カウントデータにおける空間計量経済モデル

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: tamesue.kazuki@sk.tsukuba.ac.jp

[2]正会員　筑波大学教授　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

一般的な空間計量経済学モデルは計量経済学の線形回帰モデルをベースに発展がなされてきた．ただし，データの生成過程に着目すると，分析の対象とするデータが必ずしも正規分布に従わないことも考えられる．本研究ではそのようなデータの中でも特にカウントデータを対象とした空間計量経済学の適用を考える．ポアソン分布を対象とした取り組みはLambert et al. (2010) やSellner et al. (2013)が存在するが，モデル構築や推定法はそれぞれ大きく異なり，カウントデータの空間計量経済モデルが確立されているとは言い難い．本研究では，上述の先行研究に加え，時系列カウントデータの自己回帰モデルのレビューを踏まえ，カウントデータにおける空間計量経済モデルの構築と推定法に関する議論を行う．

Key Words : *count data, spatial lag model, Poisson, spatial econometrics*

1.　序論

空間計量経済モデルには，空間ラグモデル，空間エラーモデル等複数のモデル族が存在するが，それらはすべて線形回帰モデルをベースとし，それに空間ラグ項加わる形をとっている．統計モデルで分析を行う際には，観測されたデータのデータ生成過程を仮定し，仮定された確率分布に則したモデルを構築することが一般的である．線形回帰モデルの場合は，データが正規分布に従うと暗に仮定しているが，分析の対象とするデータが必ずしも正規分布に従うとは限らない．非負の整数の値のみをとるカウントデータの場合は，連続分布ではなく離散分布をデータ生成過程として仮定することが自然であろう．

特に，LeSage and Pace (2008)が重力モデルの空間ラグモデルを提案して以降，人流や物流といったODフローデータに対する空間計量経済モデルによる分析が近年注目をあびている（例えば，Fishcer and Wang, 2011）． ODフローはカウントデータである場合が多く，LeSage and Pace (2008)を契機に空間計量経済学の分野でカウントデータを分析対象とする機会は増えたと考えられる．

例えば重力モデルにおいても，データの対数をとって線形回帰モデルで分析することは今でも行われているが，古くは Flowerdew and Aitkin (1982) で指摘されているように，人口移動や交通トリップのように非負の整数の値をとるODデータに対しては，データ生成過程として正規分布ではなくポアソン分布を仮定した重力モデルが望ましいことが知られている．

空間計量経済学では，離散選択モデルを対象とし空間的自己相関を導入した研究は一定数存在するが（例えば，McMillen, 1992; Chakir and Parent, 2009; Smirnov, 2010），カウントデータを対象とした研究は非常に少なく，その方法論が確立されているとは言い難い．そこで本論文では，カウントデータを対象とした空間計量経済モデルの確立の示唆を得るためにこれまでに取り組まれた先駆的な研究のレビューと整理を行い，カウントデータにおける空間計量経済モデルの構築と推定法に関する議論を行う．

本稿の流れは次の通りである．まず第２章で，カウントデータにおける空間ラグモデル（以下カウント空間ラグモデル）の構築を試みた既存研究のレビューを行い，モデルの定式化，パラメータ推定法，そしてパラメータ解釈に関してその特性を整理する．続いて第３章では，計量経済学における時系列カウントデータの自己回帰モデルを紹介し第２章で取り上げた既存研究との関連性を整理する．そして第４章にてまとめと今後の展望について述べる．

2.　空間計量経済学における既存研究

(1)　モデルの定式化

空間計量経済学の枠組みでの中で，カウントデータに

対するモデリングに取り組んだ研究はLambert et al. (2010)が代表的である．彼らは，観測データのデータ生成過程がポアソン分布に従うと仮定した下での空間ラグモデルの構築及びパラメータ推定法を提案した．

今，観測データを$\boldsymbol{y}=(y_1,\dots,y_n)$としたとき，$\boldsymbol{y}$が平均$\boldsymbol{\mu}$のポアソン分布に従うと仮定すると，その確率密度関数は

$$\frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!} \tag{1}$$

と表すことができる．一般的なポアソン回帰モデルにおいては，$\boldsymbol{\mu}$は説明変数行列$\mathbf{X}$とその係数パラメータ$\boldsymbol{\beta}$を用いて

$$\mu_i = \exp(\boldsymbol{X}_i\boldsymbol{\beta}) \tag{2}$$

としてモデルが構築される．Lambert et al. (2010)では，これに空間ラグ項を加えた

$$\mu_i^* = \exp(\rho\sum_{j\neq i} w_{ij}\ln y_j + \boldsymbol{X}_i\boldsymbol{\beta}) \tag{3}$$

