【土木計画学研究・論文集 No. 17 2000年9月】

空間的自己相関を記述するための重み行列の構造が分析結果に及ぼす影響*

The Influence of the Structure of Spatial Weight Matrix on Regression Analysis in the Presence of Spatial Autocorrelation *

堤 盛人**・井出 裕史***・清水 英範****

*by Morito Tsutsumi **, Hiroshi Ide*** and Eihan Shimizu *****

1.はじめに

地理的空間を対象とした分析モデルのパラメータ推定においては、誤差項に空間的自己相関が生じることが多いため、さまざまな統計学的対処法が提案されている[1)~4)]。これらの手法では、通常、空間的自己相関が空間内の相互作用に起因するものとしてモデル化が行われる。その際、空間的な相互作用は、地点間の物理距離をもとにした「空間重み行列」により構造化されるが、従来からパラメータの推定結果がこの構造化に大きく依存する問題が指摘されている。

本研究では、誤差項の空間的自己相関に対する対処法を適用する際に、重み行列の構造が分析結果に及ぼす影響について、社会資本整備政策の分析において特に重要な要因の一つである地価を対象とした回帰モデルを用いて実証的な考察を行う。

2. 空間的自己相関に対する対処法と重み行列

(1) 誤差項の空間的自己相関

空間的自己相関とは、例えば地理的空間を対象とした回帰モデル

$$y = X\beta + \varepsilon \qquad (1)$$

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

i :地点番号$(i=1,\cdots,n)$ n:データ数

における誤差項 ε についての空間的な系列相関を言う。この場合、通常の誤差項の仮定

$$Var(\varepsilon) = \sigma^2 I \qquad (2)$$

が成立しないため、これを無視すると推定量が有効性を持たない等の問題が生じる。

空間的自己相関を検出するためにはいくつかの指標が提案されているが、パラメータの推定によって得られた残差を用いて次式のように定義される Moran の I 統計量を用いる I 検定[4),5)] は、計算の容易さもあって最もよく用いられている。

$$I \equiv \frac{n}{\sum_j \sum_i w_{ij}} \frac{\varepsilon^t W \varepsilon}{\varepsilon^t \varepsilon} \qquad (3)$$

ここで、W は w_{ij} を要素とする $n \times n$ の行列で、空間重み行列 Spatial Weight Matrix と呼ばれ、2.(3)において詳しく述べるように、地点、ゾーンあるいはメッシュ間の空間的依存の度合いを表す行列である。なお、Moran の I 統計量を用いて定義される次式の Z は、観測数が十分に多いとき漸近的に正規分布 $N(0,1)$に従うことが知れており、本研究 3.においてもこの Z を用いることとする。

$$Z = \frac{I - E[I]}{\sqrt{Var[I]}} \qquad (4)$$

(2) 空間的自己相関に対する対処法

空間的自己相関に対する対処法として次の(a)~(d) [1)~4)] に示すような統計モデルが提案されている[6)]。

$$\text{(a)}\ y = \rho W y + X\beta + u \qquad (5)$$

$$\text{(b)}\ y = X\beta + WX\gamma + u \qquad (6)$$

$$\text{(c)}\ y = X\beta + \varepsilon, \quad \varepsilon = \lambda W\varepsilon + u \qquad (7)$$

$$\text{(d)}\ y = X\beta + \varepsilon, \quad \varepsilon = \lambda W u + u \qquad (8)$$

ただし、ρ, λ, γ :パラメータ　$u \sim N(0, \sigma_u^2 I)$

いずれのモデル化も、空間的自己相関の原因を空間的な相互作用に起因するものと仮定し、その相互作用の度合いを空間重み行列 W を用いて表現している。

(3) 重み行列依存性

2.(1)で述べた空間的自己相関の診断や2.(2)のモデル化手法のいずれにおいても、実際には未知である W を分析者が先験的に与える必要がある。W の構造には様々なタイプが考えられるが、各要素 w_{ij} を地点 i と地点 j の間の空間距離 d_{ij} を用いて、例えば次式のように与える[5)]。

$$w_{ij} = c_j \big/ d_{ij}^{\alpha} \ (i \neq j), \qquad w_{ii} = 0 \qquad (9)$$

