

固有ベクトル空間フィルタリングを用いた ゼロ過剰重力モデル

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 筑波大学准教授 システム情報系社会工学域 (〒305-8573 茨城県つくば市天王台1-1-1)
E-mail:tsutsumi@sk.tsukuba.ac.jp

重力モデルのパラメータ推定は対数変換した回帰モデルにOLSを適用するのが一般的であるが、ODデータが離散データである場合、対数正規分布ではなく離散分布を仮定した一般化線形回帰モデルを用いる必要がある。さらに、物流センサス等ではフロー量ゼロが多く含まれるため、ゼロ過剰な離散分布を仮定したモデリングが望ましい。一方、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なか、固有ベクトル空間フィルタリングは非常に強力な手法とされており、上述のような特殊な一般化線形回帰モデルでも適用可能である。本研究では、ゼロ過剰離散分布を仮定した重力モデルに固有ベクトル空間フィルタリングを援用したモデルを我が国の物流センサスデータに適用して実証分析を行うとともに、離散分布の種類によるパラメータ推定値の感応度の考察を行う。

Key Words : origin-destination flows, gravity model, zero-inflated, eigenvector spatial filtering

1. 序論

二地域間の人や物の流動（フロー）を説明する研究において、古くから、いわゆる空間的相互作用モデルが用いられてきた。中でも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。例えば、鉄道・航空間の機能分担を考慮した運行頻度最適化モデルの研究を行った村上ら（2006）では都市間鉄道の利便性の評価方法において、また川田ら（2008）では幹線道路の整備効果のための経済波及効果モデルの研究で分布交通量の予測において重力モデルを用いている。しかしフローデータの多くは計数データであることから、このようなデータに対して暗黙に連続型の確率分布を仮定したモデルを利用することへの統計学的な問題点が、古くから Flowerdew and Aitkin (1982) によって指摘されており、離散型確率分布を仮定したモデリングが望まれる。

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Cuny (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)でも同様の問題が確認されている。

この様な背景から、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、Griffith (2007)による固有ベクトル空間フィルタリングを用いたモデリング手法は、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。固有ベクトル空間フィルタリングは一般化線形モデルにも適用が可能な柔軟な方法論であり、Fischer and Griffith (2008) や Chun(2008) では、離散分布の一つであるポアソン分布を仮定したポアソン重力モデルに固有ベクトル空間フィルタリングを適用している。しかし実際のフローデータではフロー量がゼロの地域間が多く含まれる場合があり、仮定した分布で想定するよりもゼロ度数が過剰な際には、それを明示的にモデル化することが望まれるが、筆者らの知る限りその様な研究は見当たらない。

したがって本研究は、ゼロ過剰離散分布を仮定した重力モデル（ゼロ過剰重力モデル）に対して固有ベクトル空間フィルタリングを援用したモデルを提案し、我が国の物流センサスデータに適用して実証分析を行うとともに

に、離散分布の種類によるパラメータ推定値の感応度の考察を行う。

2. 空間従属性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。

$$y_{OD} = s x_O^\eta x_D^\lambda d_{OD}^{-\delta} \quad (1)$$

y_{OD} は発地 O から着地 D へのフロー量であり、 x_O , x_D はそれぞれ発地 O と着地 D の規模を表す変数で、 d_{OD} は O , D 間の距離である。このモデルはパラメータ s , η , λ , δ (>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。

$$\ln y_{OD} = \ln s + \eta \ln x_O + \lambda \ln x_D - \delta \ln d_{OD} \quad (2)$$

式 (2) において誤差項 ε が平均 0、分散 σ^2 の確率分布に従うと仮定し、線形モデルのパラメータ推定法である通常最小二乗法 (OLS) によって未知パラメータを求めることが線形化された重力モデルにおいて一般的に行われている。特に、確率分布として連続分布である正規分布を仮定、即ち $\varepsilon \sim N(0, \sigma^2)$ なる誤差項を仮定する場合には、対数正規重力 (log-normal gravity) モデルの名称が用いられることも多い。

(2) 固有ベクトル空間フィルタリング

空間的相関の検定統計量として式 (3) のモラン I 統計量が知られている。

$$I_{\text{moran}} = \mathbf{y}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{y} / \mathbf{y}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{y} \quad (3)$$

