

誤差の空間相関を考慮した線形回帰型地域予測に関する研究*

A Study on Linear Regression Based Regional Modeling under the Spatial Correlation of Errors*

井上亮**・清水英範***

By Ryo INOUE**・Eihan SHIMIZU***

1. 背景

地域の人口、経済、交通等の社会経済データは、地域間の社会経済活動の結果として顕在化した地域の状況を表したもので、一般に、交通所要時間や交通費用等のアクセシビリティ(以後、社会距離)に依存する系列相関を持つ。一方、社会経済データの地域モデリングでは、データ制約等の事情から、説明変数として本来指標化したい地域の状況を見逃さざるを得ないことが多い。もちろん無視された地域の状況にも系列相関が存在し、そのため地域モデリングの誤差から社会距離に依存する誤差を排除することはできない。この結果、地域モデリングの誤差は社会距離に依存する系列相関(空間相関)を持つことになる。

線形回帰型地域モデリングにおいて、誤差の空間相関を考慮する方法は、空間計量経済学の分野で議論がなされてきている¹⁾が、これまでの議論は誤差構造の同定に重きが置かれていた。地域モデリングの主目的の一つは将来予測にあるが、推定されたモデルをもとに予測を行う具体的な方法についてはほとんど論じられていない。

また、従来の議論では、誤差の空間相関を考慮する場合には、地域間の幾何距離に対して誤差が系列相関を持っているとして誤差構造を同定していた。しかし、前述のように地域モデリングでは、幾何距離よりも社会距離に対して依存する相関を持っている

とするのが合理的である。この社会距離(例えば時間距離)は幾何距離と異なり、観測時と予測時の間に関係が変化する距離指標である。このように観測時と予測時で変化する距離指標を用いて誤差の空間相関を構造化し将来予測を行うためには、予測時の空間相関構造を推定する必要があるため、更なる検討が必要である。

そこで本研究では、誤差の社会距離に対する空間相関を考慮した線形回帰型地域予測の方法について、1階の空間的自己回帰モデルを例に、問題の整理を行った。更に、現在行われている通常最小二乗法(OLS)を用いた予測手法、および、幾何・社会距離に対して空間相関を構造化した予測手法を、簡便なモデルに対して適用を行い、予測精度を実証的に比較・検討を行った。

2. 空間相関を考慮した予測手法

(1) 空間相関の構造化

1 階の空間的自己回帰モデル(SAR)¹⁾による空間相関構造化について記す。

(a) 重み行列の設定

空間相関を構造化する際には、影響の大きさを表すために空間重み行列 $\mathbf{W}$ が用いられる。 $\mathbf{W}$ には、2 値による表現など様々な与え方があるが、ここでは影響の大きさはゾーン間の距離に依存すると仮定し、空間重み行列の各要素 W_{ij} はゾーン間の距離 d_{ij} の関数 $f(d_{ij})$ で与える。

(b) 空間相関の構造化

SAR では、モデル

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (1)$$

(但し、 $\mathbf{y}$:被説明変数、 $\mathbf{X}$:説明変数、 $\mathbf{u}$:観測誤差)に対して、 $\mathbf{W}$ を用いて $\mathbf{u}$ に対して式(2)の構造化を行う。

*キーワード：公共事業評価法、整備効果計測法

**学生員，工修，

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:03-5841-6129, FAX:03-5841-7453)

***正会員，工博，

東京大学大学院 工学系研究科 社会基盤工学専攻

$$\mathbf{u} = \lambda \mathbf{W} \mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (2)$$

(但し, λ :空間相関係数)

この構造化により, 観測誤差 $\mathbf{u}$ の分散共分散行列 $\mathbf{V}$ は式(3)のように表せる.

$$\mathbf{V} = E(\mathbf{u}\mathbf{u}') = \sigma^2 [(\mathbf{I} - \lambda \mathbf{W})'(\mathbf{I} - \lambda \mathbf{W})]^{-1} \quad (3)$$

(2) 最良線形不偏予測

誤差の相関関係を考慮した予測手法に最良線形不偏予測(BLUP)²⁾がある. ここでは, BLUPに必要な仮定と導かれる予測量を示す.

