method for Spatial Distribution Models of Urban Activities

内田 敬**, 明石和之***, 宮本和明****

*by Takashi UCHIDA**, Kazuyuki AKASHI***, Kazuaki MIYAMOTO*****

1. はじめに

交通量，人口などの都市活動量に関する線形モデルの推定法を考える。線形モデルの推定では，誤差項の正規性が成立すれば通常最小2乗法(OLS)により最良一致推定量が得られる。しかし，都市活動量の空間分布を取り扱う際には正規性を阻害する要因として，空間的相関，観測の不確定性が影響する。このうち前者については，誤差に空間的自己相関構造を仮定するモデルが用いられる[1]。しかし，現実への適用においては空間的連関の程度（空間的ウェイト）の定義が問題となる。またそのようにモデル化されたときの誤差分布が正規性を満たす保証はない[2]。そこで正規性が満たされない（OLS の立場からは“外れ値”と見なされる観測値が存在する）場合に適用可能なモデルパラメータ推定法，すなわち頑健推定法を検討する。

著者らは，空間分布量の適合度指標として SFI を提案し，SFI 規準によるモデル推定法を考究してきた[3]。本研究では，SFI 規準推定法を頑健推定法として位置づけて，OLS との比較によりその特性を明らかにする。以下，まず 2., 3.で，頑健推定法，SFI 規準推定法についてまとめる。そして 4.で SFI 規準推定量を頑健推定量として解釈し，その特性を示す。5.以下では，OD 交通量重力モデル推定を例にとって，ケーススタディのための諸設定について述べる。なお，線形回帰モデル：$Y_i = x_i^t \beta + u_i$ のみを考える。

2. M 推定量

頑健推定とは，誤差項の正規性が満たされないときに，“外れ値”の影響をあまり受けることなく安定した推定を得ることができ，また正規性が満たされるときにも OLS に遜色の無い効率を持つような推定の方法，推定量である。そのひとつである M 推定量を概説する[2]。

最尤推定量は，

$$\max \log L(\theta) = \sum_{i=1}^{n} \log f(x;\theta) \tag{1}$$

ここに，$f(x;\theta)$:確率密度関数

の解として与えられるが，ここで，損失関数

$$\rho(x;\theta) = -\log f(x;\theta) \tag{2}$$

を導入すると，(1)を総損失最小化に転換できる：

$$\min \sum_{i=1}^{n} \rho(x_i;\theta) \tag{3}$$

M 推定量（一般化最尤推定量；generalized maximum likelihood estimator）は，式(3)の解として与えられる。損失関数を式(2)のように定義すれば通常の最尤推定量であるが，一般に損失関数は任意に，頑健性を高めるべく設定される。

ρ が微分可能かつ，0 のまわりで対称な凸関数のとき，M 推定量はウェイトが残差に依存する加重最小 2 乗回帰の解として求めることができる。ただし，くりかえし再加重最小 2 乗法(IRIS; Iterative Reweighted Least Squares)によることとし，ウェイトと損失関数は同時には評価されない。すなわち，

ウェイト関数：

$$w(u) = \frac{\rho'(u)}{u} = \frac{\psi(u)}{u} \tag{4}$$

$\psi(u)$：影響関数

u：誤差

損失関数（加重最小 2 乗回帰）：

$$\rho(u) = \frac{1}{2} w(u) u^2 \tag{5}$$

M 推定量の“外れ値”に対する頑健性は，ウェイト関数あるいは影響関数に端的に見ることができる。例えば，通

*　キーワーズ：パラメータ推定，OD 交通量，空間分布適合度

**　正会員，博(工)，東北大学大学院工学研究科土木工学専攻
（〒980-8579 仙台市青葉区荒巻字青葉 06）
TEL 022-217-7476, FAX 022-217-7477
E-mail: uchida@civil.tohoku.ac.jp
***　学生会員，東北大学大学院情報科学研究科
****　フェロー，工博，東北大学東北アジア研究センター

常最小 2 乗推定量(OLSE)も M 推定量の特殊例(頑健性なし)と位置づけられるが，このウェイト関数は，

$$w(u)=\frac{\psi(u)}{u}=\frac{u}{u}=1 \tag{6}$$

であり，外れ値のパラメータ推定値に対する影響力に限界がない(崩壊点=1 個，総誤差感度＝無限大)。一方，M 推定量として代表的な Huber の ψ 関数では，

$$w(u)=\begin{cases}1 & |u|\le c\\ \dfrac{c}{|u|} & |u|>c\end{cases} \tag{7}$$

ここに，c: 調整定数(tuning constant)

であり，“外れ値”の影響が軽減される。ただし，このような M 推定量は被説明変数の“外れ値”には頑健であるが，説明変数に関しては頑健性を持たない。

3. SFI 規準推定量(SFIE)

(1) 空間分布適合度指標(SFI)[3]

