エリアレベルモデルに基づく 都道府県別貧困率の推定とマッピング

九州大学　マス・フォア・インダストリ研究所
廣瀬雅代

統計数理研究所 医療健康データ科学研究センター 岡檀先生との共同研究

本研究は2019年度日本経済研究センター研究奨励金，科研費助成を受けたものである。
Special thanks to
桜美林大学 加藤直子先生
中央大 伊藤先生
統計センター, 岡山大学オンサイト施設, ROIS-DSオンサイト施設

Outline

1. 研究目的
2. 推定手法の検討
3. 都道府県別貧困率マップ

Income and Poverty counts

- SAIPE program in U.S.Census Bureau (http://www.census.gov/did/www/saipe/about/index.html)
  - the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program provides annual estimates of income and poverty statistics for all school districts, counties, and states.
  - The main objective of this program is to provide estimates of income and poverty for the administration of federal programs and the allocation of federal funds to local jurisdictions.
  - In addition to these federal programs, state and local programs use the income and poverty estimates for distributing funds and managing programs.

研究背景

- 近年, わが国でも貧困に関連した社会問題が提起
  - こどもの相対的貧困率が
    - 16.1%
      - 平成24年　国民生活基礎調査の概況
        https://www.mhlw.go.jp/seisakunitsuite/soshiki/toukei/dl/tp151218-01_1.pdf
    - 13.9%
      - 平成28年　国民生活基礎調査の概況,
        (http://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa16/index.html)
  - 2020年度前後にも, 新型コロナウィルス感染拡大の影響あり？
- 官庁・行政によって行われる貧困対策及び貧困が要因と成りうる社会的課題のための有用な対策立案の為に, 細かな区分ごとに貧困実態を把握するための**信用性の高いエビデンス資料**が非常に重要！

研究目的

【最初の試みとして、貧困率推定とそのマッピング】
- 単身世帯
- さらなる細かな層を考慮に入れてみる
  →都道府県 (47)×性別(2)×年齢階級(7)
- コロナ前により近い年のデータを用いてみる

研究背景---戸室(2016)

【懸念1】乗率の考慮？

- ●貧困率を都道府県別に推計
  - ●就業構造基本調査(1992, 1997, 2002, 2007, 2012年)
    https://www.nstac.go.jp/services/jisseki-o-shugyou.html
    - ●世帯所得の間隔に世帯数が均等分布していると仮定。
- ●都道府県別貧困率の推計
  - ●貧困率：総世帯のうち、**最低生活費**以下の収入しか得ていない世帯の割合
  - ●最低生活費:被保護者全国一斉調査と被保護者調査の「最低生活費」

出典：毎日新聞
https://mainichi.jp/articles/20160218/k00/00m/040/108000c

研究目的

【最初の試みとして、貧困率推定とそのマッピング】
- 単身世帯
- さらなる細かな層を考慮に入れてみる
  →都道府県 (47)×性別(2)×年齢階級(7)
- コロナ前により近い年のデータを用いてみる
  →就業構造基本調査 (H29), 同じ貧困線の定義, 仮定
- 【懸念1】乗率の考慮

Outline

1. 研究目的
2. 推定手法の検討
3. 都道府県別貧困率マップ

検討する推計手法

- 直接推定量:　例：男性＊年齢第１階級
  - 第i番目の都道府県、層 $h \in \{1, \ldots, H\}$ に対して、

$$\bar{y}_{ih} = \sum_{j \in s_h} \tilde{w}_{ij}\, y_{ij}$$

- y_{ij}: 第i番目の都道府県に所属し、その所属する都道府県の最低生活費を下回る(1)か上回るか(0)の値
- $\tilde{w}_{ij} = \frac{w_{ij}}{\sum_j^{n_{ih}} w_{ij}}$: weight（乗率考慮）
- n_{ih}: 第i番目の都道府県のh層のサンプルサイズ

Design一致性！

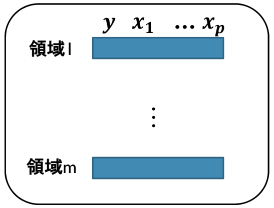
【懸念2】サンプルサイズが小さい地域も現れ、信頼性の低下が懸念

エリアレベルモデル

- 直接推定量と地域特有な補助情報との関係を表したモデル

---Fay Herriot model (1979)

$$g(\mathbf{y}) = \boldsymbol{\theta} + \mathbf{e}, \boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$



