

フロー間の空間的相関を考慮した 負の二項重力モデル

爲季 和樹[†]・堤 盛人[†]

(受付 2011年6月29日; 改訂 11月21日; 採択 12月6日)

要 旨

近年交通や物流などのODデータ(フローデータ)における空間的相関の考慮に関する研究が注目されている。中でも LeSage and Pace(2008)は、海外で空間的相関を考慮したモデルの研究として進展の目覚ましい空間計量経済学的手法により、フロー間の空間的相関を考慮した重力モデルを提案している。彼らのモデルは誤差項に正規分布を仮定しているが、離散データに対する連続分布の当てはめについては従来から統計学的な問題点が指摘されており、離散分布を仮定したモデリングが望ましい。本研究では、離散分布の一つである負の二項分布を仮定した重力モデルをもとに、フロー間の空間的相関を考慮したモデルを提案する。提案したモデルを都道府県間人口移動データに適用した結果、LeSage and Pace(2008)同様、空間的相関を考慮することで、対数尤度や平均二乗誤差が大きく改善することを確認した。

キーワード：フローデータ、負の二項分布、重力モデル、空間計量経済学、空間的相互作用。

1. 研究の背景と目的

人や物の流動(フロー)をモデル化したいわゆる空間的相互作用モデルは、古くは万有引力の法則のアナロジーによる重力モデルから始まり、その後エントロピー最大化モデルや競合着地モデルへと発展していった。空間相互作用モデルの原点である重力モデルは、発地の放出性、着地の吸収性、及び発着地間の分離性によってフローを説明するという理解が容易な構造と、モデルの対数変換によって最小二乗法で推定が行える単純さから、非常に古典的ながらも今なお多くの研究分野で使用されている。

しかしながら重力モデルは、フローの観測値がそれぞれ独立であると仮定しており、空間データ特有の空間従属性を考慮していない。そのため推定によって得られたパラメータは空間的自己相関の影響を受けて正しいパラメータ推定がなされずその信頼性が低下するという問題が生じてしまう。Griffith(2007)によれば、この問題はCurry(1972)によって初めて指摘され、その後カナダの通勤フローに関する研究を行ったGriffith and Jones(1980)でも同じ様にこの問題が確認されている。

このような背景から、フローにおける空間従属性を考慮したモデル化に関する研究が近年盛んに行われている。特にLeSage and Pace(2008)は、空間データに内在する空間従属性を考慮したモデリングとして進展の目覚ましい空間計量経済学的手法を用いたモデルを提案し、海外の

[†] 筑波大学大学院 システム情報工学研究科：〒305-8573 茨城県つくば市天王台 1-1-1

様々な実証研究においてその有用性が示されている。しかしながら彼らのモデルは誤差項に正規分布を仮定した対数正規重力モデルをベースとしている一方で、離散データに対する連続分布の当てはめに関しては古くから Flowerdew and Aitkin (1982) によって統計学的な問題点が指摘されている。

本研究は、離散分布の一つである負の二項分布を仮定した重力モデルを基に、フロー間の空間的相関を考慮したモデルを提案する。

2. フロー間の空間的相関を考慮した重力モデル

2.1 重力モデル

最も古典的な空間的相互作用モデルである重力モデル(無制約)の一般式は次式で表される。

$$(2.1) \quad T_{ij} = kV_i^\alpha W_j^\gamma d_{ij}^{-\beta}$$

T_{ij} は発地 i から着地 j へのフロー量であり、 V_i , W_j はそれぞれ発地 i と着地 j の規模を表す変数で、 d_{ij} は i , j 間の距離である。例えば人口移動の場合は、 T_{ij} に観測された移動人口数、 V_i に移動前地域の人口、 W_j に移動後地域の人口、 d_{ij} に発着地域間の距離を入れてパラメータ k , α , β , γ (>0) を解き人口移動の要因分析等を行う。より詳しい重力モデルの説明に関しては石川(1988)が挙げられる。

(2.1)式はパラメータに対して非線形であるが、両辺を対数変換することで次式のように線形モデルとして表すことができる。

$$(2.2) \quad \ln T_{ij} = \ln k + \alpha \ln V_i + \gamma \ln W_j - \beta \ln d_{ij}$$

