

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ている。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「重み行列依存性」の問題が指摘されている。

本研究では、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である地価を対象とした回帰モデルを用いて実証分析を行う。

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

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

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

$$\varepsilon = (\varepsilon_1, \dots, \varepsilon_i, \dots, \varepsilon_n)^t,$$

i : 地点番号 ($i=1, \dots, n$) 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。すなわち、通常の誤差項の仮定

$$\text{Var}(\varepsilon) = \sigma^2 I \quad (2)$$

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には、パラメータの推定によって得られた残差を用いた次式のように定義される Moran の I 統計量と呼ばれる指標が用いられる (I 検定)^{4),5)}。

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\varepsilon' W \varepsilon}{\varepsilon' \varepsilon} \quad (3)$$

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、あとで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。

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、しかも計算が比較的容易であることから最も一般的に利用される。

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

(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、(a)~(d)^{1)~4)} に示すような統計モデルが提案されている。

$$(a) y = \rho W y + X\beta + u \quad (5)$$

$$(b) y = X\beta + W X \gamma + u \quad (6)$$

$$(c) y = X\beta + \varepsilon, \quad \varepsilon = (I - \lambda W)^{-1} u \quad (7)$$

$$(d) y = X\beta + \varepsilon, \quad \varepsilon = (I + \lambda W) u \quad (8)$$

ただし、 ρ, λ : 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。

*キーワード: 計画手法論・調査論

正員、工修 *学生員 ****正員、工博

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:03-3812-2111 ext.6128 FAX:03-5689-7290)

(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} を用いて次式のように与えることが多い。

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

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 統計量は次式ようになる。

$$I = \frac{\epsilon' W \epsilon}{\epsilon' \epsilon} \quad (10)$$

また、 α は定数であり、事前に 1/2, 1, 2 などの値を与えることが多い。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る。このような問題は、「重み行列依存性」と呼ばれて、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。

そこで、本研究では公示地価を用いた回帰モデルによる実証分析を行い、重み行列の構造が分析結果に及ぼす影響について考察を行う。

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、つぎのような回帰モデルを用いた。

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

表. 1 モデルに用いた変数

y	公示地価 (円/㎡)
x_1	土地面積 (㎡)
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「駅すばあと」(株)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)

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

に最尤推定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 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(11)における残差 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)におけるパラメータを $\alpha = 0.5, 1, 2, 3, 5, 10$ としたの 6 種類を用意した。

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、表.3 に示すように、用いる重み行列によって Moran の統計量も異なる。しかし、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 1% 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認められた。

(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。なお、ここでは紙面の都合上、(c)の手法を用いた結果についてのみ示す。

表. 4 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

表. 5 パラメータ推定結果— t 値

α	0.5	1	2	3	5	10
β_1	2.70	2.42	1.94	1.80	1.72	1.70
β_2	-6.26	-6.23	-6.10	-6.22	-6.61	-6.77
β_3	-1.99	-1.63	-1.05	-0.92	-0.93	-1.04
β_4	5.24	5.39	5.15	4.83	4.53	4.39
β_0	17.9	15.7	15.3	16.2	17.5	18.1

表.4 に示すように、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 $\alpha=0.5$ と $\alpha=3$ では 1.8 倍もの違いが見られる。 t 値によるパラメータの有意性については、説明変数 x_1 は重みパラメータ $\alpha=0.5$, 1 の時には有意水準 5% でも十分有意と判断されるが、 $\alpha=10$ においては同じ有意水準において棄却される。

表. 6 重みパラメータとMoran統計量

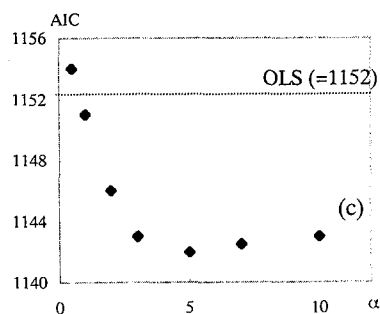
α	0.5	1	2	3	5	10
Z	4.59	3.34	0.79	0.14	-0.28	-0.34
上側確率(%)	0.00	0.08	42.8	89.0	78.3	73.4

空間的自己相関がないと仮定されている残差 u について、Moran 統計量(Z)を計算した結果は表.6 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。

4. 重みパラメータの同定に関する考察

(1) 重みパラメータの同定

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

図. 1 重みパラメータ α と AIC(手法(c))

己相関を考慮しないケースの(o)との比較のため、手法(c)についての各重みパラメータと AIC の関係をプロットした。ここでは、 $\alpha=5$ 程度において尤度が最大(AIC が最小)になった。また、(o)との比較では、重みパラメータ $\alpha=0.5$ を用いた場合には対処法を適用することにより AIC の値が(o)よりも大きくなり、パラメータが増えたことによりモデルの汎化性が低下している可能性が示唆される。