- 各地域の直接推定量: $\mathbf{y}_{m\times 1} = (\bar{y}_{1\cdot}, \cdots, \bar{y}_{m\cdot})'$
- 各地域の平均: $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)'$
- 補助変数: $\mathbf{X}_{m\times p} = (\mathbf{x}_{1\cdot}, \ldots, \mathbf{x}_{m\cdot})'$
- 回帰係数: $\boldsymbol{\beta} \in R^p$
  $rank(\mathbf{X}) = p < \infty, \sup_{i,k\geq 1} |\{\mathbf{X}\}_{ik}| < \infty$
- 地域の差異　$\mathbf{u} = (u_1, \ldots, u_m)', \mathbf{e} = (e_1, \ldots, e_m)'$
  それぞれ独立に
  $u_i \sim N(0, a),\ e_i \sim N(0, d_i),\ (d_1, \cdots, d_m : \text{既知})$

各地域の標本サイズが小さいときに比率データを用いて、各地域の割合を知りたい (SRSを仮定した場合)

- 比率データに対するArc-sin 変換によるモデル
  - 同様の変換モデル：Efron and Morris (1975), Casas-Cordero et al.(2015)

$$z = g(\mathbf{y}) = sin^{-1}(2\mathbf{y} - \mathbf{1}), \theta_i = sin^{-1}(2p_i - \mathbf{1})$$

$$d_i = 1/n_i$$

- 逆変換(NBT)：

$$\hat{p}_i^N(\hat{\theta}_i^{EB}) = \frac{1}{2}[1 + \sin(\hat{\theta}_i^{EB})],\ (i = 1, \ldots, m)$$

$$\hat{\theta}_i^{EB} = (1 - \hat{B}_i)z_i + \hat{B}_i x_i'\hat{\beta}$$

各地域の標本サイズが小さいときに比率データを用いて、各地域の割合を知りたい (SRSを仮定した場合)

- Empirical Bayes 推定量 (EB)
  - p_iに対するEBを陽に導出し、地域内サンプルサイズが小さいときのバイアスを地域数に関して漸近的に消去可 (Hirose, Ghosh, Ghosh., 2023)

$$\hat{p}_i^{EB} = \frac{1}{2}\left[1 + \exp(-g_{1i}(\hat{a})/2)\sin(\hat{\theta}_i^{EB})\right],$$

- $g_{1i} = g_{1i}(a) = ad_i/(a + d_i)$

EB推定量の予測誤差の評価

- Hirose, Ghosh, Ghosh (2023)

Def. MSE推定量：

$$\widehat{M}_i\ (\hat{\lambda}) = \begin{cases} M_i^0(\hat{\lambda}) & (0 < \widehat{M}_i^0) \\ M_{1i}(\hat{\lambda}) & (otherwise) \end{cases}$$

ここで $M_i^0(\hat{\lambda}) = M_{1i}(\hat{\lambda}) + M_{2i}(\hat{\lambda}) - b_M(\hat{\lambda})$

$$M_{1i}(\lambda) = \frac{1}{8}(1 - \exp(-g_{1i}(a)))(1 + \exp(-2a + g_{1i}(a))\cos(2x_i'\beta));$$

$$M_{2i}(\lambda) = \frac{1}{8}\exp(-g_{1i}(a))\left\{g_{2i}(a) + g_{3i}(a) + \frac{B_i^4}{4}V_a\right\}$$
$$+\frac{1}{8}\cos(2x_i'\beta)\exp(-2a + g_{1i}(a))\left\{g_{2i}(a) + g_{3i}(a) - \frac{B_i^2(B_i-2)^2}{4}V_a\right\}$$

$$b_M(\lambda) = -\frac{1}{8}\exp(-g_{1i}(a))\left(g_{3i}(a) - b_aB_i^2 + \frac{B_i^4}{2}V_a\right)$$
$$-\frac{1}{8}\exp(-2a + g_{1i}(a))\cos(2x_i'\beta)\left\{\frac{2g_{2i}(a)}{B_i^2} + g_{3i}(a) - b_a(B_i^2-2) - \frac{(B_i^2-2)^2}{2}V_a\right\}$$
$$-\frac{1}{8}\exp(-2a)\cos(2x_i'\beta)\left(2V_a - 2b_a - \frac{2g_{2i}(a)}{B_i^2}\right).$$

$$\lambda = (a, x_i'\beta),\ E[(\hat{a}-a)^2] = V_a + o(m^{-1}) \quad E(\hat{a}-a) = b_a + o(m^{-1})$$