本研究では LeSage and Pace (2008) に倣い、(2.2)式を行列表記した次式を用いる。

$$(2.3) \quad \mathbf{y} = \alpha \mathbf{1}_n + \mathbf{X}_o \boldsymbol{\beta} + \mathbf{X}_d \boldsymbol{\gamma} + \boldsymbol{\theta} \mathbf{d} + \boldsymbol{\varepsilon}$$

被説明変数 \mathbf{y} は $n \times n$ の OD 行列を vec オペレータにより $n^2 \times 1$ のベクトルに変換したものである。また \mathbf{X}_o , \mathbf{X}_d , \mathbf{d} はそれぞれ対応するフローにおける $\ln V_i$, $\ln W_j$, $\ln d_{ij}$ を要素とする行列あるいはベクトルである(より詳しい説明は LeSage and Pace, 2008 を参照)。

2.2 空間従属性を考慮した重力モデル

空間従属性を主要なテーマとしている学問分野の一つとして空間計量経済学が挙げられる。空間計量経済モデルでは、データが観測された地域間の近接性を表した空間重み行列を用いた空間ラグ付き内生変数を含む自己回帰モデルによって空間従属性を考慮する。空間計量経済モデルは主に点データや面データに対して用いられることが多く、筆者らが知る限り、フローデータでの適用は LeSage and Pace (2008) 以前には無かった。従来の点・面データでは、二地域間の近接性は n 地域であれば $n \times n$ の空間重み行列で表すことができる。しかしフローではすでに一つの観測値に発地と着地の二地域が含まれるため、二つのフロー間の近接性を表現するには計四地域を考慮しなければならず、空間重み行列での表現が非常に困難であると予想された。しかし LeSage and Pace (2008) は、フローを発地と着地のペアと考えるのではなく、一つの観測値として捉えるという発想により、フローにおける空間重み行列の作成法を提案した。

彼らによれば、フローにおける近接性は、ある発地 i から着地 j へのフローが存在するとき、(a) 発地 i 周辺から着地 j へのフロー、(b) 発地 i から着地 j 周辺へのフロー、そして(c) 発地 i 周辺から着地 j 周辺へのフロー、という三つの場合に分けることができるとしている。ここでは、(a)~(c)それぞれに対応して、「発地ベースの近接性」、「着地ベースの近接性」、「発着地ベースの近接性」と呼ぶこととする(図1参照)。

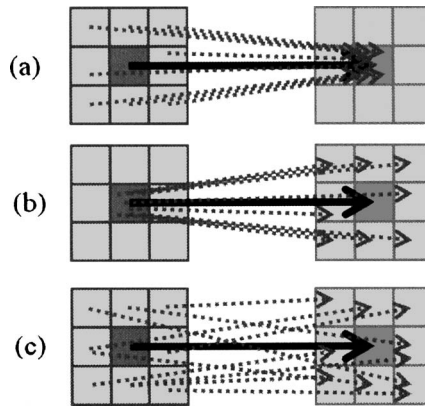


図 1. フローにおける近接性の定義.

これらの近接性は空間重み行列とクロネッカー積を用いて、発地ベース、着地ベース、そして発着地ベースのそれぞれについて

$$(2.4) \quad W_o = W \otimes I_n$$

$$(2.5) \quad W_d = I_n \otimes W$$

$$(2.6) \quad W_w = W \otimes W$$

と定義される。ここで W は n 地域での $n \times n$ の空間重み行列であり、 I_n は $n \times n$ の単位行列である。これら三つの近接性を同時に考慮した際の空間ラグモデルは

$$(2.7) \quad y = \rho_o W_o y + \rho_d W_d y + \rho_w W_w y + \alpha \mathbf{1}_n + X_o \beta + X_d \gamma + \theta d + \varepsilon$$

と表すことができる。(2.7)式の ρ に様々な制約を課すことで種々の空間従属性のパターンを考慮することが可能となるが、LeSage and Pace (2008)は ρ に制約をかけず発地ベース、着地ベース、発着地ベースの空間従属性を識別したモデルがモデル定式化において好ましいことを尤度比検定を用いて明らかにしている。