を，ポアソン分布の期待値として定式化しカウントデータにおける空間ラグモデルとして提案した．ここでw_{ij}は$n\times n$の空間重み行列$\mathbf{W}$の要素である．式（3）は，$\mathbf{I}$を$n\times n$の単位行列とし$\mathbf{A}=\mathbf{I}-\rho\mathbf{W}$とおくと

$$\mu_i^* = \exp(\boldsymbol{A}_i^{-1}\boldsymbol{X}_i\boldsymbol{\beta}) \tag{4}$$

と表現できる．$\mathbf{A}^{-1}$は空間乗数（spatial multiplier）やレオンチェフ逆行列として知られ（Anselin, 2002），観測地域間のフィードバック効果を表現する役割を持つ．これは一般的な空間ラグモデル

$$\boldsymbol{y} = \rho\mathbf{W}\boldsymbol{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{5}$$

が

$$\boldsymbol{y} = \mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{A}^{-1}\boldsymbol{\varepsilon} \tag{6}$$

として表現可能であるように，Lambertらのモデルは空間ラグモデルと同様にパラメータの意味解釈等が行える．

しかしながら，式（3）の右辺には$\ln y$が含まれているため，yがゼロをとるとき$\ln y$が定義できない問題がある．そこでLambert et al. (2010)では，アドホックに微小値（例えば0.1や0.01）を加える処置や，加える微小値そのものを一つのパラメータとして推定する方法のほかに，Burbidge et al. (1988) で提案された逆双曲線サイン（inverse hyperbolic sine; IHS）変換を処置法として提示している．IHS変換の関数は次式の通りである：

$$g(y_j;\theta) = \log(\theta y_j + \sqrt{1+\theta^2 y_j^2})\big/\theta = \sinh^{-1}(\theta y_j)\big/\theta\,. \tag{7}$$

ここでθは非負のスケーリングパラメータである．$\theta\to 0$であれば$\sinh^{-1}(\theta y_j)$はy_jに近づき，$\theta\to\infty$であれば$\sinh^{-1}(\theta y_j)$は0に近づく．なお，パラメータθは最尤法によって推定することができ（Burbidge et al., 1988），近年では例えばEl-Osta et al. (2007)における実証研究においてゼロ値問題の対処法として用いられており，その実用性・有用性が確認されている．

Sellner et al. (2013) は，LeSage and Pace (2008) の空間ラグ重力モデルを出発点として，ポアソン分布を仮定した重力モデルに空間ラグを組み込んだモデルを提案した．

以下では，前節のLambertらのモデルとの比較を容易とするために対象データをODフローではなく面（点）データとして考えるが，モデルの本質的な解釈に影響はない．

Sellner et al. (2013) では，観測データ$\boldsymbol{y}$は空間的にフィルターされた変数$\boldsymbol{y}^*$と剰余の空間変数$\tilde{\boldsymbol{y}}$に分解されると考える．すなわち$\boldsymbol{y}=\tilde{\boldsymbol{y}}+\boldsymbol{y}^*$であり，$\tilde{\boldsymbol{y}}$は

$$\tilde{\boldsymbol{y}} = \rho\mathbf{W}\boldsymbol{y} \tag{8}$$

として与えられると仮定する．さらに，$\boldsymbol{y}^*$がポアソン分布に従うと仮定し，その平均が式（2）によって構造化される．したがってSellner らのモデルは次式の様に表される：

$$y_i = \tilde{y}_i + y_i^* = \rho\mathbf{W}_i y_i + \exp(\boldsymbol{X}_i\boldsymbol{\beta})\,. \tag{9}$$