式(1)のモデルを用いて予測を行う場合を考える. まず, 観測誤差 $\mathbf{u}$ の分散共分散行列 $\mathbf{V}$ が既知であるとする.

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{V}) \quad (4)$$

このとき, 一般化最小二乗法(GLS)によりパラメータ推定量 $\hat{\boldsymbol{\beta}}$ は

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (5)$$

となる.

ここで, 予測モデルには式(6)を用いるとする.

$$\mathbf{y}_f = \mathbf{X}_f \boldsymbol{\beta} + \mathbf{u}_f \quad (6)$$

$$E(\mathbf{u}_f) = \mathbf{0} \quad (7)$$

(但し, $\mathbf{y}_f$:被説明変数の将来予測値,

$\mathbf{X}_f$:説明変数の将来値, $\mathbf{u}_f$:予測誤差)

また, 観測誤差と予測誤差間の共分散行列 $E(\mathbf{u}\mathbf{u}_f')$ が既知であるとする.

$$E(\mathbf{u}\mathbf{u}_f') = \mathbf{C} \quad (8)$$

このとき, 線形($\hat{\mathbf{y}}_f = \mathbf{A}\mathbf{y}$)かつ不偏性($E(\hat{\mathbf{y}}_f) = \mathbf{y}_f$)をもつ予測のうち, 予測誤差の分散が最小となる予測量は, $\hat{\boldsymbol{\beta}}$ を用いると, 式(9)のように表される.

$$\hat{\mathbf{y}}_f = \mathbf{X}_f \hat{\boldsymbol{\beta}} + \mathbf{C}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (9)$$

(3) 予測にあたっての検討課題

一般に $\mathbf{V}$, $\mathbf{C}$ が既知であることはない. そこで, 代わりに推定値 $\hat{\mathbf{V}}$, $\hat{\mathbf{C}}$ を用いて予測する手法(EBLUP)が提案されている. この予測量は式(10)で表され, 不偏性を持つことが知られている³⁾.

$$\hat{\mathbf{y}}_f = \mathbf{X}_f \hat{\boldsymbol{\beta}} + \hat{\mathbf{C}}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (10)$$

SAR で空間相関を構造化する場合には, $\hat{\mathbf{V}}$ は式(3)より与えることができるが, $\mathbf{C}$ を推定するためには予測誤差に対しても観測誤差と同様の構造化が必要である. そこで, 予測誤差 $\mathbf{u}_f$ を式(10)のように構造化する.

$$\mathbf{u}_f = \lambda \mathbf{W}_f \mathbf{u}_f + \boldsymbol{\varepsilon}_f, \quad \boldsymbol{\varepsilon}_f \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (11)$$

(但し, $\mathbf{W}_f$:予測時の空間重み行列)

この結果, $\mathbf{C}$ は式(12)のように表すことができる.

$$\begin{aligned} \mathbf{C} &= E(\mathbf{u}\mathbf{u}_f') \\ &= (\mathbf{I} - \lambda \mathbf{W})^{-1} E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f') \left[(\mathbf{I} - \lambda \mathbf{W}_f)' \right]^{-1} \end{aligned} \quad (12)$$

式(12)を用いて $\mathbf{C}$ の推定値 $\hat{\mathbf{C}}$ を求めるためには, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')$, $\mathbf{W}_f$ を与える必要があるが, これらに対しては複数の仮定が可能である.

以下では, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')$, $\mathbf{W}_f$ に対して, 予測を行う上で必要かつ妥当な仮定を設定する.

仮定 I $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')$ に対する仮定

(a) $\boldsymbol{\varepsilon}$ は観測毎に独立であるので, 観測誤差と予測誤差間に共分散がないと仮定する.

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f') = \mathbf{0} \quad (13)$$

(b) $\boldsymbol{\varepsilon}$ は空間相関の影響を排除した誤差であるが, 観測と予測の間に存在する時系列相関の影響は除去していない. そのため, 同ゾーンの観測誤差と予測誤差間にもみ共分散があると仮定する.

