

令和 2 年 6 月 15 日現在

機関番号：14501

研究種目：若手研究(B)

研究期間：2017～2019

課題番号：17K14738

研究課題名(和文)"Dirty"な空間データに対する統計分析手法の開発と応用

研究課題名(英文)The development of statistical models for dirty spatial data

研究代表者

瀬谷 創 (SEYA, HAJIME)

神戸大学・工学研究科・准教授

研究者番号：20584296

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：教科書な空間計量経済モデルは、データ欠損がなく、属性にエラーがなく、観測位置が正確という"clean"な空間データを仮定しており、"dirty"なデータには、そのまま適用することが難しい。本研究では、dirtyな空間データに対する統計分析手法の開発を試みた。具体的には、(1) 非ランダム欠損への対処法であるHeckmanのサンプルセレクションモデル、(2) 外れ値に頑健な(無条件)分位点回帰モデル、をそれぞれ援用して、空間データの持つ空間的自己相関という性質と不完全性の両者を考慮したモデルの構築とパラメータ推定方法の提案を行った。

研究成果の学術的意義や社会的意義

空間データ分析に用いられる空間計量経済モデルは、欠損がなく、属性にエラーがなく、観測位置が正確という"clean"な状態であることが仮定されており、"dirty"なデータには、そのまま適用することが難しい。本研究では、dirtyな空間データに対する統計分析手法の開発を試みた。具体的には、非ランダム欠損への対処法であるHeckmanのサンプルセレクションモデル、外れ値に頑健な分位点回帰モデル、をそれぞれ援用して、空間データの持つ空間的自己相関という性質と不完全性の両者を考慮したモデルの構築とパラメータ推定方法の提案を行った。開発されたモデルは、幅広い実証研究に用いることが可能である。

研究成果の概要(英文)：Standard spatial econometric models assume that data is "clean" in that spatial data is complete and accurate in terms of attributes and locations. However, of course, real spatial data is "dirty." We attempted to develop spatial econometric models which are applicable to dirty spatial data. More specifically, we proposed two new models: (1) A spatial sample selection model, (2) A spatial unconditional quantile regression model, and the parameter estimation methods for these models.

研究分野：土木計画学，空間統計学，空間計量経済学

キーワード：欠損データ サンプルセレクション 空間データ 分位点回帰 空間計量経済学 実証分析

1. 研究開始当初の背景

近年、誰でも編集可能な地図情報である OpenStreetMap のような“Volunteered”な空間情報の利活用が急激に進展している。また、携帯電話の位置情報のような“Citizens as Sensors(CAS)”による大規模な空間ビッグデータの利活用の進展も著しい(以下、あわせて V/CAS 空間データと呼称する)。わが国では、携帯電話の位置情報が、いくつかの民間会社を通して集計・販売されており、数年に1度の統計調査情報では把握が困難であった人々の時空間移動やその変動が、リアルタイムに近い形で把握できるようになりつつある(この点は、新型コロナウイルスに関する各種報道により、広く世に知られることになったであろう)。

しかしながら、これらの V/CAS 空間データは、巧妙に設計された国の各種統計や都市交通計画の基礎となるパーソン・トリップ調査とは異なり、標本抽出がランダム(あるいは系統的)に行われたものとはなっていない。また、完全性(漏れ・過剰がない)、位置正確度、主題(属性)正確度といった、地理情報標準に基づく空間データの品質基準を満たすものでもない。したがって分析者の視点からみると、V/CAS 空間データは、サンプルに偏りがあり(ランダムでない欠損を含み)、位置や属性の正確度に不安がある、“dirty”なデータである。このような新たなデータを用いなければ不可能な実証分析が多い一方、dirty であることを前提とした空間データの統計分析に関する研究は、現在までのところほとんどなされていない。

ところで、空間データの持つ空間的自己相関(距離に依拠する類似性)と呼ばれる性質をモデル化するための理論と方法は、社会科学に端を発する空間計量経済学と呼ばれる分野で発展してきた(Yamagata and Seya, 2019)。データに空間的自己相関が存在する場合に、それを無視したモデルを適用するとパラメータの信頼性が低下し、政策的な含意も必然的に異ってくる。そのため、対処法としての空間計量経済モデルは、実証分析ツールとして一定程度重要な位置を踏んでいる。