ここで、c_j は $\sum_i w_{ij} = 1$ とするための規格化定数であり、このとき Moran 統計量は次式のように単純化される。

*キーワード：調査論・地域計画・地価分析

**正員、博士（工学）　***正員、修士（工学）
****正員、工学博士
東京大学大学院 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
TEL:03-5841-6128 FAX:03-5841-7453)

$$I = \frac{\boldsymbol{\varepsilon}^t \boldsymbol{W} \boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}^t \boldsymbol{\varepsilon}} \tag{10}$$

一方、αは定数であり、分析者が事前に 1/2, 1, 2 などの値を与えることが多い。（以後、α のことを「重みパラメータ」と呼ぶこととする。）しかし、用いる重み行列により空間的自己相関の診断結果が異なり、さらにパラメータ推定の結果が非常に大きく異なる可能性があり、空間重み行列を用いた分析手法を適用する際の大きな課題の一つである。このような問題をここでは「重み行列依存性」と呼ぶこととする。重み行列依存性に関する研究は、主に仮想データを用いたシミュレーションによるものが中心で、例えば平均二乗誤差等を用いることによって重み行列の設定の誤りを検出できるかといった問題を取り扱ったものが多い[7]。これに対し本研究では、公示地価を用いた回帰モデルを用い、重み行列の構造が分析結果に実際どの程度影響を及ぼし得るかについて実証的な考察を行う。

3. モデルを用いた重み行列依存性の実証分析

(1) 実証に用いるモデルとデータ

本研究では、東京都足立区内の公示地価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合計 52 点を対象とした、つぎのような回帰モデルを作成した。

$$y = \beta_0 + \sum_{m=1}^{4} x_m \beta_m \tag{11}$$

表 1 モデルに用いた変数

y	公示地価 (円/m^2)
x_1	土地面積 (m^2)
x_2	最寄り駅までの距離 (m)
x_3	最寄り駅から北千住までの時間 (min)
x_4	容積率 (%)

x_3は Windows 版「駅すぱあと」（㈱ヴァル研究所）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。

まず、式(9)について、誤差項に共分散が存在しないと仮定したモデル(o)

$$\text{(o)} \quad \boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \boldsymbol{u} \sim N(0, \sigma_u^2 \boldsymbol{I}) \tag{12}$$

に最尤法を適用してパラメータを推定した（表 2）。自由度修正済み相関係数は 0.89 であった。

表 2 (o)のパラメータ推定結果

	推定値	t値
β_1	203	2.79
β_2	-25.8	-6.27
$\beta_3(\times 10^3)$	-2.51	-2.09
β_4	231	5.16
$\beta_0(\times 10^5)$	3.12	18.5

(2) 空間的自己相関の診断

つぎに、式(12)における残差 $\boldsymbol{u}$ に対し、Moran の統計量 (Z) を用いて空間的自己相関の診断を行なった。空間重み行列として、式(9)における重みパラメータα は α = 0.5, 1, 2, 3, 5, 10　の 6 種類を用意した。なお、重みパラメータについては $\sum_i w_{ij} = 1$ となるよう基準化を行っている。

表 3 空間的自己相関の診断

α	0.5	1	2	3	5	10
Z	5.21	4.77	4.01	3.91	3.74	3.47
上側確率(%)	*0.00*	*0.00*	*0.01*	*0.01*	*0.02*	*0.05*

当然ながら、表 3 に示すように、用いる重み行列によって Moran 統計量の値も異なるが、今回用いた例では結果としてどの重み行列に対しても有意水準 5%で「空間的自己相関が存在しない」という帰無仮説が棄却され、強い空間的自己相関が示唆された。