もし$\rho=0$であれば式（9）は従来のポアソン回帰モデルとなる．これは Lambert らのモデルでも同じである．しかしながら，Lambert et al. (2010) では観測されたデータに対してデータ生成過程を仮定しているのに対し，Sellner et al. (2013) では，観測データではなく，「潜在的なデータ生成過程」(Sellner et al., 2013) として$\boldsymbol{y}^*$にデータ生成過程を仮定している．したがって，観測データがポアソン分布に従っている必要はなく，連続変数でも適用可能である．このことから，Sellnerらのモデルは厳密にはカウント空間ラグモデルではなく，非線形空間ラグモデルとして扱うのが妥当であると言えよう．

Lambertらのモデルでは式（3）の右辺に$\boldsymbol{y}$の対数が含まれることによるゼロ値問題が存在したが，Sellnerらのモデルではそれが無いことが式（9）より確認できる．したがって，ODフローのように値がゼロを取るケース

が多いカウントデータに対しては，式（7）のようにHIS変換を行うなどの煩雑さが無いことからLambertらのモデルよりもSellnerらのモデルの方が優位である．

(2) パラメータ推定法

前節で述べた様に，両モデルの最も大きな相違点は，観測データに対してデータ生成過程の過程の有無である．これは統計モデルにおいてパラメータ推定の方法にも影響を及ぼす．

Lambert et al. (2010) では $\boldsymbol{y}$ にポアソン分布を仮定しており，その確率密度関数は

$$\frac{\exp(-\mu_i^*)(\mu_i^*)^{y_i}}{y_i!} \tag{10}$$

で表される．したがって対数尤度関数も下式のとおり導出することが可能である：

$$\ln L = \sum_i y_i \ln \mu_i^* - \mu_i^* - \ln(y_i)!. \tag{11}$$

尤度関数が導出できるため，パラメータは最尤法で推定することができる．しかしLambert et al. (2010) は自身らのモンテカルロ実験において，完全情報最尤推定では解を求めることが困難であったという理由で，制限情報最尤推定によるパラメータ推定を提案している．これは即ち空間計量経済モデルにおける空間二段階最小二乗法（Kelejian and and Prucha, 1998）に等しい．

一方で，Sellner et al. (2013)は $\boldsymbol{y}$ のデータ生成過程を仮定していないため，その尤度関数を導出することはできない．そのためSellnerらは，Lambert et al. (2010)の様に最尤法によるパラメータ推定ではなく，二段階非線形最小二乗法によるパラメータ推定を提案している．このことからも，Sellnerらのモデルが厳密にはポアソン空間ラグモデルではないことが伺える．$e_i = y_i - \boldsymbol{A}_i^{-1}\exp(\boldsymbol{X}_i\boldsymbol{\beta})$ とおいたとき，パラメータ $\boldsymbol{\delta} = (\rho, \boldsymbol{\beta})$ の非線形最小二乗推定量は $\hat{\boldsymbol{\delta}} = \min \boldsymbol{e}'\boldsymbol{e}$ と表されるが，このとき $\boldsymbol{y}$ のデータ生成過程がポアソン分布ではなく負の二項分布や正規分布であっても評価関数の勾配は等しく推定法に影響を与えない．

(3) パラメータの解釈

空間ラグモデルの特徴として，$\mathbf{A}^{-1}$ によってある地域のショックが $\mathbf{W}$ で定義された周りの地域に波及していくスピルオーバー効果を表現していることが挙げられる．また，空間ラグ項の導入によって，通常の回帰モデルと同様にパラメータを解釈することはできないことも知られている．これに関する説明や具体例については，LeSage and Pace (2009) や瀬谷・堤（2014）などが詳しいが，観測データが互いに独立であると仮定する通常の回帰モデルとは異なり，ある地域における説明変数の変化が，自地域だけではなく近隣地域にも影響を与えるのである．なお，空間エラーモデルではスピルオーバー効果を仮定しないので，パラメータの意味解釈は通常の回帰モデルと同様に行える．空間ラグモデルでは，ある地域における説明変数の変化が自地域へ与える影響を直接効果，他地域へ与える影響を間接効果と呼ばれる（e.g., 瀬谷・堤, 2014）．Lambert et al. (2010) 及び Sellner et al. (2013) のモデルにおいても，直接効果及び間接効果がそれぞれ導出されている．

