

令和 4 年 5 月 31 日現在

機関番号：12501

研究種目：若手研究

研究期間：2019～2021

課題番号：19K13667

研究課題名（和文）小地域推定における変数選択と不確実性の評価

研究課題名（英文）Variable selection problem and evaluation of measuring uncertainty in small area estimation

研究代表者

川久保 友超（Kawakubo, Yuki）

千葉大学・大学院社会科学研究院・准教授

研究者番号：80771881

交付決定額（研究期間全体）：（直接経費） 4,700,000円

研究成果の概要（和文）：本研究課題においては、小地域推定で用いられる混合効果モデルの変数選択問題に取り組んだ。小地域推定とは、標本調査において、サンプルサイズが小さくなってしまいう小地域レベルの推定を、混合効果モデルと呼ばれる統計モデルを用いて推定精度を高めようとする統計手法である。混合効果モデルに取り込む補助変数の組み合わせを選択する変数選択規準の開発に取り組んだ。

小地域推定における変数選択という大枠の中で、各論としていくつかの問題に取り組み、国際査読誌への掲載や学会での報告を行った。

研究成果の学術的意義や社会的意義

いくつかの本研究成果の共通した着眼点は、変数選択法と、予測量の不確実性の評価との関連である。小地域推定においては、各小地域の推定対象の値を言い当てること（点予測）だけでなく、その不確実性を見積もることを重視している。不確実性の評価方法として、平均二乗予測誤差（MSPE）と呼ばれる指標が一般的であるが、既存手法のほとんどは、候補モデルが真であるという仮定のもとでMSPEを評価していた。しかし本研究においては、この仮定をおかずに、変数選択の不確実性を明示的に考慮したMSPEの評価を行った。従来手法はMSPEを過小評価している可能性が高いことから、本研究成果は学術的にも社会的にも意義が大きい。

研究成果の概要（英文）：In this research project, we addressed the problem of variable selection in mixed effects models used in small area estimation. Small area estimation is a statistical method that attempts to improve the accuracy of estimation at the small area level, where the sample size is small in a sample survey, by using a statistical model called mixed effects models. We worked on the development of variable selection criteria for selecting combinations of auxiliary variables to be included in the mixed effects models.

Within the general framework of variable selection in small area estimation, we addressed several problems in each issue, which were published in international peer-reviewed journals and reported at academic conferences.

研究分野：統計学

キーワード：小地域推定 変数選択 変量効果モデル

1. 研究開始当初の背景

小地域推定とはどのような研究分野であるかを簡単に説明する。ある標本調査にもとづいて各市区町村における何らかの推定対象(平均所得・貧困率・疾病リスクなど)に対して、サンプルサイズが小さかったりゼロであったりする市区町村(小地域)においては、標本平均のようにその地域の標本調査のデータのみにもとづいた推定量は、分散が大きくなり推定が安定しない。そこで、混合効果モデルと呼ばれる統計モデルを適切に仮定し、小地域の推定対象の予測量をモデルから導出すると、この予測量は安定する。その理由は、モデルから導出された予測量は、補助情報を与える変数(補助変数)や他の近い地域の情報(地理情報)を利用しており、標本調査では十分な情報を得ることができなかった地域において他の情報を補填した予測量となっているからである。

2. 研究の目的

混合効果モデルによる小地域推定を行う際、補助変数をモデルの説明変数として利用するが、その際どのような変数を用いるのかは非常に重要な問題である。統計学では、どのような変数を用いるかをデータから決定するという考え方が一般的であるが、このような問題を変数選択問題と呼ぶ。先に述べたように、小地域推定では混合効果モデルと呼ばれる統計モデルが利用されているが、混合効果モデルにおける変数選択問題は統計の一般理論の枠組みで直近10年近く非常に盛んに研究がなされてきた。混合効果モデルは小地域推定以外にも多くの応用分野で利用可能なモデルであるため、一般理論における研究結果は重要である。しかしながら、一般理論で得られた変数選択の手法を、直接的に小地域推定に應用すると問題が生じる場合が多くある。それにもかかわらず、小地域推定への応用という観点から変数選択問題を議論してきた研究は数少ない。そこで、小地域推定における変数選択にはいったいどのような問題があるのか、そしてそれをどのように解決するか、これらを解明することが研究の目的である。

3. 研究の方法

「小地域推定における変数選択」という大枠のテーマのなかで、(1)「OBP法にもとづいた変数選択」と(2)「MSPEによる地域別変数選択法」という2件の具体的な問題に取り組んだ。これらの問題に対して、新しい変数選択の方法論を開発し、その理論的性質を調べた。また、提案手法の数値的なパフォーマンスを、シミュレーションや実データへの応用を通して調べて行った。