ここで、 n はデータ数、 I_{moran} はモラン I 統計量、 $\mathbf{y}$ は空間的相関を計測する変数のベクトル、 $\mathbf{I}$ は $n \times n$ の単位行列、 $\mathbf{1}$ は 1 を要素に持つ $n \times 1$ のベクトル、 $\mathbf{C}$ は $n \times n$ の近接行列である。但し、簡単のため、以降では $\mathbf{C}$ を隣接行列に限定して議論を進める。 I_{moran} は -1 から 1 までの値をとり、 1 に近いことは $\mathbf{y}$ が強い正の空間的相関を持つことを、 -1 に近いことは $\mathbf{y}$ が強い負の空間的相関を持つことを、それぞれ表わす。モラン I 統計量は近接性を $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ で表現する統計量とみなすことができる。

固有ベクトル空間フィルタリング (ESF) は、隣接行列 $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ の固有ベクトル(2)式を用いて空間的相関を考慮するアプローチである。

$$\{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_n\} = \text{evec}[(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)] \quad (4)$$

ここで、 $\text{evec}[\]$ は $[\]$ 内の行列の固有ベクトルの演算子である。 $\mathbf{E}_1$ は最大の固有値を持つ固有ベクトルであり、モラン I 統計量で説明可能な最も大域的な空間パターンを表す。 $\mathbf{E}_2$ は 2 番目に大きな固有値を持つ固有ベクトルであり、 $\mathbf{E}_1$ と無相関、かつモラン I 統計量で説明可能な

最も大域的な空間パターンを表す。同様に、 $\mathbf{E}_n$ は $n-1$ 番目までの全ての固有ベクトルと無相関、かつモラン I 統計量で説明される最も大域的な空間パターンを表す。以上の固有ベクトルは、ESF の先行研究に倣い（例えば、Griffith and Paelinck, 2011）、大文字で表記することとした。

重力モデルにおける ESF の基本モデルは下式で与えられる。

$$y_i = \sum_k X_{i,k} \beta_k + \sum_l E_{i,l}^* \gamma_l + \varepsilon_i \quad (5)$$

ここで添え字 i は i 番目の発地 O と着地 D のペアを表している。また簡便のため、 $\ln x_O$, $\ln x_D$, $\ln d_{OD}$ を全て含む説明変数行列を $\mathbf{X}$ とおいた。 k は説明変数、 l は固有ベクトルの添え字であり、 $X_{i,k}$, ε_i は、 i の、 k 番目の説明変数、誤差項をそれぞれ表す。 E^* は、発地の固有ベクトル $\mathbf{E}_O = \mathbf{E} \otimes \mathbf{1}$ と着地の固有ベクトル $\mathbf{E}_D = \mathbf{1} \otimes \mathbf{E}$ ベクトルで構成される行列 $E^* = (\mathbf{E}_O, \mathbf{E}_D)$ である。 β_k , γ_l はパラメータを表す。

式 (5) に投入する固有ベクトルは、(i) モラン I 統計量が 0.25 以上の固有ベクトルを予め抽出し、(ii) 抽出された固有ベクトルの中から変数選択を用いて決定することができる (Griffith, 2003)。モラン I 統計量の降順と固有値の降順は完全に一致するため、(i) は固有値の大きな固有ベクトルを抽出することを意味する。変数選択 (ii) は、自由度調整済み決定係数の最大化や、残差のモラン I 統計量の最小化等に基づいて行うことができる。前者は精度の良いモデルの同定に、後者は空間的相関の除去に有効である。

空間計量経済学・空間統計学のモデルは、空間的相関を考慮した結果として共分散が距離の関数で与えられることとなり、共分散が 0 というガウス・マルコフ定理の仮定を満たさないため、通常最小二乗推定が適用できない。一方で、ESF の基本式である式 (5) は、通常線形回帰モデルと同一であり、その共分散は 0 であるため通常最小二乗推定が適用できる。これは一般化線形モデルにおいても同様で、次章以降で紹介する分布型を仮定したモデルでも固有ベクトルを説明変数として投入することで空間的相関を考慮することが可能となる。

3. 離散分布を仮定した重力モデル

(1) 対数正規重力モデルの統計学的問題

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に等しい。しかし、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。

