

フロー間の空間的相関を考慮した 負の二項を用いた重力モデル

爲季 和樹¹・堤 盛人²

¹学生非会員 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)
E-mail:s1120539@sk.tsukuba.ac.jp

²正会員 筑波大学大学院准教授 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)
E-mail:tsutsumi@sk.tsukuba.ac.jp

近年OD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の手法により、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。

Key Words : *negative binomial gravity model, spatial econometrics, origin-destination flows*

1. 研究の背景と目的

人や物の流動（フロー）は空間的相互作用とも呼ばれ、その現象をモデル化した空間的相互作用モデルは、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。

Fotheringham (1983)の競合着地モデルは二段階目的地選択の概念を導入することでこの問題を解消したとされるが、菅沼・樋口 (2004) の実証においては二段階目的地

選択は成り立たないと指摘している。

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の手法を用いたモデルを提案し、海外の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

(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。

$$T_{ij} = kV_i^\alpha W_j^\gamma d_{ij}^{-\beta} \quad (1)$$

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。このモデルはパラメータ k 、 α 、 β 、 γ (>0) 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。

$$\ln T_{ij} = \ln k + \alpha \ln V_i + \gamma \ln W_j - \beta \ln d_{ij} \quad (2)$$

本研究ではLeSage and Pace (2008)に倣い、(2) 式を行列表記した次式を用いる。

$$\mathbf{y} = \alpha \mathbf{1}_n + \mathbf{X}_o \boldsymbol{\beta} + \mathbf{X}_d \boldsymbol{\gamma} + \boldsymbol{\theta} \mathbf{d} + \boldsymbol{\varepsilon} \quad (3)$$

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) 空間従属性を考慮した重力モデル

空間従属性を考慮する方法論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、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。

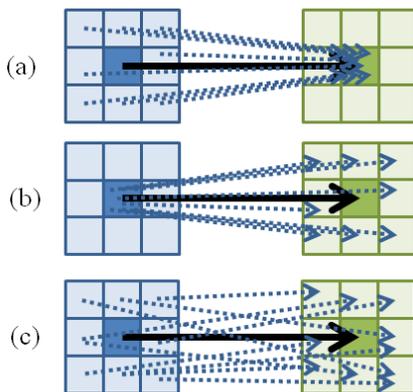


図1 フローにおける近接性の定義

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）。

これらの近接性を空間重み行列で表現することはクロネッカー積によって発地ベース、着地ベース、そして発着地ベースはそれぞれ

$$\mathbf{W}_o = \mathbf{W} \otimes \mathbf{I}_n \quad (4)$$

$$\mathbf{W}_d = \mathbf{I}_n \otimes \mathbf{W} \quad (5)$$

$$\mathbf{W}_w = \mathbf{W} \otimes \mathbf{W} \quad (6)$$

と作成することができる。ここで \mathbf{W} は n 地域での $n \times n$ の空間重み行列であり、 \mathbf{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

$$\mathbf{y} = \rho_o \mathbf{W}_o \mathbf{y} + \rho_d \mathbf{W}_d \mathbf{y} + \rho_w \mathbf{W}_w \mathbf{y} + \alpha \mathbf{1}_n + \mathbf{X}_o \boldsymbol{\beta} + \mathbf{X}_d \boldsymbol{\gamma} + \boldsymbol{\theta} \mathbf{d} + \boldsymbol{\varepsilon} \quad (7)$$

と表すことができる。

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。しかしながらこの仮定に基づいた対数正規重力モデル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) = (分散) 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、推定値は一致性をも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う。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

(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。

$$p(T_i) = \frac{\Gamma(T_i + \nu^{-1})}{T_i! \Gamma(\nu^{-1})} \left(\frac{\nu^{-1}}{\nu^{-1} + \mu_i} \right)^{\nu^{-1}} \left(\frac{\mu_i}{\nu^{-1} + \mu_i} \right)^{T_i} \quad (8)$$

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。

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。

(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

$$\mu_i = \exp(\alpha \mathbf{1}_{ni} + \mathbf{X}_{oi} \boldsymbol{\beta} + \mathbf{X}_{di} \boldsymbol{\gamma} + \theta \mathbf{1}_i) \quad (9)$$

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空間負の二項自己回帰モデル（NB-SARモデル）は

$$\mu_i = \exp(\rho_o \mathbf{W}_{oi} \mathbf{y} + \rho_d \mathbf{W}_{di} \mathbf{y} + \rho_w \mathbf{W}_{wi} \mathbf{y} + \alpha \mathbf{1}_{ni} + \mathbf{X}_{oi} \boldsymbol{\beta} + \mathbf{X}_{di} \boldsymbol{\gamma} + \theta \mathbf{1}_i) \quad (10)$$

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。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 $\mathbf{Q}_0 = [\mathbf{X}, \mathbf{W}_o \mathbf{X}, \mathbf{W}_o^2 \mathbf{X}]$ 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。 $\mathbf{W}_o \mathbf{y}$, $\mathbf{W}_w \mathbf{y}$ 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Anlesin(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。

5. 人口移動データを用いた実証

(1) 使用するデータ

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 (km^2)、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。

(2) 内々フローの説明変数

式(3)と式(4)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 nad 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。

$$y = \alpha \tilde{1}_n + \alpha_i \mathbf{1}_i + \tilde{X}_o \beta + \tilde{X}_d \gamma + X_i \psi + \theta d + \varepsilon \quad (11)$$