3. 対数正規重力モデルにおける統計学的問題点

LeSage and Pace (2008)が提案したモデルは、対数正規重力モデルに空間ラグを組み込むことで空間従属性を考慮したものである。この対数正規重力モデルは非線形の重力モデルを対数変換によって線形モデルにし、その誤差項が平均0、分散 σ^2 の正規分布に従うと仮定する。これはフロー観測値が対数正規分布に従うと仮定することと等しい。しかしながらこのような仮定に基づいた対数正規重力モデルの同定にはいくつかの統計学的な問題点が存在することがFlowerdew and Aitkin (1982)によって以下の様に指摘されている。

1. 第一の問題点は、推定の際にフロー観測値の対数が用いられることである。推定によって得られたフロー推定値の逆対数変換には、大きなフロー量を過小推定してしまい、結果としてフロー推定値の合計もまた実際の値より過小になってしまう傾向になる。

2. 誤差項の正規分布の仮定は、フローの観測値が対数正規分布に従うことを意味するが、実際にはこれは当てはまらない。フロー観測値はカウントデータであり非負であることは自明で、かつ整数の値をとることから対数正規分布という連続分布よりも離散分布を仮定する方が望ましい。しかし例外として観測単位がトンやキロといった重量で計られている場合はこの限りではない。

3. 全ての観測値の組み合わせについて等しい分散を持つという誤差項の等分散の仮定は、フローデータでは多くの場合不均一分散または異分散性が認められるため成り立たない。

4. フロー量がゼロの場合、ゼロの対数は $-\infty$ となるため、適当な小さい正の値(例えば0.05や0.1など)で置き換える必要がある。しかしながら、この値をどのように設定するかによりパラメータ推定値が左右されてしまい、適合度が変わってくる。

以上の問題点を挙げた上で、Flowerdew and Aitkin (1982)はこれらを克服するものとして、フロー観測値が離散分布の一つであるポアソン分布に従うと仮定したポアソン重力モデルを提案した。

ポアソン分布の特徴として、期待値と分散が等しいとの仮定があるが、実際には期待値に比べて分散が大きくなる現象が観測される場合が多い。その様なデータに対して期待値と分散が等しいという制約をすると、過分散の問題が引き起こされる。過分散とは、分布として期待しているよりも残差のばらつきが大きくなる状態のことで、推定値は一致性を持つが有効性を持たず標準誤差にバイアスがかかり z 値が大きくなってしまふ。これにより説明変数の有意性を過大に評価することになる。よって期待値よりも分散が大きいという、より仮定を緩くした分布によるモデリングが望ましく、それが負の二項分布である。

4. 負の二項分布を仮定した提案モデル

4.1 負の二項分布

負の二項分布では、あるフロー T_i が起こる確率は下記の式で表される。

$$(4.1) \quad p(T_i) = \frac{\Gamma(T_i + \nu^{-1})}{T_i! \Gamma(\nu^{-1})} \left(\frac{\nu^{-1}}{\nu^{-1} + \mu_i} \right)^{\nu^{-1}} \left(\frac{\mu_i}{\nu^{-1} + \mu_i} \right)^{T_i}$$

ここで $\Gamma(\cdot)$ はガンマ関数で、 ν は正のパラメータである。期待値は $E(T_i) = \mu_i$ で分散は $\text{Var}(T_i) = \mu_i + \nu \mu_i^2$ となる。 ν は定義により正の実数なので、 $E(T_i) < \text{Var}(T_i)$ と期待値より分散が大きいという仮定が成り立つ。対数尤度関数は

$$(4.2) \quad \ln L(\theta, \beta) = \sum_{i=1}^n \left\{ \sum_{j=0}^{T_i-1} \ln(j + \nu^{-1}) - \ln T_i! - (T_i + \nu^{-1}) \ln(1 + \mu_i) - T_i \ln \nu^{-1} + T_i \ln \mu_i \right\}$$