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f') = \sigma^2 \mathbf{I} \quad (14)$$

仮定 II $\mathbf{W}_f$ に対する仮定

(a) 予測時の距離指標 d_{fij} を用いて定義する.

$$W_{fij} = f(d_{fij}) \quad (15)$$

(b) 予測時と観測時の間では誤差の空間相関構造は変化しないと仮定し, 観測時の距離指標 d_{ij} で定義する.

$$W_{fij} = W_{ij} = f(d_{ij}) \quad (16)$$

仮定 I・II の組み合わせにより, 計4通りの予測手法が考えられる. まず, 仮定 II において, (a) $\mathbf{W}_f$ を予測時の距離で定義する場合, 予測量は表-1のよう表される.

表-1 予測 (仮定 II - (a) 予測時の距離で $\mathbf{W}_f$ を定義)

仮定 I	(a) $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f') = \mathbf{0}$	(b) $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f') = \sigma^2 \mathbf{I}$
予測量	$\hat{\mathbf{y}}_f = \mathbf{X}_f \hat{\boldsymbol{\beta}}$	$\hat{\mathbf{y}}_f = \mathbf{X}_f \hat{\boldsymbol{\beta}} + (\mathbf{I} - \lambda \mathbf{W}_f)^{-1} (\mathbf{I} - \lambda \mathbf{W}) (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})$

次に, 仮定 II において, (b) $\mathbf{W}_f$ を観測時の距離で定義する場合, 予測量は表-2 のように表される.

表-1・2より, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f') = \mathbf{0}$ (仮定 I - (a))とした場

表-2 予測 (仮定Ⅱ-(b) 観測時の距離で $\mathbf{W}_f$ を定義)

仮定Ⅰ	(a) $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')=\mathbf{0}$	(b) $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')=\sigma^2\mathbf{I}$
予測量	$\hat{\mathbf{y}}_f=\mathbf{X}_f\hat{\boldsymbol{\beta}}$	$\hat{\mathbf{y}}_f=\mathbf{X}_f\hat{\boldsymbol{\beta}}+(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$

合は、仮定Ⅱに関わらず予測量は $\hat{\mathbf{y}}_f=\mathbf{X}_f\hat{\boldsymbol{\beta}}$ となるため、計3通りの予測量が考えられる。

また、幾何距離のように観測時と予測時で変化する距離に対して空間相関を構造化する場合には、予測量は表-1 に示す2通りのみとなる。

(4) 現在の予測手法

現在、線形回帰モデルを用いた地域モデリングでは、通常最小二乗法(OLS)によるパラメータ推定が用いられている。将来予測を行う際には、推定されたパラメータと説明変数の将来設定値を用いて将来予測値の算出を行っている。更に、パラメータ推定時の残差を用いて補正を行う手法(以後、残差補正法)が用いられることが多い。

以下では、OLS を用いた予測手法の検証を行う。

OLS によるパラメータ推定では、モデル

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \quad (17)$$

(但し、 $\boldsymbol{\varepsilon}$:観測誤差)

に対して、 $\boldsymbol{\varepsilon}$ に式(18)の仮定を置く。

$$\boldsymbol{\varepsilon}\sim N(\mathbf{0},\sigma^2\mathbf{I}) \quad (18)$$

このとき、パラメータ推定量 $\hat{\boldsymbol{\beta}}_{OLS}$ は

$$\hat{\boldsymbol{\beta}}_{OLS}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (19)$$

と表される。

$\hat{\boldsymbol{\beta}}_{OLS}$ を用いて将来予測を行う場合、OLS は式(18)のように誤差の共分散が0という仮定の下でパラメータ推定を行っているため、 $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')=\mathbf{0}$ と仮定すべきであると考えられる。しかし、空間相関は存在しないが、同ゾーンにおいては時系列相関が存在すると仮定することも可能であり、 $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')=\sigma^2\mathbf{I}$ として予測を行うこともできる。

つまり、 $\hat{\boldsymbol{\beta}}_{OLS}$ を用いて将来予測を行う場合にも、 $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')$ に対して2通りの仮定(仮定Ⅰ)を置くことは可能であり、各仮定による予測量は表-3 のように表される。