(2) ポアソン分布と負の二項分布

ポアソン分布は二項分布の特殊系で、出現確率がきわめて小さい稀な事象の生起回数のモデリングに広く用いられている。あるフロー量 y_i がポアソン分布に従うと仮定すれば、 y_i の確率関数は $\mu_i > 0$ として次式で与えられる。

$$\Pr(y_i) = \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \quad (6)$$

ここで μ_i はポアソン過程の期待値であり、式(2)に倣い下式によって個々の特徴に依存するものとしてモデル化される。

$$\mu_i = \exp(\alpha \ln x_{O,i} + \gamma \ln x_{D,i} + \beta \ln d_{OD,i}) \quad (7)$$

ポアソン分布を仮定した重力モデルは、構造上は式

(2)の重力モデルと同じであるが、前節で述べた問題点を克服しているため、それを改良したモデルとみなすことができる。この様に、連結関数によって被説明変数と説明変数との間の関係式が簡単な線形式の形ではなく、指数型分布族と呼ばれる正規分布より広いクラスの分布が扱えるモデルが一般化線形モデルである。

ポアソン分布の特徴として、(期待値) = (分散) という仮定があるが、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課すと、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。

負の二項分布では、あるフロー y_i が起こる確率は下式で与えられる。

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

ここで $\Gamma(\cdot)$ はガンマ関数で、 ν は正のパラメータである。期待値及び分散はそれぞれ、 $E(y_i) = \mu_i$ 、 $\text{Var}(y_i) = \mu_i + \mu_i^2$ であり、 $\nu > 0$ から $E(y_i) < \text{Var}(y_i)$ となり期待値より分散が大きいという過程が成り立つ。 $\nu \rightarrow 0$ であれば、負の二項分布はポアソン分布で近似できるので、得られた ν の推定値に対し $\nu=0$ の検定を行い、棄却されればポアソン分布による仮定を採用することはできない。

4. ゼロ過剰なデータでのモデル

(1) ゼロ過剰モデル

カウントデータでは、イベントが発生しないケースに加え、発生はしても調査の方法によっては観測されないケース、つまりカウントがゼロとなるケースが多いデータも多く存在する。フローデータでは、例えば短期間におけるデータや、集計単位が市町村レベルに細かく設定されたデータ等では、フロー量がゼロとなる発着地ペアが多く発生することが考えられる。こういったカウントデータを扱う際に用いられるのがゼロ過剰モデルである。

負の二項重力モデルは条件付き期待値を変えずに条件付き分散を増加させることで、ポアソン重力モデルにおけるゼロ度数の過小評価に対処している。これに対し、ゼロ過剰モデルでは、フロー量ゼロが発生する場合を明示的にモデルで構造化する。すなわち、 $y_i = 0$ の場合と、 $y_i > 0$ の場合とで異なるモデルを仮定し、この二つを組み合わせるのである。

フロー量 y_i がポアソン分布に従うと仮定すれば、 y_i の確率関数は式(6)で与えられ、さらにフロー量ゼロが確率 ψ で発生する。 ψ はロジットまたはプロビットモデルによって次式の様に構造化される。

$$\psi_i = F(Z_i \gamma) \quad (9)$$

ここで F は正規分布もしくはロジスティック分布の累積密度関数である（詳しくはLong, 1999 参照）。 Z はカウ

ントゼロにおける説明変数である．式 (6) と式 (9) を組み合わせることで，ゼロ過剰ポアソン (ZIP) モデルの確率関数は

$$\begin{aligned}\Pr(y_i = 0) &= \psi_i + (1 - \psi_i) \exp(-\mu_i) \\ \Pr(y_i > 0) &= (1 - \psi_i) \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!}\end{aligned}\quad (10)$$

で表される．同様に，フロー量が負の二項分布に従うと仮定すれば，式 (8) と式 (9) を組み合わせることで，

$$\begin{aligned}\Pr(y_i = 0) &= \psi_i + (1 - \psi_i) \left(\frac{\nu^{-1}}{\nu^{-1} + \mu_i} \right)^{\nu^{-1}} \\ \Pr(y_i > 0) &= (1 - \psi_i) \frac{\Gamma(y_i + \nu^{-1})}{y_i! \Gamma(\nu^{-1})} \left(\frac{\nu^{-1}}{\nu^{-1} + \mu_i} \right)^{\nu^{-1}} \left(\frac{\mu_i}{\nu^{-1} + \mu_i} \right)^{y_i}\end{aligned}\quad (11)$$

とゼロ過剰負の二項 (ZINB) モデルにおける確率関数が表される．

ZIP 及び ZINB モデルにおける期待値は

$$E(y_i) = [0 \times \psi_i] + [\mu_i \times (1 - \psi_i)] = \mu_i - \mu_i \psi_i \quad (12)$$