となり、モデルにおけるパラメータ μ 及び ν の最尤推定量は反復重み付き最小二乗法の手順で求めることができる。 $\nu \rightarrow 0$ であれば負の二項分布はポアソン分布で近似することができるので、得られた ν に対し $\nu=0$ の検定を行い、棄却されればポアソン分布による仮定は採用できない。

LeSage and Pace (2008)では対数正規重力モデルを空間計量経済学のアプローチから空間ラグ付き内生変数を組み合わせることで、フローデータにおける空間従属性の考慮を行った。本研究では彼らのアプローチ法を負の二項分布を仮定した重力モデルに対してとることで、LeSage and Pace (2008)のモデルを理論的に発展させたモデルを提案する。

4.2 空間負の二項自己回帰モデルとその推定法

負の二項重力モデル(以下 NB モデル)の μ_i を

$$(4.3) \quad \mu_i = \exp(\alpha \mathbf{1}_{ni} + \mathbf{X}_{oi}\beta + \mathbf{X}_{di}\gamma + \theta d_i)$$

と定義したとき, NB モデルに空間ラグ付き内生変数を含んだ負の二項-空間自己回帰モデル (NB-SAR モデル)は

$$(4.4) \quad \mu_i = \exp(\rho_o \mathbf{W}_{oi}\mathbf{y} + \rho_d \mathbf{W}_{di}\mathbf{y} + \rho_w \mathbf{W}_{wi}\mathbf{y} + \alpha \mathbf{1}_{ni} + \mathbf{X}_{oi}\beta + \mathbf{X}_{di}\gamma + \theta d_i)$$

と表すことができる。

本提案モデルの推定方法には空間的二段階最小二乗法(以下 S2SLS)を採用する。Kelejian and Prucha (1998)による S2SLS は, 分布型に依存せず非正規性に対して頑健であるという長所があり(清水・唐渡, 2007), 空間ポアソン自己回帰モデルを用いた研究を行った Lambert et al. (2010)でも使用されている。本研究では, 具体的には以下の手順によってパラメータ推定を行う。

1. 操作変数 $\mathbf{Q}_o = [\mathbf{X}, \mathbf{W}_o\mathbf{X}, \mathbf{W}_o^2\mathbf{X}]$ を空間ラグ $\mathbf{W}_o\mathbf{y}$ に回帰して, 空間ラグの理論値を計算する。 $\mathbf{W}_d\mathbf{y}, \mathbf{W}_w\mathbf{y}$ についても同様に計算する。

2. ステップ 1 で計算された空間ラグの理論値を説明変数として代入した回帰モデルを, 反復重み付き最小二乗法で推定する。

通常, 説明変数と誤差項との間に相関が存在する場合, OLS 推定量がバイアスを持つ問題は Anselin (1988) などでも周知の事実であるが, 上記の S2SLS の様に誤差項と無相関かつ説明変数と相関がある変数 \mathbf{Q} (操作変数)を用いて回帰した内生変数の理論値は誤差項と無相関となるため, この問題を回避することができる。

上述の通り本研究では $\mathbf{Q}_o = [\mathbf{X}, \mathbf{W}_o\mathbf{X}, \mathbf{W}_o^2\mathbf{X}]$, $\mathbf{Q}_d = [\mathbf{X}, \mathbf{W}_d\mathbf{X}, \mathbf{W}_d^2\mathbf{X}]$, $\mathbf{Q}_w = [\mathbf{X}, \mathbf{W}_w\mathbf{X}, \mathbf{W}_w^2\mathbf{X}]$ と操作変数を設定しているが, (2.7)式における \mathbf{y} の期待値は

$$(4.5) \quad E(\mathbf{y}) = (\mathbf{I} - \rho_o \mathbf{W}_o - \rho_d \mathbf{W}_d - \rho_w \mathbf{W}_w)^{-1}(\alpha \mathbf{1}_n + \mathbf{X}_o\beta + \mathbf{X}_d\gamma + \theta \mathbf{d})$$