(3) 空間的自己相関に対する対処法の適用結果

つぎに、2.(2)に示した(a)~(d)の手法を用いて、各パラメータの推定を行った。パラメータの推定には最尤法を用いている[1]。本研究が対象としているような空間を対象とした 1 時点のデータセットに対する回帰分析において、誤差項の分散共分散行列をモデル化し、それに用いるパラメータも含めて同時推定する方法として最尤法を用いるには、本来、検討すべき問題が多い[8]。しかしながら、現時点でこれに代わる有力な方法が無いため、本研究においても、以後、パラメータの推定には最尤法を用いている。

(c)の手法を用いた結果について表 4 に示す。重みパラメータαによって、誤差項に直接影響を及ぼすラグパラメータλのみならず説明変数のパラメータ推定値もかなり大きく変化しており、明らかに重み行列依存性が確認される。例えば、交通政策に直接関わる説明変数 x_3 の係数パラメータ値については　α=0.5 とα=3 では 1.8 倍もの違いが見られる。なお、t 検定によるパラメータの有意性ついては、例えば、説明変数 x_1 は重みパラメータα=0.5, 1 の時には有意水準 5%でも十分有意と判断されたものが、α=10 においては同じ有意水準において棄却された。

表 4 (c)のパラメータ推定結果－推定値

α	0.5	1	2	3	5	10
λ	0.38	0.70	0.68	0.57	0.49	0.41
β_1	194	166	122	108	100	103
β_2	-26.0	-26.0	-28.0	-30.0	-31.0	-31.0
$\beta_3(\times 10^3)$	-2.42	-2.09	-1.50	-1.33	-1.34	-1.47
β_4	235	242	229	209	190	185
$\beta_0(\times 10^5)$	3.11	3.10	3.17	3.25	3.33	3.35

空間的自己相関がないと仮定されている残差 $\boldsymbol{u}$ について、Moran 統計量 (Z) を計算した結果は表 5 に示すとお

りである。これから、重みパラメータが $\alpha=3$ のときに空間的自己相関を最も改善していると判断される。一方、$\alpha=0.5, 1$ のような重みパラメータを用いた場合には、空間的自己相関が十分解消されていないと判断される。

表 5 重みパラメータとMoran統計量

α	0.5	1	2	3	5	10
Z	4.59	3.34	0.79	0.14	-0.28	-0.34
上側確率(%)	*0.00*	*0.08*	*42.8*	*89.0*	*78.3*	*73.4*

なお、ここでは(c)の手法を用いた結果についてのみについて表 4 に示しているが、(a), (b), (d)についても、大雑把にはこれと似たような挙動を示しており、特徴的な傾向等は見られなかったため、それらに関しての詳細については省略することとする。

(4) 重みパラメータの同定と重み行列依存性に関する考察

つぎに、各手法(a)〜(d)について尤度を最大化する重みパラメータ α を求めた。図 1 に、誤差項の空間的自己相関を考慮しないケース(o)（$\alpha=0$）も含め、手法(c)についての各重みパラメータ α とそれに対応する最大集中対数尤度をプロットして示す（パラメータ α に関しては、ここではいくつかの整数値のみについて計算を行った）。本ケースでは、$\alpha=5$ 程度において対数尤度が最大になった。一方、表 5 からは、$\alpha=3$ 程度において残差の空間的相関が最も良く解消されており、対数尤度の最大化が必ずしも残差の空間的自己相関の解消に結びつくとは限らないことが分かる。