であり， $\mu\psi$ の分だけモデルの期待値が減少していることが分かる．

さらに，分散に関しては，ZIP モデルでは

$$\text{Var}(y_i) = \mu_i (1 - \psi_i) (1 + \mu_i \psi_i) \quad (13)$$

となり， $\psi = 0$ であればポアソン重力モデルと等しく， $\psi > 0$ であれば期待値よりも分散が大きい仮定となることが分かる．そして ZINB モデルでは，

$$\text{Var}(y_i) = \mu_i (1 - \psi_i) [1 + \mu_i (\psi_i + \nu)] \quad (14)$$

となり， $\psi = 0$ であれば負の二項重力モデルと等しく， $\psi > 0$ であれば負の二項重力モデルよりも大きな分散を仮定する．

(2) 最尤法

いま，

$$\varpi_i = \begin{cases} 0 & (y_i = 0) \\ 1 & (y_i > 0) \end{cases}$$

なる変数を定義すると，ZIP モデルの対数尤度関数は次式で与えられる．

$$\begin{aligned}\log L &= \sum_i (1 - \varpi_i) \log \{ \psi_i + (1 - \psi_i) \exp(-\mu_i) \} \\ &+ \sum_i \varpi_i \{ \log(1 - \psi_i) - \mu_i + y_i \log \mu_i - \log(y_i!) \}\end{aligned}\quad (15)$$

同様に，ZINB モデルの対数尤度関数は，

$$\log L = \sum_i (1 - \varpi_i) \log \left\{ \psi_i + (1 - \psi_i) \left(\frac{\nu^{-1}}{\nu^{-1} + \mu_i} \right)^{\nu^{-1}} \right\}$$

$$\begin{aligned}&+ \sum_i \varpi_i \{ \log(1 - \psi_i) - \log(y_i!) - \log \Gamma(\nu^{-1}) + y_i \log \mu_i \\ &+ \log \Gamma(y_i + \nu^{-1}) + \nu^{-1} \log \nu^{-1} - (\nu^{-1} + y_i) \log(\nu^{-1} + \mu_i) \}\end{aligned}\quad (16)$$

と表される．これらの対数尤度関数を最大にする最尤解を求めることで，それぞれのモデルのパラメータ推定値が得られる．

5. 実証分析

(1) 用いるデータ

本章では，提案したゼロ過剰重力モデルを実際のデータに適用し，ポアソンと負の二項によるパラメータ推定値の感応度の考察を行う．使用するフローデータは，全国貨物純流動調査（物流センサス）の2010年の品類別都道府県間流動量（件数）である．なお，域内流動が域間流動に比べて値が高く，パラメータが域間ではなく域内流動に依存してしまうことを避けるため，本研究では域間流動のデータのみを用いることにした．発着地の規模を示す説明変数には，2009年次の人口，面積，対人口就業者比率，そして2009年の一人当たり県民所得のそれぞれ対数を取ったものを選択した．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の対数を用いた．さらに，ESFにおける隣接行列 **C** の定義には，距離の逆数ベースを使用する．パラメータ推定には数値計算による最尤推定を用いるため，ステップワイズ法で全ての固有ベクトルから変数選択を行うと収束せずに解が求まらない状況に陥ってしまう．したがって，発地と着地の固有ベクトル上位五つをそれぞれ選択し，強制投入法によって推定を行った．

(2) 推定結果

a) ゼロの割合による推定結果の感応度

まず初めに，ゼロが含まれる割合によって推定結果にどのような影響があるかを考察するため，物流センサスの品類別流動量の全8品類の中から，鉱産品，農水産品，金属機械工業品と，それぞれゼロの割合が異なる3品類のデータで推定を行う．各データにおけるゼロの割合及び記述統計量は表1の通りである．ゼロの割合が最も多いのは鉱産品で，データの81.7%をゼロが占めている．

表1 3品目の記述統計量

	鉱産	農水産	金属機械工業
ゼロ割合	81.7	28.6	4.2
第1四分位	0.0	0.0	70.0
第2四分位	0.0	34.0	363.5
第3四分位	0.0	192.8	1,691.3
第4四分位	10,206.0	28,947.0	108,098.0