であり, レオンチェフ展開を行うと例えば空間ラグ $\mathbf{W}_o\mathbf{y}$ の操作変数の候補には, $\mathbf{W}_o^2\mathbf{X}$ の他にも $\mathbf{W}_o\mathbf{W}_d\mathbf{X}$ といった他の空間重み行列と組み合わせることで発地と着地相互依存性を考慮したものも考えられる。しかしながら, (2.4)~(2.6)式より $\mathbf{W}_w = \mathbf{W}_o\mathbf{W}_d = \mathbf{W}_d\mathbf{W}_o$ であることから, 他の空間重み行列と組み合わせた操作変数を用いるとそれぞれの空間ラグ(特に $\mathbf{W}_w\mathbf{y}$)を識別することができなくなる。したがって各空間ラグにおける操作変数には, それぞれの空間重み行列のみを使用することが必要である。

5. 人口移動データを用いた実証

5.1 使用するデータ

本章では, 提案した NB-SAR モデルを実際のデータに適用し, 従来のモデルとの結果の違いについて考察する。使用する OD データは, 住民基本台帳人口移動報告によって得られた 2006 年の 47 都道府県間人口移動である。発着地の規模を表す説明変数は LeSage and Pace (2008) で用いたものに沿う様に, 2005 年時の人口の対数, 面積(km²), 15 歳未満人口比率, 完全失業者人口比率, 第三次産業就業者人口比率, 及び役員人口比率を選択した。また, 距離変数には, 各県の代表点を人口重心とし, その代表点間の直線距離(m)の対数を用いた。空間重み行列は, 各都道府県間の距離の逆数の二乗を各要素とした距離ベースの重みで定義し, その上で各行和が 1 となるように標準化を行ったものを使用する。

5.2 内々フローの説明変数

(2.3)式と(2.4)式のどちらも地域間流動(内外フロー)と地域内流動(内々フロー)を区別していない。そのためパラメータ推定値は多くの場合、内外フローよりも通常大きな値が観測される内々フローにつられ、パラメータ推定値は内々フローを説明する様な値になる恐れがある。この問題を避けるため、LeSage and Pace (2008)は一つのモデルでありながら、内外フローと内々フローを区別するモデル(以下「調整済みモデル」)の使用を提案している。

$$(5.1) \quad \mathbf{y} = \alpha \tilde{\mathbf{1}}_n + \alpha_i \mathbf{1}_i + \tilde{\mathbf{X}}_o \beta + \tilde{\mathbf{X}}_d \gamma + \mathbf{X}_i \psi + \theta \mathbf{d} + \varepsilon$$

$\mathbf{1}_i$ と \mathbf{X}_i における各要素は、それぞれ $\mathbf{1}_n$ と \mathbf{X}_o の内々フローに対応する行以外がゼロとなっている。チルダ(˜)のついた行列はその逆で、内々フローに対応する行における要素のみゼロとなっている。この内々フローを説明する変数 \mathbf{X}_i は、内外フローの説明変数と同じ変数を含む必要はない。なぜなら調整済みモデルでは、 \mathbf{X}_i によって内々フローを説明することが目的ではなく、内外フローを説明するパラメータに内々フローの情報が影響しないようにすることが目的であるためである。よって本研究では人口の対数、65歳以上人口比率、完全失業者人口比率の三つを内々フローの説明変数として用いる。

5.3 推定結果

表1にNBモデルと、NB-SARモデルのパラメータ推定結果を載せた。空間的自己相関の強弱を示す空間パラメータは発地ベースと着地ベースでそれぞれ $\rho_o = 0.627$, $\rho_d = 0.633$ と同程度の値になっており、強い正の相関を示している。また $\rho_w = -0.593$ は負の空間的自己相関を

表1. NBモデルとNB-SARモデルのパラメータ推定結果.

