

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 6月 5日現在

機関番号：13901

研究種目：基盤研究(C)

研究期間：2009～2011

課題番号：21530199

研究課題名（和文）

空間誤差モデルに関する実験的・経験的研究

研究課題名（英文）

Experimental and Empirical Investigations on the Spatial Error Model

研究代表者

加藤 尚史 (KATO TAKAFUMI)

名古屋大学・環境学研究科・准教授

研究者番号：70233799

研究成果の概要（和文）：攪乱項が空間自己相関を有する回帰モデルは空間誤差モデルと呼ばれる。自己相関の特定方法に基づいてモデルは2つのタイプに分類される。第1のタイプは相関関数を用いて直接的に定義したものであるのに対して、第2のタイプはウェイト行列を用いて間接的に定義したものである。本研究の目的は、実験的な分析や経験的な分析を通して、空間誤差モデルのパフォーマンスを調べることにある。自己相関を正確に特定することは難しいことから、相対的にパフォーマンスの良いモデルを見つけることは合理的である。推定と予測について方法論的な検討を加えたうえで、2つのタイプのモデルを併用することが有用であることを示す。

研究成果の概要（英文）：A regression model is called a spatial error model when spatial autocorrelation is present in the disturbance. A spatial error model is classified into one of two types on the basis of the method of specification of spatial autocorrelation. The first type is defined using a correlation function, which is a device to specify the autocorrelation in a direct manner, whereas the second type is defined using a weight matrix, which is a device to specify it in an indirect manner. The purpose of the present study is to evaluate the performance of spatial error models using experimental and empirical approaches. As errors cannot be assumed to be absent in any specification, choosing a relatively well-performing model is a good policy in order to allow for such errors. Methodological issues are addressed with regard to both estimation and prediction, and some results are presented that support the combined use of the two types of models.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	500,000	150,000	650,000
2010年度	500,000	150,000	650,000
2011年度	500,000	150,000	650,000
総計	1,500,000	450,000	1,950,000

研究分野：社会科学

科研費の分科・細目：経済学・経済統計学

キーワード：空間自己相関・推定精度・予測精度

## 1. 研究開始当初の背景

回帰モデルにおいて推定や予測を行うときには、攪乱項の自己相関を考慮する必要がある。攪乱項が関係し合うかもしれないこと

は、時系列データを用いる場合、以前から関心を集めていて研究の対象にされてきた。系列自己相関と呼ばれるそうした関係をどのように特定すれば良いかについては、コンセ

ンサスが得られるに至っている。注意すべきことは、横断面データを用いる場合であっても、攪乱項が関係し合うことはあり得るということであろう。空間自己相関と呼ばれるそうした関係が関心を集めるようになったのは近年になってからのことである。

空間自己相関をどのように特定すれば良いかについては、それが系列自己相関と同様なものであるとすると、研究の対象にすることなく、コンセンサスを得ることは容易である。系列自己相関に関する研究の成果を利用することができるからである。残念なことは両者に違いがあることであろう。主な違いは、攪乱項が、時系列データを用いる場合には、或る時点と次の時点の間で一方向的な関係を有するのに対して、横断面データを用いる場合には、複数の地点の間で多方向的な関係を有すると見られることにある。空間自己相関の特定を研究の対象にすることが必要になる。

空間自己相関を考慮した回帰モデルは、空間誤差モデルと呼ばれる。空間自己相関の特定については直接的な方法と間接的な方法が考案されていて、各方法に基づいている様々なモデルが提案されてきた。注目すべきことは、ア・プリオリにモデルを選択することは難しいと見られていることであろう。相対的にパフォーマンスの良いモデルを見つけることは、重要であるにもかかわらず、最近まで試みられてこなかった。

## 2. 研究の目的

直接的な方法と間接的な方法に基づいて空間自己相関を特定したモデルを、それぞれ、タイプ1とタイプ2のモデルと呼ぶことにしよう。先行研究は、各タイプについていくつかのモデルを取り上げて、それらのモデルの中で最も予測精度に優れたモデルを選択して併用することを提案している。推定精度はモデルにかかわらず確保される一方、予測精度はタイプによって優劣を付け難くモデルによって異なることが明らかになったからである。

本研究の目的は、空間誤差モデルについて、実験的・経験的な分析を行うことで、推定や予測の精度を比較することにある。先行研究においては、線形のモデルを対象にして実験的な分析が行われているので、本研究においては、経験的な分析を行ったり非線形のモデルを対象にしたりすることが期待されることになる。

## 3. 研究の方法

空間誤差モデルを  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}$  で表わすことにしよう。 $\mathbf{y}$  と  $\mathbf{b}$ 、 $\mathbf{u}$  は、それぞれ、被説明変数と係数パラメータ、攪乱項に関するベクトルであり、 $\mathbf{X}$  は説明変数に関する行列であ