表2 品目別各モデルのAIC

品目	ポアソン	負の二項	ZIP	ZINB
鉱産				
non-spatial	398,653	5,808	264,849	5,537
spatial	378,593	5,805	246,011	5,497
農水産				
non-spatial	1,788,553	22,757	1,617,418	22,251
spatial	1,583,259	22,601	1,412,804	21,997
金属機械				
non-spatial	5,217,473	32,778	5,153,405	32,578
spatial	4,416,646	32,669	4,359,938	32,429

金属機械工業品はゼロの割合が最も低く、4.2%のみとなっている。また、3品類全てのデータにおいて、中央値（第2四分位）と最大値（第4四分位）が大きく離れており、データの分散が非常に高いことが表1から推察できる。

ポアソン、負の二項、ZIP、そしてZINBモデルの、空間的相関を考慮しない（non-spatial）モデルと、ESFによって空間的相関を考慮した（spatial）モデル、計8つのモデルで推定を行った。表2はそれぞれのモデルのAICである。

3品目すべてのデータによる推定結果において、ポアソン、ZIP、負の二項、ZINB、の順にAICが低くなっており、フロー量ゼロが含まれる割合に関わらずZINBモデルが最も良いモデルであるということが示唆されている。さらに、non-spatialとspatialで比べた場合には、spatialモデルの方が全ての推定結果においてAICが低い。ため、ESFによって空間的相関を考慮したことでモデルの性能が向上したことが伺える。

ポアソンと負の二項二つの分布型でAICに大きな差が生じた原因の一つは、データの分散度合であることが考えられる。表1からどのデータも分散が大きいことが示唆されており、ポアソンの（期待値）＝（分散）という強い仮定と整合せず過分散の問題が起きたのであろう。

b) パラメータ推定結果

全てのデータにおける全モデルのパラメータ推定結果を載せることは難しいため、本節では表2の結果を踏まえ、最もAICの低いZINBモデルを中心に考察を行う。また、品目による大きな差異は見られなかったため、代表的に農水産品のデータにおける結果のみを記す。

まず初めに空間的相関の有無によるパラメータ推定値の差異を考察するために、表3と表4にそれぞれZINBのnon-spatialとspatialでのパラメータ推定結果を示した。簡便のため投入した固有ベクトルのパラメータ係数は省いている。表中におけるO_iは発地の説明変数、D_iは着地

表3 ZINBモデルの推定結果

	係数	係数(ゼロ)
定数項	0.326 ***	−53.245 ***
O _i 県民所得	0.532	12.719 ***
O _i 人口	0.349 ***	−3.087 ***
O _i 面積	1.854 ***	0.943 ***
O _i 就業率	0.014 ***	−4.690 ***
D _i 県民所得	0.848	1.369
D _i 人口	0.018 ***	−1.441 ***
D _i 面積	0.706	0.122
D _i 就業率	−0.886	−1.656
距離	−0.870 ***	1.314 ***
ν	0.419 ***	

(*:5%有意 **1%有意 ***0.1%有意)

表4 ZINBモデルの推定結果

	係数	係数(ゼロ)
定数項	−21.886 ***	−26.187 *
O _i 県民所得	1.309	8.795 ***
O _i 人口	0.264 *	−2.772 ***
O _i 面積	0.191 *	0.528 **
O _i 就業率	2.353 ***	−3.643 **
D _i 県民所得	0.052	1.312
D _i 人口	1.011 ***	−1.430 ***
D _i 面積	−0.153	0.154
D _i 就業率	0.152	−1.815
距離	−0.944 ***	1.400 ***
ν	0.490 ***	

の説明変数であることを表している。また、「係数（ゼロ）」は、フロー量ゼロにおける係数である。spatial及びnon-spatialいずれのモデルでも、フロー量>0の推定値（以下「推定値」と記載）では、発地の一人当たり県民所得と、着地の一人当たり県民所得、面積、就業率率が有意ではない結果となった。また、spatialに比べ、non-spatialでは発地の人口と面積の有意性が大きく評価されている。フロー量ゼロのパラメータ推定値（以下、「推定値（ゼロ）」と記載）も同様の結果となっている。

続いて分布型によるパラメータ推定値の違いを考察するため、ZIP（spatial）モデルのパラメータ推定結果を表5に示す。ZIPモデルでは推定値はすべて1%有意であるが、ZINBモデルでは、発地の一人当たり県民所得と着地の一人当たり県民所得、面積、就業率率が有意ではない結果となっている。これは第3章で述べたように、過分散によってバイアスが生じ、ZIPモデルでは説明変数の有意性が過大に評価されたためと考えられる。負の二項分布の分散を示すパラメータνは0.49で、1%有意となっていることから、ポアソン分布を仮定したモデリングは望ましくないことが示唆されている。推定値（ゼロ）では有意性の違いは見受けられなかったが、これは