Lmabertらのモデルでは，直接効果は

$$\frac{\beta_k}{n}\sum_i a_{ii}^{-1}\mu_i^*, \tag{12}$$

間接効果は

$$\frac{\beta_k}{n}\sum_i \sum_{j\neq i} a_{ij}^{-1}\mu_j^* \tag{13}$$

である．ここで a_{ij}^{-1} は $\mathbf{A}^{-1}$ の要素である．一方で，Sellnerらのモデルにおける直接効果は

$$\frac{\beta_k}{n}\sum_i \frac{a_{ii}^{-1}\mu_i}{\sum_j a_{ij}^{-1}\mu_i}, \tag{14}$$

間接効果は

$$\frac{\beta_k}{n}\sum_i \sum_{j\neq i} \frac{a_{ij}^{-1}\mu_j}{\sum_q a_{iq}^{-1}\mu_q} \tag{15}$$

である．二つのモデルで共通する部分が多いが，通常の空間ラグモデルにおける直接・間接効果と比較して非線形性の高い定式化となっている．

3. 時系列カウントデータの自己回帰モデル

空間計量経済モデルは，計量経済学における時系列モデルのアナロジーとして時間方向の系列相関を空間方向に置き換えたモデルとしてみることができる．そこで本節では，時系列カウントデータにおける自己回帰モデルの研究に着目し，そのモデル定式化とLambert et al. (2010) と Sellner et al. (2013) のモデルとを比較しその位置付けの整理を行う．

カウント自己回帰モデルのレビューに関しては Cameron and Trivedi (2013) の一節で詳しく取り上げられている．まず最も単純なカウント自己回帰モデルとして考えられるのは，条件付き期待値においてラグ項 y_{t-1} を加え

$\exp(\rho y_{t-1} + \boldsymbol{X}_t \beta)$ として定式化することである．ただし，$\rho > 0$ のとき $\rho y_{t-1} \geq 0$ となり条件付き期待値が時間と共に増加していてしまう（Blundell et al., 2002）．そこでラグ項を $\ln y_{t-1}$ とすることが考えられるが，この場合はLambert et al. (2010) と同じ問題として，$y_{t-1} = 0$ のときに調整を行う必要がある．この様に，単純にラグ項を加えることが困難であるために様々なカウント自己回帰モデルが開発されてきた．現時点で確立されたモデルは存在しないと Cameron and Trivedi (2013) は述べてはいるが，その代表的な研究としてZeger and Qaqish (1988) を挙げている．

Zeger and Qaqish (1988) ではより一般的な議論として一般化線形回帰モデルにおける p 次のラグを想定しているが，ここでは単純化のために一次のラグの場合のみを考える．その場合のZeger-Qaqishモデルの条件付き期待値は

$$\mu_{t|t-1} = \exp(\rho \ln y^*_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta}) \tag{16}$$

となる．これは式（3）のLambert et al. (2010)のモデルと同様の定式化であることが分かる．Cameron and Trivedi (2013)は

$$y^*_{t-1} = \max(c, y_{t-1}) \qquad 0 < c < 1 \tag{17}$$

として $y_{t-1} = 0$ のときのみ値を置き換える場合と，全ての値に

$$y^*_{t-1} = y_{t-1} + c \qquad c > 0 \tag{18}$$

として加える場合の対処法を紹介しているが，Lambertらのモデルと同じくゼロ値問題を抱えている点が課題となっている．

この様に式（17）や式（18）のようにアドホックな修正を避けるために，

$$\mu_{t|t-1} = \rho y_{t-1} + \exp(\boldsymbol{X}_t \boldsymbol{\beta}) \tag{19}$$

としたモデルも提案されている．これにより $\rho > 0$ ならば条件付き期待値は正となり，また，Blundell et al. (2002) で指摘されているような時間と共に条件付き期待値が爆発的に増加することはない．こちらのモデルは式（9）のSellner らのモデルと類似していることが伺える．

以上で紹介した通り，空間計量経済モデルが時系列の自己回帰モデルから発展していったように，Lambert et al. (2010) と Sellner et al. (2013)それぞれのモデルも，計量経済学におけるカウント自己回帰モデルがベースとなっていることが分かる．ただし通常の線形回帰モデルとは異なり，カウントデータモデルにおいては自己回帰モデルの派生形が複数存在し，その方法論が確立されていない．

