

## 論文の内容の要旨

論文題目 Mixed Modeling Approach to Small Area Estimation  
(混合効果モデルによる小地域推定へのアプローチ)

氏 名 菅澤 翔之助

本論文は、小地域推定問題に対して混合効果モデル (mixed effects models) を用いた手法の開発に取り組んだものである。小地域推定における代表的なモデルとして Fay-Herriot (FH) モデル (Fay and Herriot, 1979) と nested error regression (NER) モデル (Battese et al, 1988) がある。両モデルは非常に広く用いられているが、大きな欠点としてどちらのモデルも比較的強い仮定が要求される点が挙げられる。実際のデータ解析では、このような仮定が満たされないケースは多く存在し、そのようなケースにおけるモデルベース推定量の精度は良くない。そこで本論文では、既存の混合効果モデルを優越する柔軟な推定手法を提案し、それに付随する理論解析を行った。

まず第1章では introduction として、小地域推定の背景と各章のモチベーションと概要を与えた。そして第2章から続く7つの章で研究成果がまとめられている。

第2章、第3章では FH モデル、NER モデルにおける欠点として、データに正規性を仮定している点を言及した。現実のデータ解析では、所得や消費などの正值データを扱うケースが多くあるが、このようなデータの分布は非対称であることが多く正規分布を用いたモデルでは適切でない。代替策として対数変換が用いられることが多いが、必ずしも対数変換が妥当であるとは限らない。そこで、Yang (2006) によって提案された dual power (DP) 変換を用いて FH モデルの拡張を行った。この変換は1つのパラメータを持ち特殊ケースとして対数変換を含む。その変換パラメータをデータから推定することで、データの情報から適切な変換を選択し柔軟に推定を行うことを可能にした。数値実験の結果から、対数変換を利用したモデルや変換を施さないモデルでは、変換が誤特定されたケースにおいて推定精度が悪くなることが観察された。一方、DP 変換を用いた提案手法では、全てのケースで既存手法の性能に対して同等か優越することが確認された。

第4章では、小地域推定量のリスク尺度について考察した。従来、平均2乗誤差 (MSE: Mean Squared Error) を用いるが、Booth and Hobert (1998) で示されたように、地域ごとにデータを条件付けた MSE である条件付き MSE (CMSE: Conditional MSE) を用いる方が好ましい。FH モデルや NER モデルのように正規性を仮定したモデルにおいて、CMSE と MSE は漸近的に等しくなるが、非正規のもとでは必ずしも等しくないことが知られている。そこで本研究では、Ghosh and Maiti (2004) で提案された、カウントデータや2値データに対するモデルを含む広いクラスにおいて CMSE の考察を行なった。その結果、CMSE は観測値の関数で近似することができ、正規分布以外のケースでは MSE と漸近的に異なることが示された。また、実際に CMSE を利用するため CMSE の漸近不偏推定量を与えた。

第5章では分散不均一なデータに対するモデルの開発を行なった。NER モデルにおいて観測値の分散はすべてのデータで等しいことを仮定しているが、Jiang and Nguyen (2012) では、その仮定は必ずしも成り立たないことを実データの例を通して示した。Jiang and Nguyen (2012) は分散不均一な NER モデルを提案しているが、そのモデルのパラメータ数は地域の数に応じて増大してしまうため安定な推定を行うことが困難である。そこで、不均一分散は共変量で説明されるケースが多くあることに注目し、不均一分散が共変量の関数で与えられる構造を持つ NER モデルを提案した。このモデルに対してモーメント法による一致推定量を提案し、小地域パラメータに対する経験最良線形不偏予測量 (EBLUP: Empirical Best Linear Unbiased Predictor) を導出した。また EBLUP のリスク評価のための MSE の漸近不偏推定量も導出した。数値実験および実データ解析によって、この提案手法と Jiang and Nguyen (2012) の手法を含む様々な既存手法との比較を行い、その結果、提案手法があらゆるシナリオで良いパフォーマンスを示すことが確認された。

第6章では平均分散同時縮小モデルの開発を行なった。FH モデルにおいて、サンプリング分散は既知と仮定してモデルを推定するが、実際に使用する場合は推定したサンプリング分散を用いる。しかし、サンプリング分散の推定値は標本数が小さいと不安定になり、それを既知として利用したモデルから得られる推定

値も不安定なものになってしまう。この問題に対し、Maiti et al. (2014) はサンプリング分散もモデルから安定的に推定をする同時縮小モデルを提案した。Maiti et al. (2014) では周辺尤度最大化によるパラメータ推定の方法を提案しているが、モデルの複雑さゆえにパラメータ推定値が安定しない。そこで同時縮小モデルの階層表現を利用して、マルコフ連鎖モンテカルロ法 (MCMC) による推定方法を提案した。導出された MCMC は簡単な分布からの乱数発生の一連の繰り返しのみで実行できることがわかり、数値実験においても Maiti et al (2014) の手法を優越することが確認された。

第 7 章、第 8 章では変量効果の不確実性を扱った。地域によっては変量効果が不必要なケースがあり、Datta and Mandal (2015) では、変量効果がある確率でモデルに含まれないことを許すモデルを FH モデルの枠組みで提案し、FH モデルに対して良い性能をもつことを示した。第 7 章では、このアイデアを NER モデルに適用し、MCMC による推定手法を提案した。さらに、有限母集団に対する推定問題への適用も行い、有限母集団内のサンプル数が多くなるほど提案手法が既存手法よりも優越する結果を数値実験から得ることができた。第 8 章では、第 4 章でも扱った Ghosh and Maiti (2004) によるモデルの枠組みで Datta and Mandal (2015) のアイデアを適用することで、変量効果の不確実性の議論をカウントデータや 2 値データを含む枠組みまで広げることが考えた。そのために、小地域パラメータの事前分布として、Ghosh and Maiti (2004) で利用された共役事前分布とある 1 点分布の混合分布を用いることを提案した。この設定のもと、Expectation-Maximization (EM) アルゴリズムによるパラメータ推定法を提案し、小地域平均の推定量を導出した。そして、数値実験を通して提案手法が既存手法よりも良い性能をもつことが確認された。

## 参考文献

- Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28-36.
- Booth, J. S. and Hobert, P. (1998). Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association*, **93**, 262-272.
- Datta, G. S. and Mandal, A. (2015). Small area estimation with uncertain random effects. *Journal of the American Statistical Association*, **110**, 1735-1744.
- Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269-277.
- Ghosh, M. and Maiti, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika*, **91**, 95-112.
- Jiang, J. and Nguyen, T. (2012). Small area estimation via heteroscedastic nested-error regression. *Canadian Journal of Statistics*, **40**, 588-603.
- Maiti, T., Ren, H. and Sinha, A. (2014). Prediction error of small area predictors shrinking both means and variances. *Scandinavian Journal of Statistics*, **41**, 775-790.
- Yang, Z. L. (2006). A modified family of power transformations. *Economics Letters*, **92**, 14-19.