表5 ZIPモデルの推定結果

	係数	係数(ゼロ)
定数項	-36.686 ***	-20.044 *
O_県民所得	3.383 ***	7.924 ***
O_人口	0.396 ***	-2.204 ***
O_面積	0.053 ***	0.325 *
O_就業率	0.846 ***	-4.179 ***
D_県民所得	0.664 ***	0.824
D_人口	0.751 ***	-1.340 ***
D_面積	-0.078 ***	0.160
D_就業率	0.497 ***	-1.465
距離	-0.865 ***	1.304 ***

フロー量が全てゼロであるため(期待値) = (分散) の仮定が保たれ過分散の問題が生じなかったためと考えられる。係数(ゼロ)はフロー量が発生しない要因であることから、一般的に推定値(ゼロ)は推定値と符号が逆転するが、一人当たり県民所得は係数(ゼロ)においても正の値となっている。つまり、発地または着地の一人当たり県民所得が高いほどフロー量は増えるが、同じくフロー量がゼロになる要因にもなりうる、という興味深い結果となっている。

6. 結論

本研究は、カウントデータにおいてゼロが多く含まれるデータを扱う際に用いられるゼロ過剰モデルに、空間的相関を考慮する手法の一つである固有ベクトル空間フィルタリングを援用したモデルを提案した。古くから重力モデルを対数変換し最小二乗法で推定することへの統計学的な問題は指摘されており、離散型確率分布を仮定した重力モデルが使用されてきたが、ゼロ過剰なデータに着目した研究は行われてこなかった。

ゼロの割合が異なる3種類の物流センサスのデータに本研究のモデルを適用した結果、

本研究では、域内流動が域間流動と比べて異質であるためデータから除外して推定を行ったが、今後は域内流動をモデル内でいかに対処するかが課題となる。ゼロ過剰重力モデルでは、フロー量ゼロが発生する確率を構造化しているが、この手法を援用して域内流動が発生する

確率を明示的にモデルに組み込むことが考えられる。またその際には、域内の距離をどのように設定するかも今後の課題である。

参考文献

- 1) 村上直樹, 竹内太郎, 奥村誠, 塚井誠人: 航空との補完的サービスを考慮した最適鉄道運行計画, 土木計画学・論文集, Vol.23, No.3, pp.629-634, 2006.
- 2) 川田圭吾, 廣島康裕, 宮田譲, 中西仁美: 三遠地域における道路整備による経済波及効果の計測手法の開発, 土木計画学研究・論文集, Vol.25, No.2, pp.363-372, 2008.
- 3) Griffith, D.: Spatial structure and spatial interaction: 25 years later, *The Rev Reg Stud*, Vol.37, No. 1, pp.28-38, 2007.
- 4) Curry, L.: A spatial analysis of gravity flows, *Reg Stud*, Vol.6, No.2, pp.131-147, 1972.
- 5) Griffith, D., Jones K.: Explorations into the relationships between spatial structure and spatial interaction, *Env Plann A*, Vol.12, No.2, pp.187-201, 1980.
- 6) 岩崎学: カウントデータの統計解析, 朝倉書店, 2010.
- 7) Flowerdew R., Aitkin, M.: A method of fitting the gravity model based on the Poisson distribution, *Journal of Regional Science*, Vol.22, pp.191-202.
- 8) Scott, L. J., Regression models for categorical and limited dependent variables, Sage Publications, California, 1997.
- 9) Chun, Y.: Modeling network autocorrelation within migration flows by eigenvector spatial filtering, *Journal of Geographical Systems*, Vol.10, No.4, pp.317-344, 2008.
- 10) Fischer, M. M., Griffith, D. A.: Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union, *Journal of Regional Science*, Vol.48, No.5, pp.969-989, 2008.
- 11) Greene, W. H.: Accounting for excess zeros and sample selection in Poisson and negative binomial regression models, Working Paper No.94-10. New York: Stern School of Business, New York University, Department of Economics.
- 12) Lambert, D.: Zero-inflated Poisson regression, with an application to defects in manufacturing, *Technometrics*, Vol.34, No.1, pp.1-14, 1992.

(2012.5.7 受付)