

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

東京大学大学院 学生会員 ○井上 亮
東京大学大学院 正会員 清水 英範

1. はじめに

線形回帰型地域モデリングにおいて誤差の空間相関をどのように考慮するかは、これまで空間計量経済学の分野で議論がなされている¹⁾。しかし、これまでの議論は誤差構造の同定に重きを置いており、推定されたモデルを用いて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。

本稿では、1階の空間的自己回帰モデルを例に、誤差の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、問題の整理と実証的な検討を行っている。

2. 誤差の空間相関構造化手法の概要

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の概要について記す。SARでは、モデル

$$(但し, y:被説明変数, X:説明変数, u:観測誤差)$$

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

に対して、各要素がゾーン間の距離 d_{ij} の関数 $W_{ij}=f(d_{ij})$ として表される空間重み行列 W を用い、

$$u = \lambda Wu + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (2)$$

のように構造化を行う。この構造化により、観測誤差 u の分散共分散行列 V は式(3)のように表せる。

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

また、パラメータ β, λ, σ^2 は最尤法で推定する。

3. 予測手法の概要

誤差に相関が存在する場合の予測手法²⁾について説明する。予測モデル(但し、 y_f :被説明変数の将来予測値、 X_f :説明変数の将来値、 u_f :予測誤差)

$$y_f = X_f \beta + u_f \quad (4)$$

に対して、

・予測誤差の期待値

$$E(u_f) = 0 \quad (5)$$

・観測誤差と予測誤差間の共分散行列

$$E(uu_f') = C \quad (6)$$

という条件の下で予測を行う。

このとき、予測誤差の分散が最小(最良線形不偏予測)となる予測値は式(7)のように表される。

$$\hat{y}_f = X_f \hat{\beta} + C'V^{-1}(y - X\hat{\beta}) \quad (7)$$

4. 予測にあたっての検討課題

式(7)を用いて予測を行うためには、 V, C の推定が必要である。SARで誤差項の分散共分散を構造化する場合、誤差の分散共分散行列 V は式(3)で与えられる。次に、 C を推定するために、観測誤差の場合(式(2))と同様に、予測誤差モデルを構造化する。

(但し、 W_f :予測時の空間重み行列)

$$u_f = \lambda W_f u_f + \varepsilon_f \quad \varepsilon_f \sim N(0, \sigma_f^2 I) \quad (8)$$

この結果、 C は式(9)のように表すことができる。

$$C = E(uu_f') = (I - \lambda W)^{-1} E(\varepsilon \varepsilon_f') [(I - \lambda W_f)']^{-1} \quad (9)$$

しかし、 W_f および $E(\varepsilon \varepsilon_f')$ に対して仮定を置かなければ C を推定することはできない。実用上、以下の仮定が考えられる。

仮定 I. W_f に対する仮定

- (1) 予測と観測の間では空間相関構造は変化しないと仮定し、観測時の距離で定義する。
- (2) 予測時の距離を用いて定義する。

仮定 II. $E(\varepsilon \varepsilon_f')$ に対する仮定

- (1) ε 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。

$$E(\varepsilon \varepsilon_f') = 0 \quad (10)$$

- (2) ε 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の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のみ共分散があると仮定する。

$$E(\varepsilon \varepsilon_f') = \sigma^2 I \quad (11)$$

仮定 I・II により、計4通りの予測手法が考えられる。まず、仮定 I において(1) W_f を観測時の距離で定義する場合、予測値は表1のように表される。

キーワード 将来予測, 線形回帰, 空間相関, 空間的自己回帰モデル

連絡先 〒113-8656 東京都文京区本郷7-3-1 東京大学大学院工学系研究科 社会基盤工学専攻 TEL 03-5841-6129

表1 予測値(仮定 I-(1) 観測時の距離で W_f を定義)

仮定 II	(1) $E(\varepsilon\varepsilon'_f) = \mathbf{0}$	(2) $E(\varepsilon\varepsilon'_f) = \sigma^2 \mathbf{I}$
予測値	$\hat{y}_f = X_f \hat{\beta}$	$\hat{y}_f = X_f \hat{\beta} + (y - X\hat{\beta})$

また, 仮定 I において(2) W_f を予測時の距離で定義する場合, 予測値は表2のように表される.

