分散関数を用いた Heteroscedastic Nested Error Regression モデル

東京大・経済・院　菅澤 翔之助
東京大・経済　久保川 達也

1　はじめに

官庁統計等で十分な情報 (サンプルサイズ) が得られない地域の平均を推定する際, 単純な算術平均では分散が大きくなってしまう問題がある. その問題に対し, 線形混合モデルの一種である Nested Error Regression (NER) モデル を当てはめ, 経験最良線形不偏予測量 (EBLUP) によって平均を推定することで推定精度を高める方法が実用的によく用いられている．NER モデルは以下のように定義されるモデルである.

$$y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, m.$$

ここで $v_i \sim (0, \tau^2), \varepsilon_{ij} \sim (0, \sigma^2)$ である. 確率変数 X に対して $X \sim (a, b)$ は $\mathrm{E}(X) = a, \mathrm{Var}(X) = b$ を表すとする.

このモデルは $\mathrm{Var}(y_{ij}) = \tau^2 + \sigma^2$ のように分散均一な構造を仮定しているが,Jiang and Nguyen (2012) は現実のデータには分散不均一性を示唆するものがあり, 既存の NER モデルは不適切であることを主張した. そこで Jiang and Nguyen(2012) では誤差項に正規性を仮定して,$v_i \sim N(0, \lambda\sigma_i^2)$, $\varepsilon_{ij} \sim N(0, \sigma_i^2)$ なる分散不均一構造をもつ Heteroscedastic NER (HNER) モデルを提案した. しかし,σ_i^2 のパラメータ数は地域数 m のオーダーで増加していくので,$m \to \infty$ の漸近論を考えると n_i が有界のうちは一致性を持たないという問題がある.

2　分散関数を用いた HNER モデル

現実のデータでは分散不均一構造がある共変量に影響を受けているケースが多く見受けられる. そこで Sugasawa and Kubokawa (2015) では分散関数を用いて分散不均一構造を特定化する以下のような誤差項の構造を提案した.

$$v_i \sim (0, \tau^2), \quad \varepsilon_{ij} \sim (0, \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})).$$

ここで $\sigma^2(\cdot)$ は分散関数であり,z_{ij} は分散不均一構造に影響を与えるであろう共変量である. また Jiang and Nguyen(2012) のモデルとは異なり, 誤差項に正規性を仮定していない. 本報告では, このモデルのパラメータ推定,EBLUP, その MSE および MSE の漸近２次不偏推定量について報告し, シミュレーションや実データ解析を通してこのモデルの有用性を示す.

参考文献

[1] Jiang, J. and Nguyen, T. (2012). Small area estimation via heteroscedastic nested-error regression. *Canad. J. Statist.*, **40**, 588-603.

[2] Sugasawa, S. and Kubokawa, T. (2015). Heteroscedastic nested error regression models with variance functions. ArXiv:1505.07369.