説明変数	NB-SARモデル		NBモデル	
	係数	z値	係数	z値
定数項	-9.741	-8.148	-25.391	-33.046
I_定数項	4.067	1.335	-8.001	-2.555
O_人口	0.337	7.748	0.905	37.28
O_面積	0.072	2.965	0.204	8.362
O_15歳未満人口	1.254	0.913	6.839	4.884
O_失業者	12.181	4.284	34.202	12.477
O_第3次産業就業者	6.971	5.774	19.231	18.422
O_役員人口	-15.245	-3.427	-33.631	-7.447
D_人口	0.388	8.09	1.021	42.063
D_面積	0.037	1.573	0.116	4.761
D_15歳未満人口	3.665	2.538	12.750	9.101
D_失業者	8.377	3.049	26.252	9.576
D_第3次産業就業者	6.525	5.416	18.52	17.742
D_役員人口	-12.300	-2.805	-25.159	-5.572
I_人口	0.0637	0.408	1.255	8.537
I_65歳以上人口	0.926	0.204	-1.275	-0.265
I_失業者	18.935	1.394	20.450	1.421
距離	-0.393	-7.534	-1.224	-69.788
ρ_o	0.627	14.983		
ρ_d	0.633	15.322		
ρ_w	-0.593	-14.443		
v	3.583	33.802	3.181	34.236
対数尤度	-14650.560		-14790.678	

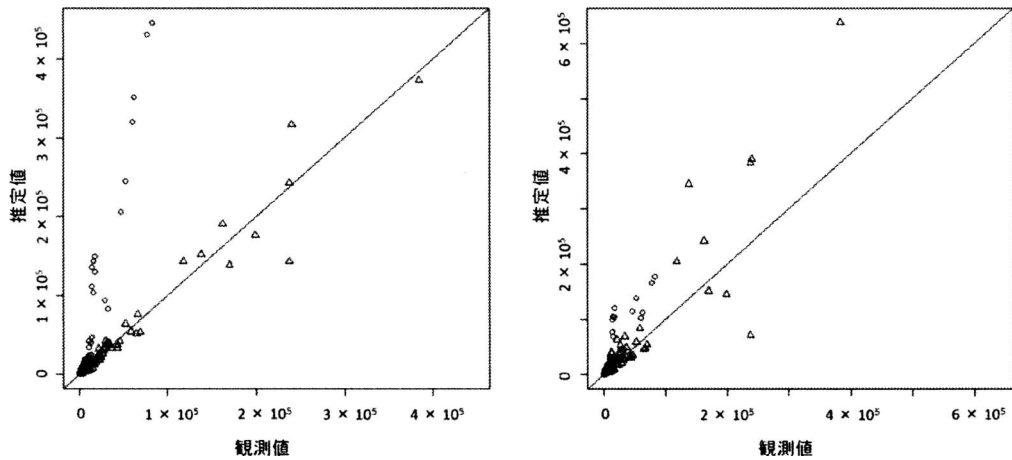


図 2. 推定値と観測値のプロット. (左)NB モデル (右)NB-SAR モデル.

示している. そして ρ_o, ρ_d, ρ_w すべてで 1%水準で有意であることから, 発地ベース, 着地ベース, 発着地ベースの空間従属性を考慮することが適当であることが示唆されている.

距離に係るパラメータは NB モデルと NB-SAR モデルどちらでも 1%水準で有意となっており, 距離減衰パラメータとして機能していることが伺える. 分散パラメータ ν は NB モデルでは 3.181, NB-SAR モデルでは 3.583 と, 空間従属性を考慮することで若干の上昇がみられたが, どちらのモデルでも z 値を見ると帰無仮説 $\nu = 0$ が 1%有意水準で棄却されることを示しており, 本研究で用いた人口移動データにおいて「(期待値) = (分散)」であるポアソン分布を仮定したモデリングは正しくなく, 負の二項分布が採用されるべきであることが分かる.

5.4 推定精度に関する考察

図 2 は観測値に対する推定値をプロットしたものととも 45 度線を引いており, 内外フローが ○ 印, 内々フローが △ 印となっている.

NB モデルの図を見ると, 内外フローの一部がかなり観測値からかけ離れた推定になっている. これらの乖離は, 東京・神奈川間, 東京・埼玉間, 東京・千葉間, とすべて首都圏の都県間におけるフローにおいて見られるという興味深い結果が得られた. この過大推定の原因として, 首都圏ではその他の地方に比べて県が密集しており県間距離が小さいためであると考えられる. それに比べて NB-SAR モデルによる推定では, どの内外フローの推定値も 45 度線付近に沿っており, 首都圏内のフローも観測値に近い推定がなされている. これは, NB モデルでは県間距離と都道府県の説明変数のみでフローを説明するのに対して, 空間従属性を考慮した NB-SAR モデルではそれらに加えて周辺フローの情報が与えられることで, 推定精度が向上したと考えられる.