る。 $\mathbf{u}$  は  $N(\mathbf{0}, \sigma^2 \mathbf{K}(\mathbf{D}; \mathbf{c}))$  に従うと仮定すると、空間自己相関は  $\sigma^2 \mathbf{K}(\mathbf{D}; \mathbf{c})$  の非対角成分に反映されることになる。 $\sigma^2$  は標準的なパラメータであり、 $\mathbf{D}$  は地点間の距離を成分とする行列である。 $\mathbf{K}(\cdot; \mathbf{c})$  は同次数の行列をインプットとアウトプットにとる関数であり、 $\mathbf{c}$  は2つのパラメータからなるベクトルである。それらのパラメータを  $c_1$  と  $c_2$  で表わすことにしよう。

タイプ1のモデルは、 $\mathbf{K}(\mathbf{D}; \mathbf{c})$  の第  $ij$  成分を  $K(D_{ij}; \mathbf{c})$  によって捉えようとするものである。 $K(\cdot; \mathbf{c})$  は相関関数と呼ばれるものであり、 $D_{ij}$  は第  $i$  地点と第  $j$  地点の間の距離である。空間自己相関は、相関関数を定義すれば、直接的に特定されることがわかる。一方、タイプ2のモデルは、 $\mathbf{K}(\mathbf{D}; \mathbf{c})$  を  $\{(\mathbf{I} - c_1 \mathbf{W}(\mathbf{D}; c_2))'(\mathbf{I} - c_1 \mathbf{W}(\mathbf{D}; c_2))\}^{-1}$  によって捉えようとするものである。 $\mathbf{W}(\mathbf{D}; c_2)$  はウェイト行列と呼ばれるものをその行和で基準化したものである。ウェイト行列の第  $ij$  成分  $\mathbf{W}(\mathbf{D}; c_2)_{ij}$  は第  $i$  地点と第  $j$  地点の近接性を表わす。空間自己相関は、ウェイト行列を定義すれば、間接的に特定されることがわかる。

相関関数やウェイト行列の定義を変えれば、いろいろなモデルを提案することが可能になる。モンテ・カルロ法を適用して実験的な分析を行うことは、推定や予測の精度がモデルによってどのように異なるかを検討するにあたって有用である。地点を定めて、 $\sigma^2$  と  $\mathbf{c}$  を与えると、 $\mathbf{u}$  を抽出することができるので、 $\mathbf{X}$  と  $\mathbf{b}$  を与えると、 $\mathbf{y}$  を計算することができる。各モデルを使って人工的なデータを作成したうえですべてのモデルを使って推定や予測を行えば、モデルを比較することが可能になる。推定や予測に際しては一般化最小2乗法を適用することが考えられる。 $\mathbf{c}$  を最尤推定すれば、この方法を実行することができる。

現実的なデータを収集して経験的な分析を行うことは、実験的な分析の結果を検証するにあたって有用である。不動産価格や人口密度の分布については、横断面データを用いた推定や予測が行われていて、空間自己相関が考慮されるようになってきている。不動産価格や人口密度を被説明変数として不動産の特性や都心からの距離を説明変数とし、データを収集したうえでモデルを当てはめることになる。

$\mathbf{y}$  と  $\mathbf{X}$  は、それぞれ、変換後の被説明変数に関するベクトルと説明変数に関する行列を表わし、変換前の被説明変数に関するベクトルと説明変数に関する行列とは異なることにしよう。変換をパラメトリックに定義すれば、 $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}$  を非線形の空間誤差モデルとして扱うことができる。ボックス・コックス変換は、パラメータの値に応じていろいろな変換になるので、モデルをフレキシブルな

ものになることになり、有用である。非線形のモデルについても、人工的なデータや現実的なデータを用いれば、実験的な分析や経験的な分析を行うことができる。推定や予測に際しては最尤推定量や確率論的な予測量を採用することが考えられる。

#### 4. 研究成果

先行研究は、 $K(D_{ij}; c)$  の値を

$$\begin{cases} 1 & [D_{ij} = 0] \\ c_1 \exp(-D_{ij}/c_2) & [D_{ij} > 0] \end{cases}$$

$$\begin{cases} 1 & [D_{ij} = 0] \\ c_1 \exp(-(D_{ij}/c_2)^2) & [D_{ij} > 0] \end{cases}$$

$$\begin{cases} 1 & [D_{ij} = 0] \\ c_1(1 - 3D_{ij}/2c_2 + D_{ij}^3/2c_2^3) & [0 < D_{ij} < c_2] \\ 0 & [D_{ij} \geq c_2] \end{cases}$$

とする一方、 $\mathbf{W}(\mathbf{D}; c_2)_{ij}$  の値を

$$\begin{cases} 0 & [D_{ij} = 0] \\ 1 & [D_{ij} > 0, S(D_{ij}) \leq c_2] \\ 0 & [D_{ij} > 0, S(D_{ij}) > c_2] \end{cases}$$

$$\begin{cases} 0 & [D_{ij} = 0] \\ 1/D_{ij}^{c_2} & [D_{ij} > 0] \end{cases}$$

$$\begin{cases} 0 & [D_{ij} = 0] \\ 1 & [0 < D_{ij} \leq c_2] \\ 0 & [D_{ij} > c_2] \end{cases}$$

$$\begin{cases} 0 & [D_{ij} = 0] \\ \exp(-D_{ij}/c_2) & [D_{ij} > 0] \end{cases}$$