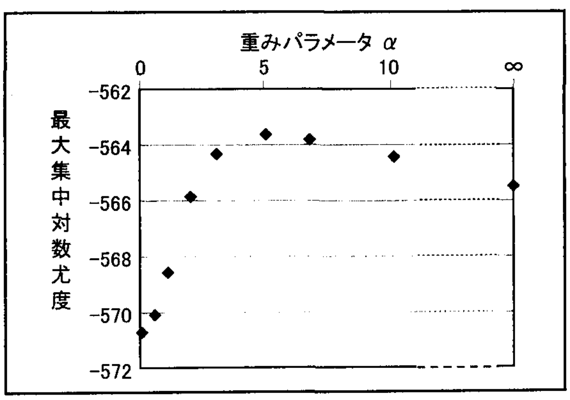

図 1 重みパラメータ α と最大集中対数尤度(モデル(c))

尤度関数の重みパラメータ α に対する単峰性若しくは多峰性については、解析的な扱いが困難であるため直接確認していない。ただし、いくつかの整数値 α に関し、α を固定しておいて対数尤度を最大化した集中（集約）対数尤度を求めた限りにおいては、極く大雑把には、単峰性を有していると仮定しても大きな問題はないものと判断される。

なお、(o)との比較では、(o)よりもパラメータが ρ, λ の計 2 つ増えたことによって当然最大対数尤度は大きくなる。パラメータが増えたことによってモデルの汎化性が低下している可能性があるため、(o)と(c)の比較も行う際には赤池情報量基準（AIC）のような基準を導入すべきであろうが、仮に AIC を用いる場合に、単純に［−2×（最大対数尤度）+2×（パラメータ数）］としてよいか否かについては現時点では著者らは十分な知識を有しない。そのため、ここでは対数尤度の値のみを用いて示している。

(a)〜(d)の各手法において、対数尤度を最大化する重みパラメータは、(a) $\alpha=3$ (b) $\alpha=1$ (c) $\alpha=5$ (d) $\alpha=5$ と求められた。なお、手法(a), (b), (d) に関しても図 1 と同様のプロットを行ったが、パラメータの変化に対する挙動は概ね同様の傾向を示した。

一般に、計算の手間などの問題から、重みパラメータをアプリオリに与える場合が少なくなく、グラビティモデルのアナロジーから特に $\alpha=2$ などが用いられることが多い。しかし、本研究で用いたいくつかの対処法においては、対数尤度から判断するとより大きな重みパラメータが適切であると判断される結果が得られた。そこで、比較のため、$\alpha=\infty$ とした空間重み行列を用いてパラメータの推定を行い、対数尤度最大化の観点から最適な重みパラメータによる推定結果と比較した（表 6）。これによれば、$\alpha=3, 5$ の推定結果は $\alpha=\infty$ の推定結果にかなり近く、しばしばアプリオリに用いられる $\alpha=1, 2$ とは重み行列の構造が大きく異なる。何故なら、$\alpha=1, 2$ では分散共分散行列の要素には w_{ij} によって重み付けられた他のあらゆる地点の変数が入ってくるのに対し、$\alpha=\infty$ のケースでは最も距離の近い点のみの作用を考慮し、それ以外の作用を無視することとなるからである。

表 6 各手法によるパラメータの推定結果

モデル	(o)	(a)		(b)	
α	—	3	∞	1	∞
対数尤度	-570.2	-565.3	-566.1	-557.7	-556.7
ρ	—	0.40	0.23		
β_1	203	153	144	308	173
β_2	-25.8	-17.8	-21.0	-23.3	-28.0
$\beta_3(\times 10^3)$	-2.51	-1.64	-1.68	-1.74	-1.39
β_4	231	204	195	256	213
$\beta_0(\times 10^5)$	3.12	1.74	2.38	-0.62	2.97
γ_1				3.40×10^3	184
γ_2				132	6.86
$\gamma_3(\times 10^4)$				-4.48	-0.27

表 6 続き

モデル	(c)		(d)	
α	5	∞	5	∞
対数尤度	-563.9	-565.6	-565.6	-567
λ	0.49	0.37	0.38	0.28
β_1	100	90	123	119
β_2	-31.0	-31.3	-29.7	-29.9
$\beta_3(\times 10^3)$	-1.34	-1.48	-1.69	-1.76
β_4	190	178	204	196
$\beta_0(\times 10^5)$	3.33	3.38	3.28	3.31