NB-SAR モデルによって内外フローの推定精度は向上したものの, 内々フローの推定値は NB モデルのものより 45 度線から離れればつきが増している. 特に東京都や千葉県内のフローは NB モデルに比べてかなり当てはまりが悪く過大推定となっている.

NB モデル及び NB-SAR モデルの推定精度を平均二乗平方根誤差 (RMSE) で測ったものが表 2 に表されているが, やはり内外フローの推定誤差は空間従属性を考慮した本提案によって半分以下に抑えられているが, 反面内々フローの推定誤差は悪化していることが分かる. それ

表 2. NB モデルと NB-SAR モデルの RMSE.

	NBモデル	NB-SARモデル
全体	16242.51	11292.11
内外	16142.33	5980.77
内々	20324.12	65936.74

表 3. 対数正規分布を仮定したモデルのパラメータ推定結果.

説明変数	係数	z 値
定数項	-12.777	-10.602
I_定数項	3.466	1.122
O_人口	0.403	9.178
O_面積	0.089	3.641
O_15歳未満人口	3.997	2.904
O_失業者	11.543	4.047
O_第3次産業就業者	7.351	6.056
O_役員人口	-14.279	-3.203
D_人口	0.461	9.546
D_面積	0.055	2.337
D_15歳未満人口	6.214	4.292
D_失業者	9.065	3.292
D_第3次産業就業者	7.148	5.901
D_役員人口	-11.519	-2.621
I_人口	0.140	0.885
I_65歳以上人口	-0.042	-0.009
I_失業者	19.224	1.396
距離	-0.420	-7.983
ρ_o	0.559	14.983
ρ_d	0.573	15.322
ρ_w	-0.522	-14.443

にも関わらず全体の推定精度が向上しているのは、都道府県間人口移動データでは内々フローが47であるのに対し内外フローは2,162と、圧倒的に内外フローの数が多いため、その推定精度の向上につられて全体の RMSE 値も下がったためである。

5.5 対数正規分布を仮定したモデリングとの比較

表 3 は LeSage and Pace (2008) の対数正規分布を仮定したモデル (LL-SAR モデル) のパラメータ推定結果を示している。空間パラメータは $\rho_o = 0.559$, $\rho_d = 0.573$, そして $\rho_w = -0.522$ と NB-SAR モデルの空間パラメータと同様の傾向を見せているが、NB-SAR モデルの方がどの ρ においても絶対値が大きくそれぞれの空間的自己相関の度合いが強いことを示唆している。またその他の説明変数のパラメータ推定値は、有意なものに関しては NB-SAR モデルの方が小さいという結果となっている。LL-SAR モデルと NB-SAR モデルの推定結果の大きな違いとしては、発地の「15歳未満人口」と着地の「面積」の説明変数が NB-SAR モデルでは有意ではなかったにも関わらず、対数正規分布を仮定することでそれぞれ 1% 及び 5% 水準で有意という結果になったことであろう。

LL-SAR モデルと本提案モデルは、連続確率分布と離散確率分布の違いから対数尤度の計算法が異なるため、モデルの良さを測る対数尤度値による直接的な比較を行うことができない。そのため本研究では代わりに RMSE 値で LeSage and Pace (2008) のモデルと本提案モデルによる比較を行う。LL-SAR モデルにおける RMSE 値は全体では 8086.69 と、NB-SAR モデル