(2) 重みパラメータに関する考察

表.7 に示すように、各手法について AIC を最小化する重みパラメータはおおよそ、(a) $\alpha=3$ (b) $\alpha=1$ (c) $\alpha=5$ (d) $\alpha=5$ と求められた。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、特に $\alpha=1, 2$ 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、AIC 基準から判断すると、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、AIC 基準において最適とされたパラメータ推定結果と比較した(表 7)。

表. 7 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	∞
AIC	1152	1145	1146	1133	1151
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
$\gamma_3(\times 10^3)$				-4.48	-0.27

モデル	(c)		(d)	
α	5	∞	5	∞
AIC	1142	1145	1145	1148
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

本研究で用いたデータにおける、 $\alpha=1$ と $\alpha=\infty$ のそれぞれに対応した空間重み行列の一部を示すと以下のとおりである。

$$\begin{matrix} \alpha=1 & & \alpha=\infty \\ \begin{pmatrix} 0 & 0.38 & 0.33 & 0.29 \\ 0.20 & 0 & 0.16 & 0.64 \\ 0.37 & 0.33 & 0 & 0.30 \\ 0.17 & 0.68 & 0.15 & 0 \end{pmatrix} & \Rightarrow & \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

これからも分かるように、 $\alpha=\infty$ 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に他ならない。表7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り、結果としてパラメータの推定結果が異なる可能性が高い。

最後に、参考までに、重み行列を余り大きくないパラメータ α で表した場合と $\alpha=\infty$ とした場合における誤差項 ε の分散共分散行列を比較して表8に記す。

表 8 誤差項 ε の分散共分散行列の要素

		$\alpha \ll \infty$	$\alpha = \infty$
a	$k=l$	$\rho^2(\sum_j w_{kj} y_j + u_k)^2$	$\rho^2(y_A + u)^2$
	$k \neq l$	$\rho^2(\sum_j w_{kj} y_j + u_k) \times (\sum_j w_{lj} y_j + u_l)$	$\rho^2(y_A + u)(y_B + u)$
b	$k=l$	$(\sum_j \sum_p w_{kj} x_{jp} \gamma_p + u)^2$	$(\sum_p x_{Ap} \gamma_p + u)^2$
	$k \neq l$	$(\sum_j \sum_p w_{kj} x_{jp} \gamma_p + u_k) \times (\sum_j \sum_p w_{lj} x_{jp} \gamma_p + u_l)$	$(\sum_p x_{Ap} \gamma_p + u) \times (\sum_p x_{Bp} \gamma_p + u)$
c	$k=l$	$1 + \lambda^2 \sum_j w_{kj}^2$	$1 + \lambda^2$
	$k \neq l$	$-2\lambda w_{kl} + \lambda^2 \sum_j w_{kj} w_{lj}$	$-2\lambda, \lambda^2$
d	$k=l$	$1 + \lambda^2 \sum_j w_{kj}^2$	$1 + \lambda^2$
	$k \neq l$	$2\lambda w_{kl} + \lambda^2 \sum_j w_{kj} w_{lj}$	$2\lambda, \lambda^2$

(k, l) は誤差項 ε の分散共分散行列の行番号・列番号を表す。また、 $\alpha=\infty$ においては、地点 k から最小距離の地点を A、地点 l から最小距離の地点を B とする。 u は平均 0、分散 0 の正規分布に従う確率変数。

分散共分散行列の要素には、表左側の一般型では 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いるのに対し、右側の $\alpha=\infty$ では最も距離の近い点の影響、もしくは他の点の影響を全く考慮せずラグパラメータのみに依存した型になっている。

なお、 $\alpha=\infty$ の場合を例にすると、手法(c)・(d)の分

散共分散行列の要素はラグパラメータのみに依存した形であり、本研究では扱わなかった誤差項分散不均一に対する改善効果は小さいことが分かる。他方、手法(b)は誤差項を説明変数の線形和として表しており、分散不均一の対処法の1つともなっている。このことは、実際の計算結果により確認された。

5. おわりに

誤差項に空間的自己相関が生じている場合のパラメータ推定手法について、これまで多くの対処法が提案されている。そこでは、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」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問題となる。

本研究では、実際の地価データを用いた回帰モデルを対象に、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 AIC 基準に基づき選択した。その結果、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パラメータ値が選択され、空間上の物理距離をほとんど反映しない全く構造の異なる重み行列が得られた。

なお、本研究では重み行列の構造化に用いる関数形としてもっと単純な指数関数を用いたが、これ以外の関数と重み依存性との関係については引き続き考察を行いたい。

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*, 1973.