表2 予測値(仮定 I-(2) 予測時の距離で W_f を定義)

仮定 II	(1) $E(\varepsilon\varepsilon'_f) = \mathbf{0}$	(2) $E(\varepsilon\varepsilon'_f) = \sigma^2 \mathbf{I}$
予測値	$\hat{y}_f = X_f \hat{\beta}$	$\hat{y}_f = X_f \hat{\beta} + (\mathbf{I} - \lambda W_f)^{-1} (\mathbf{I} - \lambda W)(y - X\hat{\beta})$

以上のように $E(\varepsilon\varepsilon'_f) = \mathbf{0}$ (仮定 II-(1))とした場合は, 仮定 I に関わらず, 予測値は $\hat{y}_f = X_f \hat{\beta}$ となる.

また, 幾何距離のように観測時と予測時で変化しない距離に対して空間相関を構造化する場合には, 予測値は表1に示す2通りとなる.

現在, 線形回帰モデルを用いた予測では, 通常最小二乗法(OLS)によりパラメータ推定・将来予測が行われている. この手法は $\lambda=0$ とし予測を行うことに相当する. なかでも, OLS のパラメータ推定時の残差を用いて補正をおこなう手法(仮定 II-(2)に相当)(以後, 残差補正法)がしばしば用いられている.

以下では, OLS に基づく将来予測手法も含め, 実証的な分析を試みる.

5. 実験および考察

関東地方(東京都心部・島嶼部は除く)の26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, 85・90・95年の人口予測を行った. 用いたモデルは式(12)である.

$$\ln(\text{人口密度}) = \beta_0 + \beta_1 \times \ln(\text{ACC}) + u \quad (12)$$

ACC: アクセシビリティ指標

($=1/(\text{東京からの鉄道による所要時間})$)

また, 空間重み行列 W は幾何距離および時間距離を用いて定義(式(13))し, 予測精度の比較を行った.

$$W_{ij} = (1/(\text{生活圏間距離 or 鉄道所要時間}))_{ij}^\alpha \quad (13)$$

パラメータ推定結果を表3に, その結果をもとにした将来予測結果を表4に示す.

まず, $E(\varepsilon\varepsilon'_f) = \mathbf{0}$ (仮定 II-(1))では高い予測精度が得られないことから, 同ゾーンの観測と予測の間に相関関係が存在することが分かる.

一方, $E(\varepsilon\varepsilon'_f) = \sigma^2 \mathbf{I}$ (仮定 II-(2))による予測結果を

表3 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	-29.8	55.7	6.61 (48.8)	1.98 (8.20)	0.433		
SAR 幾何距離	-23.2	50.3	6.54 (55.7)	2.15 (10.1)	0.330	713 (3.25)	10.0
SAR 時間距離	-23.9	51.8	6.10 (18.3)	1.60 (4.70)	0.322	0.036 (7.52)	1.4

表4 将来予測結果 (相関係数)

パラメータ推定法および 予測手法	1985 年	1990 年	1995 年
OLS $\hat{y}_f = X_f \hat{\beta}$	0.829	0.826	0.822
OLS 残差補正法 $\hat{y}_f = X_f \hat{\beta} + (y - X\hat{\beta})$	0.971	0.974	0.953
SAR 幾何距離 $\hat{y}_f = X_f \hat{\beta}$	0.829	0.826	0.822
SAR 幾何距離 $\hat{y}_f = X_f \hat{\beta} + (y - X\hat{\beta})$	0.966	0.969	0.959
SAR 時間距離 $\hat{y}_f = X_f \hat{\beta}$	0.829	0.826	0.822
SAR 時間距離 $\hat{y}_f = X_f \hat{\beta} + (y - X\hat{\beta})$	0.982	0.983	0.978
SAR 時間距離 $\hat{y}_f = X_f \hat{\beta} + (\mathbf{I} - \lambda W_f)^{-1} (\mathbf{I} - \lambda W)(y - X\hat{\beta})$	0.927	0.915	0.904

比較すると, 時間距離に対する空間相関を考慮し観測時の時間距離を用いた予測(仮定 I-(1))が, 最もよい結果を示した. 更なる検討が必要であるが, 空間相関構造は観測時から将来に向けて不変であるとの仮定が妥当であるのかもしれない. また, 幾何距離より時間距離で空間相関を定義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.

残差補正法による予測は, 幾何距離による空間相関を考慮したモデルと同等の精度が得られた. 残差補正法の実用性の高さが示唆される. ただし, 時間距離による空間相関を考慮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. (1988): *Spatial Econometrics: Methods and Models*, Kluwer Academic.
- 2) Goldberger, A. (1962): Best Unbiased Prediction in the Generalized Linear Regression Model, *American Statistical Association Journal*, Vol.57, pp.369-375.