表-3 予測 (OLSによるパラメータ推定の場合)

仮定Ⅰ	(a) $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')=\mathbf{0}$	(b) $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_f')=\sigma^2\mathbf{I}$
予測量	$\hat{\mathbf{y}}_f=\mathbf{X}_f\hat{\boldsymbol{\beta}}_{OLS}$	$\hat{\mathbf{y}}_f=\mathbf{X}_f\hat{\boldsymbol{\beta}}_{OLS}+(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{OLS})$

この OLS によるパラメータ推定・将来予測を行う手法は、誤差の空間相関が存在しないという仮定の下で予測を行うことに相当する。例えば、SAR によるパラメータ推定の結果、空間相関の存在が棄却された($\lambda=0$)場合、式(5)(19)より

$$\hat{\boldsymbol{\beta}}_{OLS}=\hat{\boldsymbol{\beta}} \quad \text{if } \lambda=0 \quad (20)$$

となることから、表-1 に示されている予測量は表-3 と等しくなる。つまり、従来の OLS を用いた予測手法は、本研究で提案した予測手法のうち、空間相関が存在しない場合($\lambda=0$)と位置づけることができる。

また、残差補正法は、仮定Ⅰ-(b)に基づく予測手法に相当することが分かる。同ゾーンにおける誤差が時系列相関を持っていると考えられる場合には、妥当性を持った予測手法であるといえる。

3. 実験と考察

OLS に基づく将来予測手法も含め、本研究で提案した予測手法による予測精度の違いに関して実証的な分析を試みる。

(1) 実験に用いたデータとモデル

東京都心から100km以内の23生活圏に注目し、国土交通省整備のデータベース(TRANET)から、1980・85・90・95年の人口・面積・生活圏間の鉄道所要時間のデータを使用し、人口予測を通して予測精度の比較を行った。用いたモデルは式(21)である。

$$\ln(\text{人口密度})=\beta_0+\beta_1\times\ln(ACC)+u \quad (21)$$

ACC:アクセシビリティ指標

(=1/(東京からの鉄道による所要時間))

また、空間重み行列 $\mathbf{W}$ は幾何距離および時間距離を用いて定義(式(22))し、予測精度の比較を行った。但し、 α は0.1刻みで格子点探索を行った。

$$W_{ij}=(1/(\text{生活圏間距離 or 鉄道所要時間}_{ij}))^\alpha \quad (22)$$

(2) 予測手法による精度の違い

表-4にOLS、幾何距離および社会距離に対して空間相関を構造化した際の、1980年データによるパラメータ推定結果を示す。表-4より、時間距離

表-4 パラメータ推定結果 (()内は t 値)

	尤度	AIC	β_0	β_1	σ^2	λ	α
OLS	-19.4	42.8	6.74 (54.9)	1.70 (7.61)	0.306		
SAR 幾何 距離	-17.5	41.1	6.61 (48.9)	1.84 (8.53)	0.255	1.63 (2.92)	1.3
SAR 時間 距離	-16.0	37.9	5.06 (10.0)	0.78 (2.25)	0.189	0.0527 (26.3)	1.1

定(仮定Ⅱ-(a))による予測手法では、RMSE で比較した場合、「OLS によるパラメータ推定・仮定Ⅰ-(a)を用いた予測」よりも低い予測精度しか得られなかったことから、予測時の距離指標を用いて空間相関を定義する手法には妥当性が無いと考えられる。