4　結論

カウントデータにおける空間計量経済モデルの適用として，Lambert et al. (2010) や Sellner et al. (2013) の先駆的な取り組みが存在するが，各々のモデルは古くから計量経済学で取り組まれてきたカウント自己回帰モデルをベースとしたモデルとなっている．データ生成過程に着目した際には，観測データに対して確率分布を仮定しているLambert et al. (2010) のモデルがカウントデータにおける空間ラグモデルとして自然な定式化であると考えられる．しかし空間ラグ項が対数をとっているため，ゼロ値問題への対処としてIHS変換等の対処を行う必要がある．Sellner et al. (2013) ではゼロ値問題が無いが，観測データに対してデータ生成過程を仮定できない．この点は，パラメータ推定のバイアスを調べるためにモンテカルロ実験を行う時等に影響を及ぼす．本論文では紹介したカウント自己回帰モデル以外にも，Shepard (1995) やDavis et al. (2003) 等多くのモデル特定化が存在するため，カウント空間ラグモデルの確立に向けて今後はこれらのモデルも含めて体系化を行う必要性があると考える．ただしCameron and Trivedi (2013) が述べているようにカウント自己回帰モデルも現時点で確立されたモデルが存在しないことから，十分な検討が必要である．また，本論文ではカウントデータのデータ生成過程としてポアソン分布を取り上げたが，モデル構築にあたっては負の二項分布等でも同様であり，今後カウントデータにおける方法論が確立することで一般化線形回帰モデルの空間計量経済モデルの構築や推定法の確立へと発展していくことが期待される．

謝辞：本研究は JSPS 特別研究員奨励費 251786 の助成を受けたものである．

参考文献

1) Anselin, L.: Under the hood issues in the specification and interpretation of spatial regression models, *Agricultural Economics*, 17, pp.247 – 267, 2002.
2) Blundell, R., Griffith, R., Windmeijer, F.: Individual effects and dynamics in count data models, *Journal of Econometrics*, 108, pp.113 – 131, 2002.
3) Cameron, A.C., Trivedi, R.K., *Regression Analysis of*

Count Data scond edition, Cambridge University Press, 2013.

4) Chakir, R., Parent, O.: Determinants of land use changes: A spatial multinomial probit approach, *Papers in Regional Science*, 88, pp.327 – 344, 2009.

5) Davis, R.A., Dunsmuir, W.T.M., Streett, S.B.: Observation-driven models for Poisson counts, *Biometrika*, 90, pp.777 – 790, 2003.

6) El-Osta, H., Mishra, A., Morehart, M.: Determinants of economic well-bieng among U.S. farm operator households, *Aglicultural Economics*, 36, pp.291 – 304, 2007.

7) Fischer, M.M. Wang., J.: *Spatial data analysis: Models, Methods and Techniques*, Berlin, Heidelberg, and NewYork: Springer, 2011.

8) 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.

9) Kelejian, H.H., Prucha, I.R.: A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics*, 17, pp.99 – 121, 1998.

10) Lambert, D.M., Brown, J.O., 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 Economis*, 40, pp.241 – 252, 2010.

11) LeSage, J.P., Pace, K.: Spatial econometric modeling of origin-destination flows, *Journal of Regional Science*, 48, pp.941 – 967, 2008.

12) LeSage, J.P., Pace, K.: *Introduction to Spatial Econometrics*. Boca Raton, London and New York: CRC press, Taylor & FrancisGroup.

13) McMillen, D.P.: Probit with spatial autocorrelation, *Journal of Regional Science*, 32, pp.335 – 348, 1992.

14) Sellner, R., Fischer, M.M., Koch, M.: A spatial autoregressive Poisson gravity model, *Geographical Analysis*, 45, pp.180 – 201, 2013.

15) Shepard, N.: *Generalized linear autoregression*, Nuffield College, Oxford University, pre-print, 1995.

16) Smirnov, O.A.: Modeling spatial discrete choice, *Regional Science and Urban Economics*, 40, pp.292 – 298, 2010.

17) Zeger, S.L., Qaqish, B.: Markov regression models for time series: a quasi-likelihood approach, *biometrics*, 44, pp.1019 – 1031, 1988.

18) 瀬谷創，堤盛人：空間統計学―自然科学から人文・社会科学まで―，朝倉書店，2014.

（2015. 4. 24 受付）