とすることで、7つのモデルを比較している。 $S(D_{ij})$  は第  $i$  地点とその他の地点の距離を昇ベキの順にソートしたときの  $D_{ij}$  の順位である。先行研究が課題としたものを検討することで本研究をスタートさせることにした。

第1の課題は、予測サンプルに含まれる情報を利用することが推定精度を高めるか否かということである。空間自己相関によって推定サンプルと予測サンプルが関係し合うとすれば、そのような関係を推定に反映させることは推定精度を高めることになるかもしれないし、推定精度が高まれば、予測精度も高まることになるかもしれない。肯定的な見方を示す文献の問題点を考慮して実験的な分析を行ったところ、否定的な見方をすることはできるとわかった。この知見については、推定サンプルに含まれる情報だけを利用する一般的なアプローチを支持すると同時に、そのようなアプローチを採用した先

行研究の成果が有用であることを含意することから、国際的な学術雑誌を通して発表されることになった。

第2の課題は、被説明変数が変換されたモデルにおいて被説明変数の予測精度を確保するためにはどうしたら良いかということである。変換をパラメトリックに定義するか否かにかかわらず、変換前の被説明変数の予測精度を確保することは重要である。不動産価格や人口密度の分布を扱った文献をサーベイしてみると、被説明変数を対数変換したモデルが多用されているものの、対数変換前の被説明変数の予測精度は確保されていないことが明らかになった。いくつかの予測量を提案して実験的な分析を行ったところ、或る予測量が精度を確保するにあたって有用であることがわかった。この知見については、不動産価格や人口密度の分布を扱う場合に限らず利用できると同時に、変換後の被説明変数の予測精度を検討することとどまった先行研究の成果を補完することから、国際的な学術雑誌を通して発表されることになった。

第3の課題は、先行研究が併用することを提案しているモデルに優るものを見出すことはできるか否かということである。タイプ1に分類されるモデルは相対的に限られているので、 $\mathbf{W}(\mathbf{D}; c_2)_{ij}$  の値を

$$\begin{cases} 0 & [D_{ij} = 0] \\ \exp(-(D_{ij}/c_2)^2) & [D_{ij} > 0] \end{cases}$$

$$\begin{cases} 0 & [D_{ij} = 0] \\ (1 - 3D_{ij}/2c_2 + D_{ij}^3/2c_2^3) & [0 < D_{ij} < c_2] \\ 0 & [D_{ij} \geq c_2] \end{cases}$$

$$\begin{cases} 0 & [D_{ij} = 0] \\ \max(1 - D_{ij}/c_2, 0) & [D_{ij} > 0] \end{cases}$$

とすることで、タイプ2に分類される3つのモデルを7つのモデルに加えて比較したところ、否定的な見方をすることはできるとわかった。この知見については、先行研究の提案を支持することから、論文にまとめて国際的な学術雑誌に投稿することで、評価を受けることにした。

不動産価格の分布に比べて扱い易いことを考慮して、人口密度の分布を取り上げ、経験的な分析を行うことにした。文献をサーベイしたうえで、線形の空間誤差モデルにおいてはスプライン関数を利用する一方、非線形の空間誤差モデルにおいてはボックス・コックス変換を高次で利用することにした。名古屋市総務局のデータを用いて予備的な分析を行ったところ、線形と非線形にかかわらず、先行研究が併用することを提案した2つ

のモデルのパフォーマンスは良いことがわかった。この結果は、線形のモデルに関する実験的な分析の結果と整合していて、非線形のモデルについても有用であると考えられる。2つのモデルが選択されることを実験的に確認することができれば、そのような有用性を一般化することもできる。残念なことは、実験的な分析を行うにあたって、いくつかの疑問に答えなければならないことが明らかになったことであろう。それらの中には、推定が被説明変数の測定単位に左右されるかもしれないことや、線形のモデルの場合と同様に、決定論的な予測量を採用することができるかもしれないことが含まれる。疑問に答えることが可能になれば、非線形のモデルを対象にして実験的な分析を行ったり、総務省統計局のデータを用いつつ、経験的な分析を進展させたりすることも可能になる。それらの分析の結果については、論文にまとめて国際的な学術雑誌に投稿することで、評価を受けることにしたい。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計2件)

- ① Kato, Takafumi (2012). "Usefulness of the Information Contained in the Prediction Sample for the Spatial Error Model," *Journal of Real Estate Finance and Economics*, in press. 査読有.
- ② Kato, Takafumi (2012). "Prediction in the Lognormal Regression Model with Spatial Error Dependence," *Journal of Housing Economics*, Vol.21, pp.66-76. 査読有.

#### 6. 研究組織

##### (1) 研究代表者

加藤 尚史 (KATO TAKAFUMI)

名古屋大学・環境学研究科・准教授

研究者番号：70233799

##### (2) 研究分担者なし

##### (3) 連携研究者なし