次に、これらのことを本研究で用いたデータに即して具体的に考察する。図2に示すように、本研究で用いたデータにおいては、最寄りの公示地点間の距離はおおむね500m～1km 程度である。最適な重みパラメータが$\alpha=\infty$に近いということは、式(11)の回帰モデルで説明しきれない要因における空間的自己相関が、500m～1km 程度の範囲内に起因するものであることを示唆している。仮に$\alpha=1$ や$\alpha=2$といった重みパラメータを用いると、数 km 程度の割合広い範囲との相関を想定することになり、結果的に大きく値の異なるパラメータを得る。本研究で対象とした地域に関しては、用途地域(特に工業系地域の存在)が残差項に何らかの影響を及ぼしている可能性がある。(なお、用途地域用途地域に関しては、ダミー変数として説明変数に加えてみたが、有意には効かなかった。)本研究で得られた結果が、仮に、誤差項の空間的自己相関をモデル化することによりそれらがうまくモデルに取り込めたものであると考えると、用途地域の空間的規模を考えた場合には 500m～1km という空間的スケールはある程度妥当なものであろう。もちろん、用途地域以外の要因も誤差項に影響を与えていると考えられるため、このことだけによって求めた重みパラメータが本研究の分析対象において最も適当なものであると即断すべきではない。

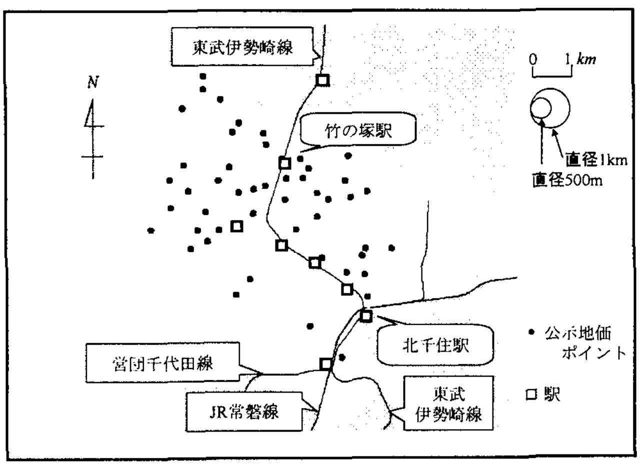

図2 分析に用いたデータの空間的分布とスケール

4. おわりに

本研究では、回帰モデルの誤差項に空間的自己相関が生じている場合の対処法において、誤差項の空間的自己相関を記述するために導入され、通常、空間上の物理距離によって規定される「空間重み行列」が、分析者によってどのように与えられるかによりパラメータの推定結果が大きく異なる可能性があるという問題を取り上げた。まず、実際の地価データを用いた回帰モデルを作成し、この重み行列を変えることによるパラメータの推定結果について考察し、それが重み行列に大きく依存していることを確認した。次に、重み行列を決定するパラメータを尤度最大化に基づき同定した結果、いくつかの対処法においては、通常用いられることの多い値とは大きく異なる重みパラメータの値が選択された。具体的に言えば、比較的狭い地理的範囲の影響のみを考慮した、通常の設定とは大きく構造の異なる重み行列が選択された。

最後に、本研究は、ここで示したような方法によって重みパラメータを同定することを主張するものではないことを断っておきたい。そもそも、3.(3)でも述べたように、本研究が対象としているような空間を対象とした1時点のデータセットに対する回帰分析において、誤差項の分散共分散行列をモデル化し、それに用いるパラメータも含めて同時推定する方法として最尤法を用いるためには、検討すべき問題が多い[7]。重要なことは、通常、分析者があまり関心を示さない重み行列の設定が、分析結果に大きな影響を及ぼす可能性があり得ることを十分に認識のうえで、分析対象となる現象や地域の特性、実際に用いる回帰モデル、さらには分析者が有する事前の情報に応じてモデル分析を行う必要があるということに他ならない。