SFI は空間分布量の予測値の適合度を残差の空間分布に基づいて定量化する。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均す(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。SFI は対象とする数量のタイプにより具体形は異なってくるが，集計量の場合には移送コスト最小の輸送問題を解くことで算出できる(次項参照)。

(2) SFI 規準パラメータ推定法 [4]

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。

[SFIE 問題]

$$S=\sum_i\sum_j d_{ij}T_{ij}\quad\rightarrow min. \qquad \text{(目的関数)} \tag{8a}$$

$$\sum_j T_{ij}\le R_i^+ \qquad \text{(供給制約)} \tag{8b}$$

$$\sum_j T_{ji}\ge R_i^- \qquad \text{(需要制約)} \tag{8c}$$

$$R_i^+-R_i^-=Y_i-A_i \qquad \text{(残差の定義)} \tag{8d}$$

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

$$Y_i=\sum_k \alpha_k X_{ki}+(\alpha_0-m) \qquad \text{(予測モデル)} \tag{8f}$$

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

ここに，

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_{ki} : 予測モデルの第 k 説明変数，

α_k : 予測モデルのパラメータ，

m: 予測モデルの定数項について非負条件 $\alpha_0\ge 0$ を成立させるための調整項(定数，>0)

4. 頑健推定としての SFIE

SFI 規準推定量は，広義の M 推定量のひとつとみなすことができる。一般に M 推定量は，式(4)に示したようにウェイトを残差に逆比例させる。一方，SFIE の損失関数は，

$$\rho_{SFI}(x;\theta)=func(R_i(x;\theta),d_i(x;\theta)) \tag{9}$$

であり，残差 R のみならずサンプル点の空間配置を反映する移送距離 d も考慮している。移送距離は LP 問題(8)を解くことによりパラメータと同時に決定される。

SFIE の M 推定量としての特徴をみるために，目的関数を損失関数，ウェイト関数に関連付ける。まず，サンプル点ペアごとに定められている損失を，個々のサンプル点に対応するように分割・統合する：

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

ここに，

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

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

中継移送がない($R_i^+\bullet R_i^-=0$，$\bar{d}_i^+\bullet\bar{d}_i^-=0$)と仮定すると，

$$S = \frac{1}{2}\sum_i R_i \bar{d}_i \tag{11}$$

ここに,

$$R_i = R_i^+ + R_i^- \quad (絶対残差),$$

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

したがって, 損失関数, ウェイト関数は,

$$\rho(u) = \frac{1}{2} R_i \bar{d}_i \tag{12}$$

$$w(u) = \frac{\rho'}{R_i} = \frac{1}{2}\frac{\bar{d}_i}{R_i} \tag{13}$$

式(4)と(13)を比較すると, ウェイトが残差に逆比例する点で SFIE は M 推定量(広義)のひとつとみなすことができよう。ただし, 分子を見ると, SFIE における平均移送距離 $\bar{d}$ は LP 解から得られるものであって解析的には表現できない。そして, 基本的には $\rho'(u)$ と同様の意味合いを持つもののサンプル点相互間の関係にも依存する点が異なる。

ウェイト関数(6), (13)を比較することにより, OLSE と対比したときの SFIE の特徴として,

(i) 残差の大きなものほど軽視する(外れ値に影響されにくい),

(ii) 平均移送距離の小さなものほど軽視する,

ことがわかる。ただし, 平均移送距離はサンプル点の空間配置のみならず, それらにおける残差の大きさにも依存するから, (i), (ii)の総合的な効果は一概には判断できない。例えば, 大きな残差を持つサンプル点が, その残差の移送先を近傍で見つけられる可能性は大きくはないから平均移送距離は大きくなりがちであり, ウェイトに対して相反する効果が作用する。

なお, SFIE と同様に LP により推定問題を構成している例として O'Kelly *et al.*[5]がある。目的関数に重み付き絶対偏差和を設定しており SFIE 問題と類似している。しかし, 重みの与え方は恣意的であり, むしろ主観的・個人的判断を取り入れられることを特長と主張している。ウェイトの意味付けを試みている点で, 本研究はより発展性を有すると評価することができよう。

5. 移送距離の設定

式(13)に示されるように, SFIE のウェイトは, 平均移送距離, したがって残差の分布パターンおよびゾーン間距離に依存する。M 推定量においては, 残差の大きさに基づくウェイトの定義内容(関数形)によって頑健性を制御する。SFIE においても同様の方法が可能ではあるが, そのほかに移送距離 d_{ij} の設定により頑健性を変化させることができる。そもそも SFI は空間分布の適合度指標であるから, 移送距離 d_{ij} の設定如何で, 空間分布としての適合度と頑健性のバランスが良い推定値を得ることも可能であろう。