$\mathbf{1}_i$ と X_i における各要素は、それぞれ $\mathbf{1}_n$ と 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 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、 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。

(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示して

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	係数
定数項	-9.741 ***	-25.391 ***
I_定数項	4.067	-8.001 *
O_人口	0.337 ***	0.905 ***
O_面積	0.072 **	0.204 ***
O_15歳未満人口	1.254	6.839 ***
O_失業者	12.181 ***	34.202 ***
O_第3次産業就業者	6.971 ***	19.231 ***
O_役員人口	-15.245 ***	-33.631 ***
D_人口	0.388 ***	1.021 ***
D_面積	0.037	0.116 ***
D_15歳未満人口	3.665 *	12.750 ***
D_失業者	8.377 **	26.252 ***
D_第3次産業就業者	6.525 ***	18.52 ***
D_役員人口	-12.300 **	-25.159 ***
I_人口	0.0637	1.255 ***
I_65歳以上人口	0.926	-1.275
I_失業者	18.935	20.450
距離	-0.393 ***	-1.224 ***
ρ_o	0.627 ***	
ρ_d	0.633 ***	
ρ_w	-0.593 ***	

(*** : 1%有意, ** : 5%有意, * : 10%有意)

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0からかなり離れていることから、(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いことが分かる。

両モデル間での説明変数の有意性の評価に着目してみると、空間従属性を考慮してないNBモデルでは発地及び着地における説明変数はどれも1%有意と評価されていたのに対し、NB-SARモデルでは発地の15歳未満人口比率と、着地の面積が説明変数として有意ではないとされている。

対数尤度値は、空間従属性を考慮することによって-14790.7から-14650.6に上昇し、空間従属性を考慮したことでモデルの良さが向上したことが明らかとなった。

(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、内外フローを青色、内々フローを赤色で色分けしている。

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度線から離ればらつきが増している。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。

表2は以上の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)で表したものである。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

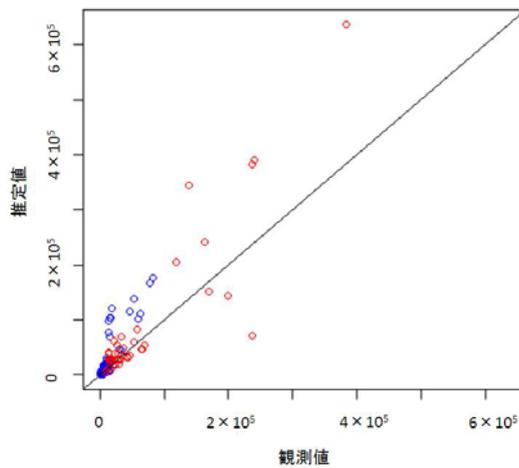
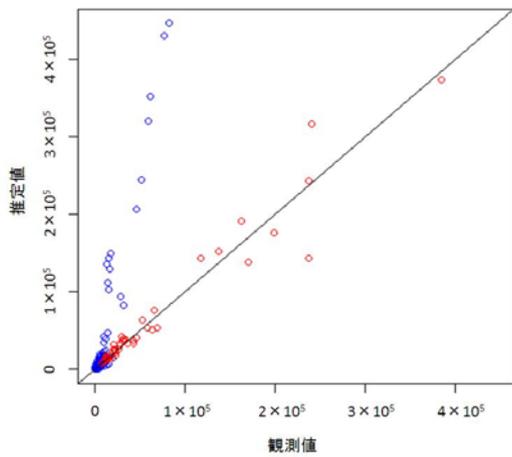


図2 推定値と観測値のプロット
(上) NBモデル (下) NB-SARモデル

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	16242.5	14067.5
内外	16142.3	7251.3
内々	20324.1	82959.6

ローの推定誤差は悪化していることが分かる。それ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。

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的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。また、LeSage and Pace (2008)と本提案モデルでは分布が異なるため、対数尤度を用いて直接モデルの当てはまり具合を比較することができなかったが、今後、例えばモンテカルロシミュレーションによるパラメータ推定を通じ、両モデルの推定誤差を比較することも検討したい。

参考文献

- 1) Griffith, D. A. :Spatial Structure and Spatial Interaction: 25 Years Later, *The Review of Regional Studies*, Vol.37, No.1, 2007.
- 2) Curry, L. :A spatial analysis of gravity flows, *Regional Studies*, Vol.6, 1972.
- 3) Griffith, D. A. and Jones, K. G. :Explorations into the relationship between spatial structure and spatial interaction, *Environment and Planning A*, Vol.12, 1980.
- 4) LeSage, J. P. and Oace, R. K. : Spatial Econometric Modeling of Origin-Destination flows, *Journal of Regional Science*, Vol.48, No.5, pp.941-967, 2008.
- 5) Flowerdew, R. and Aitkin, M. : A method of fitting the gravity model based on the Poisson distribution, *Journal of Regional Science*, Vol.22, pp.191-202, 1982.
- 6) Kelejian, H. H. and Prucha, I. R. :A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Economics*, Vol.17, No.1, 1998.
- 7) 清水千弘, 唐渡広志 : 不動産市場の計量経済分析, 朝倉書店, 2007.
- 8) Lambert, D. M., Brown, J. P., and Florax, R. J. G. M. :A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application, *Regional Science and Urban Economics*, Vol.40, 2010.
- 9) Anselin, L. :Spatial Econometrics: Methods and Models, Kluwer Academic, Dordrecht, 1988.

(2011.5.6 受付)