一方、残差補正法による予測は、幾何距離に依存する空間相関を考慮したモデルよりも優れた予測

表-5 1980年データによる将来予測結果

パラメータ 推定	仮定Ⅰ	仮定Ⅱ	予測量	1985 年	1990 年	1995 年
OLS	(a) $E(\varepsilon\varepsilon_f')=\mathbf{0}$		$\hat{y}_f = \mathbf{X}_f \hat{\beta}_{OLS}$	0.575 (0.863)	0.596 (0.856)	0.592 (0.862)
	(b) $E(\varepsilon\varepsilon_f')=\sigma^2 \mathbf{I}$		$\hat{y}_f = \mathbf{X}_f \hat{\beta}_{OLS} + (\mathbf{y} - \mathbf{X} \hat{\beta}_{OLS})$	0.274 (0.973)	0.270 (0.974)	0.301 (0.967)
SAR 幾何距離	(a) $E(\varepsilon\varepsilon_f')=\mathbf{0}$	(b) $W_{fij} = W_{ij} = f(d_{ij})$	$\hat{y}_f = \mathbf{X}_f \hat{\beta}$	0.559 (0.863)	0.585 (0.856)	0.578 (0.862)
	(b) $E(\varepsilon\varepsilon_f')=\sigma^2 \mathbf{I}$		$\hat{y}_f = \mathbf{X}_f \hat{\beta} + (\mathbf{y} - \mathbf{X} \hat{\beta})$	0.298 (0.968)	0.292 (0.970)	0.326 (0.962)
SAR 時間距離	(a) $E(\varepsilon\varepsilon_f')=\mathbf{0}$	(b) $W_{fij} = W_{ij} = f(d_{ij})$	$\hat{y}_f = \mathbf{X}_f \hat{\beta}$	1.82 (0.863)	1.87 (0.856)	1.88 (0.862)
	(b) $E(\varepsilon\varepsilon_f')=\sigma^2 \mathbf{I}$		$\hat{y}_f = \mathbf{X}_f \hat{\beta} + (\mathbf{y} - \mathbf{X} \hat{\beta})$	0.124 (0.994)	0.139 (0.994)	0.164 (0.991)
	(b) $E(\varepsilon\varepsilon_f')=\sigma^2 \mathbf{I}$	(a) $W_{fij} = f(d_{fij})$	$\hat{y}_f = \mathbf{X}_f \hat{\beta} +$ $(\mathbf{I} - \hat{\lambda} \mathbf{W}_f)^{-1} (\mathbf{I} - \hat{\lambda} \mathbf{W})$ $(\mathbf{y} - \mathbf{X} \hat{\beta})$	0.499 (0.991)	0.626 (0.990)	0.741 (0.984)

上段 RMSE

下段 ()内 相関係数

に対して空間相関を構造化したモデルでは AIC に改善が見られる。

次に、表-4 の結果をもとにして将来予測を行った予測結果を表-5 に示す。

まず、 $E(\varepsilon\varepsilon_f')=\mathbf{0}$ (仮定Ⅰ-(a))では、パラメータ推定法に関わらず高い予測精度が得られない。今回使用したデータでは、同ゾーンの観測と予測の間に相関関係が存在すると仮定し予測を行うことが妥当性を持つことが分かる。

次に、 $E(\varepsilon\varepsilon_f')=\sigma^2 \mathbf{I}$ (仮定Ⅰ-(b))を用いた場合を比較する。時間距離に対して空間相関の構造化を行い、観測時の時間距離を用いた予測(仮定Ⅱ-(b))が、最も精度の高い予測結果を示した。幾何距離より時間距離に対して空間相関を構造化した方が高い予測精度が得られたことから、今回使用したデータは時間距離に依存する空間相関を持っているといえる。また、予測時の距離指標を用いて $\mathbf{W}_f$ を定義する仮

精度が得られた。残差補正法の実用性の高さが示唆される。ただし、時間距離に対して空間相関を構造化したモデルの予測精度は残差補正法よりも高い。パラメータ λ にどの程度の有意性があれば空間相関を考慮した将来予測を行うべきか、興味深い検討課題である。

参考文献

- 1) 例えば Anselin, L. : Spatial Econometrics: Methods and Models, Kluwer Academic, 1988.
- 2) Goldberger, A. : Best Unbiased Prediction in the Generalized Linear Regression Model, American Statistical Association Journal, Vol.57, pp.369-375, 1962.
- 3) Zimmerman, D.L. and Cressie, N. : Mean Squared Prediction Error in the Spatial Linear Model with Estimated Covariance Parameters, *Annals of the Institute of Statistical Mathematics*, Vol.44, pp.27-43, 1992.