4. 研究成果

(1) OBP法にもとづいた変数選択

小地域推定モデルの未知パラメータを推定する方法として、observed best prediction (OBP)法と呼ばれる手法が近年提案された。OBP法ではモーメント条件のみしか確率分布の仮定をおいていないことから、モデルが真の構造をとらえ誤っている状況下において、従来手法よりも予測リスクが小さくなると主張されている。そこで、OBPと同様のリスクの測り方で、未知パラメータの推定のみならず、変数選択を行う手法を考えた。また提案手法による予測量のリスクを、変数選択の不確実性も考慮に入れた平均二乗予測誤差(Mean squared prediction error, MSPE)で測る手法を提案した。本研究は、国際学会での発表や国際査読誌への掲載がなされた。

(2) MSPEによる地域別変数選択法

小地域推定においては、先述のMSPEを最小にするように、地域内平均などの小地域パラメータの予測量を構成している。一方で、変数選択のステップにおいては、AICやBICなどの情報量規準を用いており、規準によって選ばれた統計モデルを真のモデルと仮定した上で評価したMSPEを最小とする予測量を用いている。そこで、より直接的に、予測量のMSPEが最小となるモデルを選択する手法を考えた。MPSEは地域ごとに評価するものであるため、地域ごとに最適なモデルが変わっても良い。つまり、真のモデルを選ぶのではなく、ターゲットに応じて最適なモデルが変わり、それを選択するという考え方を提案する。具体的には、小地域推定において最もよく用いられるFay-Herriotモデルと呼ばれる線形混合モデルの一種において、候補モデルが真であるという仮定をおかずに、予測量のMSPEを評価する。そして、そのMSPEが最小となるようなモデルを地域ごとに選択する手続きを提案する。提案手法は、モデル選択に用いられるのみ

らず、小地域推定で重視される予測量の誤差評価にも用いることができる。しかも候補モデルが真であるという仮定をおかずに誤差評価しているため、より頑健である。本研究の成果を、学会で発表した。

(3) その他関連する研究成果

その他、本研究課題に関連する研究成果としては、「パラメータが空間的に変化する自然指数型分布族による小地域推定」、「グループデータからの地域別所得分布推定のための空間統計モデリング」、「時系列グループデータを用いたロレンツ曲線のベイズ推定」に関する研究を行い、それぞれ国際査読誌に掲載された。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4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 Kobayashi Genya, Yamauchi Yuta, Kakamu Kazuhiko, Kawakubo Yuki, Sugasawa Shonosuke	4. 巻 40
2. 論文標題 Bayesian Approach to Lorenz Curve Using Time Series Grouped Data	5. 発行年 2021年
3. 雑誌名 Journal of Business & Economic Statistics	6. 最初と最後の頁 897-912
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/07350015.2021.1883438	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 Sugasawa, S., Kawakubo, Y. and Datta, G.S.	4. 巻 173
2. 論文標題 Observed best selective prediction in small area estimation	5. 発行年 2019年
3. 雑誌名 Journal of Multivariate Analysis	6. 最初と最後の頁 383-392
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jmva.2019.04	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 Sugasawa, S., Kobayashi, G. and Kawakubo, Y.	4. 巻 145
2. 論文標題 Estimation and inference for area-wise spatial income distributions from grouped data	5. 発行年 2020年
3. 雑誌名 Computational Statistics & Data Analysis	6. 最初と最後の頁 Article 106904
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.csda.2019.106904	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 Sugasawa, S., Kawakubo, Y. and Ogasawara, K.	4. 巻 90
2. 論文標題 Small area estimation with spatially varying natural exponential families	5. 発行年 2020年
3. 雑誌名 Journal of Statistical Computation and Simulation	6. 最初と最後の頁 1039-1056
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/00949655.2020.1714048	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7件（うち招待講演 6件 / うち国際学会 5件）

1. 発表者名 川久保友超
2. 発表標題 線形混合モデルの変数選択問題
3. 学会等名 東京大学・応用統計ワークショップ（招待講演）
4. 発表年 2020年

1. 発表者名 Yuki Kawakubo
2. 発表標題 Conditional Akaike information under covariate shift with application to small area estimation
3. 学会等名 EcoSta 2019（招待講演）（国際学会）
4. 発表年 2019年

1. 発表者名 Yuki Kawakubo
2. 発表標題 Small area estimation for grouped data
3. 学会等名 Eastern Asia Chapter of ISBA（招待講演）（国際学会）
4. 発表年 2019年

1. 発表者名 Yuki Kawakubo
2. 発表標題 Observed best selective prediction in small area estimation
3. 学会等名 Conference on Current Trends in Surevey Statistics（招待講演）（国際学会）
4. 発表年 2019年

1. 発表者名 Yuki Kawakubo
2. 発表標題 Small area estimation of general finite-population parameters based on grouped data
3. 学会等名 Workshop on Bayesian Modelling of Income and Wealth (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Yuki Kawakubo
2. 発表標題 Variable selection problem for linear mixed model under covariate shift
3. 学会等名 INDSTATS 2019 (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 川久保友超
2. 発表標題 線形混合モデルにおける平均二乗予測誤差による変数選択
3. 学会等名 統計関連学会連合大会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Yuki Kawakubo's Website
<https://sites.google.com/site/ykawakubostat/>

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
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------