もちろん、実際の分析において残差に空間的自己相関が認められた場合には、まず、可能な限りデータの追加や加工、場合によっては関数形の変更等によりモデルを変更することも必要であり、本研究で扱ったような対処法は、あくまでそれらの手段を尽くしたのちに適用されるべきものであることを改めて強調しておきたい。

なお、2.で述べたように、重み行列依存の問題に関する研究は仮想データを用いたシミュレーションによるものが多い[7]。しかしながら、現時点においては、それらにおいても重み行列の誤った設定を回避する決定的な方法は示されていない。一方、実際のデータを用いた実証研究は皆無に近いが、本研究で示したように重み行列をアプリオリに与えるのではなく、データを用いて推定することにより、現象分析そのものにも貢献する余地は十分にあるものと考える。

謝辞: 貴重な修正意見を賜った匿名の査読者に、この場を借りて謝意を表します。

参考文献

1) Anselin, L.: *Spatial Econometrics : Methods and Models*, Kluwer Academic, 1988.
2) Florax, R.J.G.M. and Folmer, H. : Specification and estimation of spatial linear regression models, *Regional Science and Urban Economics*, Vol.22, pp.405-432, 1992.
3) Anselin, L. and Florax, R.J.G.M.: Small sample properties of tests for spatial dependence in regression models, *New Directions in Spatial Econometrics*, Anselin, L. and Florax, R.J.G.M., eds., Springer, pp.21-74, 1995.
4) Kelejian, H. and Robinson, D.: A Suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model, *Papers in Regional Science*, Vol.72,

pp.297-312, 1993.
5) Cliff A.D. and Ord J.K. : *Spatial Autocorrelation*, Pion, 1973.
6) 堤盛人・清水英範・福本潤也・井出裕史：誤差項に空間的自己相関が存在する回帰モデルのパラメータ推定手法に関する考察，土木計画学研究・論文集，No.15，pp.49-56，1998.
7) Anselin, L. and Rey, S.: The Impacts of Misspecified Spatial Interaction in Linear Regression Modesl, *New Directions in Spatial Econometrics*, Anselin, L. and Florax, R.J.G.M., eds., Springer, pp.111-135, 1995.
8) Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee T. C.: *The Theory and Practice of Econometrics*, Second Edition, John Wiley and Sons, 1980.

空間的自己相関を記述するための重み行列の構造が分析結果に及ぼす影響

堤盛人・井出裕史・清水英範

地理的空間を対象とした回帰モデルの誤差項に空間的自己相関が生じている場合の対処法においては、自己相関を記述するために導入される「空間重み行列」をどのように与えるかによって、パラメータの推定結果が大きく異なる可能性が指摘されている。本研究では、実際の地価データを用いた回帰モデルを作成し、このことを実証的に確認した。次に、重み行列を決定するパラメータを対数尤度最大化に基づき同定した結果、いくつか対処法においては、通常用いられることの多い値とは大きく離れた重みパラメータの値が選択された。具体的に言えば、比較的狭い地理的範囲の影響のみを考慮した、通常の設定とは大きく構造の異なる重み行列が選択された。

The Influence of the Structure of Spatial Weight Matrix on Regression Analysis in the Presence of Spatial Autocorrelation

Morito Tsutsumi・Hiroshi Ide・Eihan Shimizu

Many remedial methods for the spatial autocorrelation of error terms in regression model have been developed. However, the spatial weight matrix, which may play an important role in estimating the parameters, is often designed by the analysts *a priori*. The empirical results suggest that the parameter estimates depend heavily on the spatial weight matrix. They also suggest that the spatial weigh matrix which presents the strong correlation among neighborhoods and is very different from what is lead from the analogy of gravity law and generally used, may lead to the maximum of log-likelihood.