よりも低く推定精度が高いことが伺える。内外フロー及び内々フローの RMSE 値もそれぞれ 4,391.33, 46,759.95 と、やはり NB-SAR モデルと比べて値が低い。しかしながら先に述べた様に、RMSE はモデルの良さではなく当てはまりの良さの指標であるため、これは NB-SAR モデルが LL-SAR モデルよりモデルとして劣るという訳ではなく、本研究で用いた人口移動データに関しては対数正規分布を仮定した方が当てはまりは良くなるという結果である。また、LL-SAR モデルでも内々の RMSE 値が空間従属性を考慮していない NB モデルと比べて高いことから、分布型の仮定に関わらず空間従属性を考慮することで内々の推定精度が悪化するという結果が得られた。

6. 結論と今後の課題

本研究は LeSage and Pace (2008) によるフロー間の空間的相関を考慮した重力モデルの統計学の問題点を克服するモデルとして、負の二項分布を仮定した空間モデルを提案した。本提案モデルを都道府県間人口移動データに適用した結果、LeSage and Pace (2008) 同様、対数尤度値が改善され、また推定精度の向上も見ることができた。

しかし、内々フローのみに着目した場合、推定精度は本提案モデルによって悪化する現象が見られた。本研究の目的は、空間的自己相関によるパラメータ推定値の歪みを除去することであり、モデルの当てはまりに関してはまた別の議論が必要となるが、今後内々フローの当てはまりを良くする方法論も検討していきたい。

本提案モデルは LeSage and Pace (2008) のモデルの理論的な発展であるため、空間的な相関性を考慮したのみに留まっており、時間的な相互作用を考慮していない。例えば人口移動流は、経済状況などその年の社会的要因に左右されることが十分に考えられるため、空間のみならず時間による影響も考慮する必要がある。したがって本提案モデルの更なる発展として、空間的相関及び時間的相関の両方を考慮した STAR (Spatio-temporal Autoregressive) model へと改良し、フローのパネルデータ分析を今後の課題としたい。

参 考 文 献

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Kluwer Academic, Dordrecht.
- Curry, L. (1972). A spatial analysis of gravity flows, *Regional Studies*, **6**, 137–147.
- Flowerdew, R. and Aitkin, M. (1982). A method of fitting the gravity model based on the Poisson distribution, *Journal of Regional Science*, **22**, 191–202.
- Griffith, D. A. (2007). Spatial structure and spatial interaction: 25 years later, *The Review of Regional Studies*, **37**(1), 28–38.
- Griffith, D. A. and Jones, K. G. (1980). Explorations into the relationship between spatial structure and spatial interaction, *Environment and Planning A*, **12**, 187–201.
- 石川義孝(1988). 『空間的相互作用モデル：その系譜と体系』, 地人書房, 京都.
- Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Economics*, **17**(1), 99–121.
- Lambert, D. M., Brown, J. P. and Florax, R. J. G. M. (2010). A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application, *Regional Science and Urban Economics*, **40**, 241–252.
- LeSage, J. P. and Pace, R. K. (2008). Spatial econometric modeling of origin-destination flows, *Journal of Regional Science*, **48**(5), 941–967.
- 清水千弘, 唐渡広志(2007). 『不動産市場の計量経済分析』, 朝倉書店, 東京.

Modeling Spatial Dependence in Origin-destination Flows Based on a Negative Binomial Gravity Model

Kazuki Tamesue and Morito Tsutsumi

Graduate School of Systems and Information Engineering, University of Tsukuba

Spatial econometric modeling for origin-destination flows proposed by LeSage and Pace (2008) has succeeded in overcoming the problem of the classical gravity model: non consideration of spatial dependence. The model is based on the log-normal gravity model, which assumes log-normal distribution of observed flow data; however, assuming contiguous distribution over discrete data leads to statistical problems, which results in biased and inefficient estimates, as Flowerdew and Aitkin (1982) pointed out, and modeling that assumes discrete distribution is preferable. This study proposes spatial econometric modeling for origin-destination flows based on a negative binomial gravity model: the spatial autoregressive negative binomial model. The estimation result using the 2006 interregional migration flow data of Japan shows that, compared to the nonspatial negative binomial gravity model, the proposed model improves the log likelihood and root mean squared error. An increase in the log-likelihood value of the model was also seen in LeSage and Pace (2008), indicating that their model is better than the commonly used log-normal gravity model. This similar result suggests that the proposed spatial negative binomial gravity model is credible.