

令和 3 年 5 月 11 日現在

機関番号：12602

研究種目：研究活動スタート支援

研究期間：2019～2020

課題番号：19K23242

研究課題名（和文）高次元多変量線形混合効果モデルと小地域推定への応用

研究課題名（英文）High dimensional multivariate linear mixed model and application to small area estimation

研究代表者

伊藤 翼 (Ito, Tsubasa)

東京医科歯科大学・M&Dデータ科学センター・助教

研究者番号：90849001

交付決定額（研究期間全体）：（直接経費） 2,200,000円

研究成果の概要（和文）：観測次元の二乗と地域数の比が定数に収束するような設定の下で、多変量小地域推定問題を考察し、特に各地域の集計データのみが得られているFay-Herriotモデルを考えた。この設定ではモデル内のパラメータの次元も発散するが、標準的な仮定の下で推定量が真の値に収束していくレートを示した。小地域平均ベクトルはそのベイズ推定量にパラメータの推定量を代入した経験最良不偏予測量で予測し、予測リスクとして平均二乗誤差と信頼区間を近似誤差が地域数に関して2次のオーダーになるように求めた。平均二乗誤差と信頼区間ともに、観測次元が固定されているときの結果と比べて、追加的な項が現れることが確認できた。

研究成果の学術的意義や社会的意義

標本数が小さい地域の平均所得などを推定する際、小地域推定の手法による予測量はリスクを改善するため有用であり、相関をもった多次元データに対しては、多変量モデルから得られる予測量によりリスクを改善する。一方で高次元モデルについての研究はない。小地域推定の手法は漸近近似に基づくが、観測値が高次元の場合は近似誤差が大きくなるため、従来の方法では誤った結果を得る恐れがある。本研究では、より高次元のデータをモデルに組み込むことでさらに予測リスクを改善でき、高次元データに対しても頑健なリスク評価を行っているため誤った結果を得る恐れが少ないため、公共機関が小地域に対してより正しい政策決定をする際の助けとなる。

研究成果の概要（英文）：Multivariate small area estimation problem, especially the Fay-Herriot model in which we only obtain aggregated data for each area is considered in the setting where the ratio of the square of the dimension of the observed vectors and the number of areas converges to a constant. The convergence rates of the estimators of the parameters in the model are shown, noting that the dimensions of these parameters also increase in this setting. Small area mean vectors are predicted by the empirical best linear unbiased predictors, which can be obtained by substituting the estimators of the parameters into its bayes estimator. As its prediction risks, the mean squared error matrix and the confidence interval are constructed such that the approximation errors are of second order regarding the number of areas. It is shown that the additional terms appear in both compared with the results obtained when the dimension of the observed vectors is fixed.

研究分野：数理統計学

キーワード：小地域推定 Fay-Herriotモデル 高次元データ 平均二乗誤差 信頼区間 2次補正

1. 研究開始当初の背景

サンプルサイズが小さい地域に対して、平均所得などの何らかの平均値を推定する際、標本平均では不安定になる。そこで、地域効果を変量効果として組み込んだ混合効果モデルをあてはめて得た推定量を小地域の予測量として用いることで、予測リスクを改善することができる。従来の小地域推定の研究は観測値が1次元のモデルに対するものが多い一方で、貧困率と失業率などのように相関をもった多次元データが観測されたとき、それらすべてを同一のモデルに組み込んだ多変量モデルについての研究も進んでおり、多変量モデルから導かれる相関を考慮した予測量が、よりリスクを改善することが知られている。また近年ではデータが膨大になり、小地域推定の文脈でも、観測ベクトルの次元が大きい観測値が得られるケースが見受けられる。小地域推定の理論的成果は漸近近似に基づくものが主流であるが、観測値の次元が大きい場合は近似誤差が大きくなるため、従来の方法を高次元データに当てはめると誤った推定結果を得る恐れがある。小地域推定の手法から得られる予測量は、公共機関などがこれら小地域に対して所得配分などの政策決定をするうえで参考とされるものであるため、このような誤った推定結果は誤った公共政策の原因になる恐れがあり問題である。