しかし、申請者らの既往研究を含めて、教科書な空間計量経済モデルは、データ欠損がなく、属性にエラーがなく、観測位置が正確という“clean”な空間データを仮定しており、V/CAS 空間データのような“dirty”なデータには、そのまま適用することが難しい。このような問題意識は、空間計量経済学分野における唯一の国際学会である Spatial Econometrics Association の学会長である Giuseppe Arbia 氏らが 2016 年に発表した論文“Dirty spatial econometrics”のそれとも共通するものである。しかし、当該論文は、注意喚起を促すにとどまっており、dirty な空間データの統計分析のための方法論の提案は行われていない。また、空間データの欠損補完に関する研究は空間計量経済学の分野でもいくつか行われてきたが、それらのほとんどは統計学で Missing at Random として知られるランダムな欠損パターンを前提としたものであり、V/CAS 空間データのような非ランダムな欠損データには適用できない。昨今の V/CAS 空間データの蓄積を鑑みれば、空間的自己相関と不完全性を持つ dirty な空間データに対する統計分析手法の開発は喫緊の課題であるといえる。

2. 研究の目的

このような研究背景の認識のもとに、本研究では、dirty な空間データに対する統計分析手法の開発を試みた。具体的には、(1) 非ランダム欠損への対処法である Heckman のサンプルセレクションモデル、(2) 外れ値に頑健な分位点回帰モデル、をそれぞれ援用して、空間データの持つ空間的自己相関という性質と不完全性(ここでは、clean でないという意味)の両者を考慮したモデルの構築とパラメータ推定方法の提案を行った。なお、(2) については、近年標準的な条件付分位点回帰モデルの解釈上の困難性が指摘されているため、Firpo et al. (2009) [Econometrica, 77 (3)] によって提案された無条件分位点回帰モデルを用いることとした。

3. 研究の方法

(1) 非ランダムな欠損データのための空間計量経済モデルの開発

空間データは、過去→未来という一方向的な影響関係がある時系列データと異なり、互いに影響を及ぼしあうという特徴がある。このような相互作用を、空間計量経済モデルでは、観測されたデータ(標本)間の関係として、空間重み行列を用いてモデル化する。したがってデータ欠損が存在する場合、実現象における相互作用とモデル化される相互作用が異なることとなる。例を挙げると、今 A, B, C という3標本があり、A と B, B と C 間に依存関係があるとすると、このとき、実現象では B を介して A と C の間にも依存関係が存在するが、データにおいて B が欠損していると、A と C の間の依存関係がモデルから捨象されてしまう(除外ネットワークの問題)。したがって、本研究では B の欠損における非ランダム性を考慮しつつ、B を復元して A, B, C 全体でパラメータを推定する方法を採用した。具体的には、計量経済学や医学統計学における代表的な非ランダム欠損への対処法であるサンプルセレクションモデルを援用することで、両問題を解決するモデルとパラメータ推定手法の提案を行った。

本テーマに関連する既往研究として、Wang and Lee (2013) [The Econometrics Journal, 16 (1)] と、

Flores-Lagunes and Schnier [Journal of Applied Econometrics, 27 (2)] が挙げられる。前者は、除外ネットワークの問題を考慮したパラメータ推定法を提案したが、欠損パターンの非ランダム性を考慮していない。後者は、空間計量経済モデルにおいてサンプルセレクションの問題を考慮したが、除外ネットワークの問題を考慮していない。これらでは、パラメータ推定法として標準的な計量経済学的手法（GMM や非線形最小二乗法）が用いられている。一方、両問題を考慮したパラメータ推定方法として、Doğan and Taspınar (2018) [Oxford Bulletin of Economics and Statistics, 80 (1)] が挙げられる。このアプローチは、ベイズ MCMC 法により欠損データの補完とパラメータ推定（パラメータの条件付き事後分布からのサンプリング）を繰り返すものであるが、サンプルセレクションを考慮する相関パラメータがほとんど再現できないという課題があった。そこで本研究では、Doğan and Taspınar (2018) の MCMC アルゴリズムの修正を試みた。

(2) 外れ値に頑健な空間計量経済モデルの開発

V/CAS データの中でも特に、ユーザが直接属性値を回答する形式の場合、往々にして外れ値が生じ得る。標準的な回帰モデルは、分布の（条件付）期待値を推定するものであるため、外れ値の存在に脆弱である。これに対して、条件付中央値などの分位点を推定する分位点回帰が、近年実証経済学や情報処理分野でロバストな計量方法として確立し、著しく進展している。本研究では、中でも解釈上の利点がある Firpo et al. (2009) によって提案された無条件分位点回帰モデルの空間計量経済モデルへの援用を試みた。

4. 研究成果

(1) 非ランダムな欠損データのための空間計量経済モデルの開発

3.(1) で述べた通り、本研究では、Doğan and Taspınar (2018) のアルゴリズムを修正する形で、除外ネットワークとサンプルセレクションの両者を考慮したモデルの構築を行った。モデルは、空間計量経済学の代表的なモデルである空間ラグモデルと、データ欠損有無に関するプロビットモデルを誤差相関を考慮しながら同時推計するシンプルなものである。なお、空間計量経済モデルとして空間ラグモデルを用いている点は、空間誤差モデルを用いた Doğan and Taspınar (2018) との相違点である。Doğan and Taspınar (2018) では、潜在変数の埋め込みにおいて誤差相関が考慮されていない部分があったため、この点の改良を試みた。