ここでは, 2つの制御パラメータを導入して, 移送距離を次式のように定義する。

$$d_{ij} = \left(l_{ij}/l_{i0}\right)^{\omega}, \tag{14}$$

ここに,

$l_{ij} = l_2[i,j]$: i-j 間の距離(l_2 ノルム)

$l_{i0} = \min_j l_2[i,j]$

ω : 距離換算指数

なお, ω=0 の場合には, 残差分布パターンによらず等ウェイトであり, 最小絶対偏差法(LAR; Least Absolute Residuals)に一致する。このとき, 位置母数推定としてはメジアンを推定することになる[6]。

6. ケーススタディの設定

(1) 適用対象

具体例として, 空間相互作用量(交通量)モデルのひとつである片側制約重力モデル(production-constrained gravity (PCG) mode[7])を取り上げる。PCG は対数変換により線形モデルとなる:

$$I_{ij} = KM_i^{\alpha} M_j^{\beta} d_{ij}^{\gamma} u_{ij} \tag{15}$$

$$\ln I_{ij} = \ln K + \alpha \ln M_i + \beta \ln M_j + \gamma \ln d_{ij} + \ln u_{ij} \tag{16}$$

サンプルの空間配置は, 不定形ゾーンによるものとする。具体的には仙台 PT のゾーンを用いる。データも現実の PT データを用いる。

説明変数は, 通常用いられるものに加えて, 近隣条件を表すものも導入する。近隣条件を考慮することによる, 残差の空間的自己相関問題の回避可能性を検討する。近隣条件変数の具体例としては, 土地利用バランス[8]や街路パターン[9]が挙げられる。

OD 交通量に関して SFIE を求めるには, 式(14)に示したゾーン間距離を OD ペア相互間の距離に拡張することが必要である。しかし片側(発生)制約の場合には, 残差の移送を, 同じ発生点(i)を持つ OD ペア間に限定することにより, 問題(8)の微修正で対応できる。すなわち,

$$d_{ij\to ik} = d_{jk}$$

とし，式(8e)総量制約を発生ゾーンごとに与えれば良い。

(2) 比較対象

式(16)は通常最小2乗法(OLS)で推定できる。ただし，OLS を適用することは，誤差項の分布に関して$\ln u_{ij}$が正規分布に従うことを仮定していることになる。これに対して M 推定量のひとつである Huber の M 推定量は正規分布よりも裾の長い分布の場合に有効である。これら2つの推定量と SFIE を比較する。

なお制御パラメータとして，Huber の M 推定量は調整定数cを，SFIE は距離換算指数ωと基準距離l_{i0}を持つ。これら制御パラメータに関して複数のケースを設定して比較を行う。

(3) 評価基準・方法

大きく2つの観点から推定法の評価を行う。すなわち，

i) 残差の空間分布

推定されたモデルパラメータの下での残差の空間分布を，SFI と空間的自己相関検定の代表的指標である Moran's I index[10]とで評価する。

ii) 頑健性

サイズの限られたサンプルに基づくパラメータ推定の安定性を，実データに対する操作によって見る。

①サンプル除去やノイズ付加によるパラメータ推定値の変動をみる。

②予測の適合度についてクロスバリデーションにより評価する。

7. おわりに

本稿では，SFI 規準推定量の頑健推定としての位置づけを示し，その特性の評価のためのケーススタディの枠組みを示した。実データに基づく特性評価の結果は講演時に示す。

参考文献

1) Anselin, L. & R.J.G.M.Florax ed.: New Directions in Spatial Econometrics, Springer, 1995.
2) 蓑谷千凰彦：計量経済学における頑健推定，多賀出版，1992.
3) 宮本和明，三浦良平：空間分布適合度指標(SFIs)の提案，土木計画学研究・論文集，No.10, pp.135-142, 1992.
4) 宮本和明，内田　敬，明石和之：空間分布適合度指標の発生，分布交通量モデル推定への適用，第 17 回交通工学研究発表会論文報告集，pp.189-192, 1997.
5) O'Kelly, M.E., W. Song & G. Shen: New estimates of gravitational attraction by linear programming, *Geographical Analysis*, Vol.27, No.4, pp.271-285, 1995.
6) 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.
7) Fik, T.J. & G.F. Mulligan: Functional form and spatial interaction models, *Environment and Planning* A, Vol. 30, pp.1497-1507, 1998.
8) Kockelman, K.M.: Travel Behavior as Function of Accessibility, Land Use Mixing, and Land Use Balance, *Transportation Research Record*, No.1607, pp.116-125, 1998.
9) Crane, R. & R.Crepeau: Does Neighborhood Design Influence Travel?: A Behavioral Analysis of Travel Diary and GIS Data, *Transportation Research*-D, Vol.3, No.4, pp.225-238, 1998.
10) Florax, R. & H. Folmer: Specification and estimation of spatial linear regression models—Monte Carlo evaluation of pre-test esitimators, *Regional Science and Urban Economics*, Vol.22, pp.405-432, 1992.