2. 研究の目的

本研究は、観測次元がサンプルサイズに対して大きい多変量線形混合効果モデルについて、パラメータの推定方法や推定量の性質を考察する。また、小地域推定においては予測量のリスク評価がパラメータの推定と同等に重要視されるため、小地域予測量のリスクとして、観測次元がサンプルサイズに対して大きい設定に対して頑健な平均二乗誤差や信頼区間を構成する。

3. 研究の方法

観測値の次元の二乗と地域数の比が定数に収束するような設定の下で、多変量小地域推定問題を考察する。モデルとしては、各地域の集計データのみが得られているFay-Herriotモデルを考え、変量効果と誤差項にそれぞれ独立に多変量正規分布に従うと仮定した階層ベイズモデルを考える。パラメータとしては回帰係数と変量効果の共分散行列になるが、階層ベイズモデルから得られる観測値ベクトルの周辺尤度を用いて、それぞれ最小二乗推定量とモーメント推定量で推定し、各推定量の漸近的な性質も漸近展開にもとづいて求める。小地域平均ベクトルは、階層ベイズモデルから導かれるそのベイズ推定量にパラメータの推定量を代入した経験最良線形不偏予測量で予測し、そのリスクとして平均二乗誤差と信頼区間を漸近近似に基づいて評価する。なお、小地域推定問題で要求されるように、この近似は誤差が地域数に関して2次のオーダーになる(地域数で乗じても0に収束する)ように行う。最後に、数値実験や実データ解析により、高次元の予測量がより予測リスクを改善することや、もともと平均二乗誤差と信頼区間の精度を確認し、特に従来の観測値の次元を固定したときの結果と比べて真のリスクに近いことを確認し、本研究結果の有用性を示す。

4. 研究成果

(1) パラメータの推定量の性質について

観測値の次元の二乗と地域数の比が定数に収束するような設定では、回帰係数と変量効果の共分散行列の次元も地域数とともに発散するが、標準的な仮定の下で推定量が真の値に収束していくレートを示した。観測値の次元が固定されている場合は、回帰係数と変量効果の共分散行列の推定量は標準的な地域数の平方根の逆数のオーダーで収束するが、本研究の設定では、共分散行列の推定量は観測値の次元の平方根だけ収束レートが遅くなることが示された。一方で、回帰係数の推定量の収束レートは従来の結果と同じであることが示された。また、従来の設定においても、共分散行列のモーメント推定量は2次バイアスをもつことが知られているが、本研究の場合、観測値の次元だけバイアスがさらに大きくなることが示され、そのバイアスを差し引くことで、2次不偏な推定量を構成した。

(2) 予測量のリスクについて

予測量のリスクとして、平均二乗誤差を漸近近似により地域数に関して2次のオーダーで評価した。従来の正規分布を仮定した小地域推定問題においては、この平均二乗誤差は、真の小地域平均値とそのベイズ推定量の差の平均二乗誤差、ベイズ推定量と最良線形不偏予測量の差の平均二乗誤差、最良線形不偏予測量と経験最良線形不偏予測量差の平均二乗誤差の3つの項に書き表せることがよく知られているが、本研究の設定では、最後の項が、観測値の次元が大きくなる分だけ近似式に追加的な項が現れることが示された。

その他の予測リスクとして、小地域平均ベクトルとある定数ベクトルの積(例えば、平均所得と平均支出額の差など)に対しての信頼区間を地域数に関して2次のオーダーで評価した。なお、小地域平均ベクトルとある定数ベクトルの積は、地域数に依存しない定数のオーダーであることを考えている。従来の正規分布を仮定した小地域推定問題においては、小地域平均ベクトルと

ある定数ベクトルの積の信頼区間は、小地域平均ベクトルの予測量とその平均二乗誤差行列を用いて、統計学において標準的に構成される信頼区間を2次バイアス補正したものとして得られることが知られているが、本研究の設定においては、平均二乗誤差の評価と同様に、その2次バイアス補正項に追加的な項が現れることが示された。なお数値実験は現在も行っているところではあるが、従来の観測値の次元を固定したときに得られる平均二乗誤差は、本研究の設定では、近似誤差が許容できないほどの大きさであることは確認できている。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔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
--	---------------------------	-----------------------	----

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