比較実験の結果、空間的自己相関が強すぎない場合（空間パラメータが 0.5 以下）であれば、データ欠損の割合が比較的高くても（40%）、良好な精度で、誤差相関を含む各パラメータが推定できることが明らかになった。これによって、除外ネットワークとサンプルセレクションの問題を同時に抱えるようなデータを用いた実証研究への一つのアプローチが提示できたと考えている。モデルの構造や、実験結果については、Seya et al. (2020) を参照されたい。

(2) 外れ値に頑健な空間計量経済モデルの開発

3.(2) で述べた通り、本テーマでは、Firpo et al. (2009) によって提案された無条件分位点回帰モデルの空間計量経済モデルへの援用を試みた。Firpo et al. (2009) が指摘する通り、従来の条件付分位点回帰モデルには、推定されたパラメータの解釈が難しいという課題があり、その点の改善を試みたのが、無条件分位点回帰モデルである。本研究では、(a) ランダム効果固有ベクトル空間フィルタリングを用いたアプローチ（Murakami and Seya, 2019）と、(b) 空間ラグモデルを用いたアプローチ（Seya et al., 2020）の 2 つを考案した。(a) と (b) の 2 つのアプローチは、空間的自己相関の考慮方法とパラメータの推定法が違っている。前者の空間的自己相関の考慮方法については、(a) は空間的自己相関を排除してパラメータからバイアスを取り除くことを志向したアプローチであり、(b) は、空間的自己相関の発生構造をモデル化するアプローチである。一方、後者のパラメータ推定方法について、(a) は制限付き最尤法で、(b) はベイズ MCMC 法でパラメータの推定を行っている。

(a)、(b) のアプローチを、それぞれ実証分析に用いた。具体的には、前者は洪水リスクのヘドニック分析に、後者はコインパーキング間の空間的価格競争に適用した。(b) のアプローチでは、分位点ごとに空間パラメータの推定値が得られるため、空間的な競争の度合いのマーケットによる差異をとらえることができる。したがって、土木計画学分野だけでなく、経済学等他分野における実証研究においても有用なアプローチであると考えている。モデルの構造や、実験結果については、Murakami and Seya (2019) および Seya et al. (2020) を参照されたい。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 1件/うちオープンアクセス 0件）

1. 著者名 Daisuke Murakami and Hajime Seya	4. 巻 30
2. 論文標題 Spatially filtered unconditional quantile regression: Application to a hedonic analysis	5. 発行年 2019年
3. 雑誌名 Environmetrics	6. 最初と最後の頁 e2556
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1002/env.2556	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Hajime Seya	4. 巻 27
2. 論文標題 Parameter estimation of the spatial panel stochastic frontier model with random effects	5. 発行年 2020年
3. 雑誌名 Applied Economics Letters	6. 最初と最後の頁 248-253
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1080/13504851.2019.1613491	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Hajime Seya, Kay W. Axhausen, Makoto Chikaraishi	4. 巻 -
2. 論文標題 Spatial unconditional quantile regression: Application to Japanese parking price data	5. 発行年 2020年
3. 雑誌名 The Annals of Regional Science	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1007/s00168-020-00987-3	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Hajime Seya, Masashi Tomari, Shohei Uno	4. 巻 -
2. 論文標題 Parameter estimation in spatial econometric models with non-random missing data	5. 発行年 2020年
3. 雑誌名 Applied Economics Letters	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1080/13504851.2020.1758618	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Daisuke Murakami and Hajime Seya	4. 巻 -
2. 論文標題 Spatially filtered unconditional quantile regression: Application to a hedonic analysis	5. 発行年 2018年
3. 雑誌名 Environmetrics	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1002/env.2556	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計3件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 瀬谷創・安田悠人
2. 発表標題 商店街の空き店舗率をどの程度に抑えるべきか
3. 学会等名 土木計画学研究・講演集
4. 発表年 2018年

1. 発表者名 泊将史・瀬谷創
2. 発表標題 大規模小売店舗の出店が既存店舗の売上げに与える影響に関する実証的検証
3. 学会等名 第26回 地理情報システム学会 学術研究発表大会
4. 発表年 2017年

1. 発表者名 瀬谷創・泊将史・力石真
2. 発表標題 大規模小売店舗の出店が既存店舗の 撤退・売上げに与える影響
3. 学会等名 応用地域学会2017年度大会
4. 発表年 2017年

〔図書〕 計1件

1. 著者名 Yoshiki Yamagata and Hajime Seya	4. 発行年 2019年
2. 出版社 Academic Pres	5. 総ページ数 302
3. 書名 Spatial Analysis Using Big Data: Methods and Urban Applications	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----