EB推定量の予測誤差の評価：狭義正2次漸近不偏MSE推定量

定理 (Hirose, Ghosh, Ghosh; 2023)

いくつかの正則条件下、十分大なるmに対して以下が成立する

(i) $E[\hat{M}_i - M_i] = o(m^{-1}),$

(ii) $P(\hat{M}_i > 0) = 1,$

ここで $M_i = M_i(\lambda)$

➡ 狭義正2次漸近不偏MSE推定量を陽に導出

【懸念1】乗率の考慮

乗率の考慮

- 仮定モデル

$$g(\mathbf{y}) = \boldsymbol{\theta} + \mathbf{e}, \boldsymbol{\theta} = \mathbf{X\beta} + \mathbf{u}$$
$$u_i \sim N(0, a), e_i \sim N(0, d_i)$$

$$z_i = g(y_i) = sin^{-1}(2y_i - 1), \theta_i = sin^{-1}(2p_i - 1),$$
$$y_i = \bar{y}_{ih}, \qquad d_i = \sum_{j \in s_h} \widetilde{w}_{ij}^2$$

- 上記 y_i, d_i を変えてもHirose et al. (2023) の数理的結果が成り立つ
- EB推定量 $\hat{p}_i^{EB}$：$Design$一致性
  - 直接推定量 $\bar{y}_{ih} = \sum_{j \in s_h} \widetilde{w}_{ij}\, y_{ij}$ (for large n_{ih})

Outline

1. 研究目的
2. 推定手法の検討
3. 都道府県別貧困率マップ

モデルに基づくアプローチとして、都道府県別の関連のありそうな変数を検討
H22. 国勢調査：都道府県別高等教育機関卒率
H27.国勢調査：都道府県別単身世帯数

注意：この分析結果は、調査票情報を基に利用者が独自に作成・加工したものである。そのため、公表している統計とは異なる。

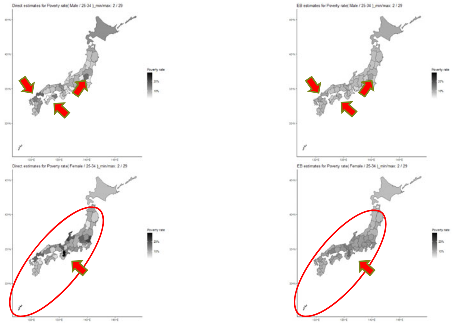
Age Group 2 [25~34 years old]

Direct

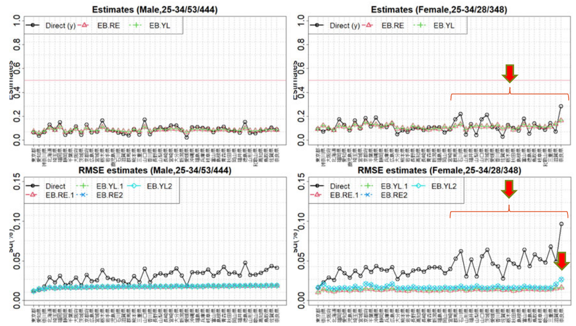
EB+RE

Male

Female



Age Group 2 [25~34 years old]



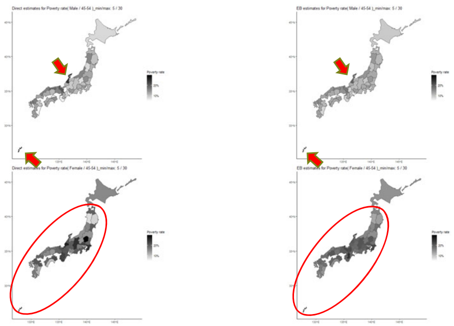
Age Group 4 [45~54 years old]

Direct

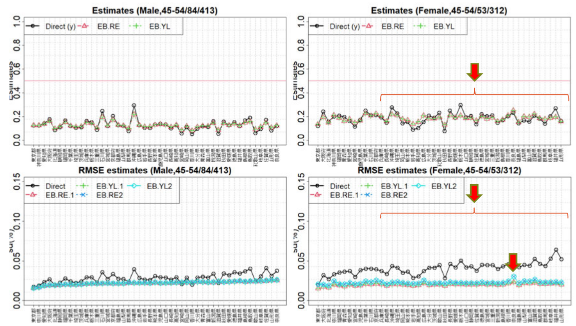
EB+RE

Male

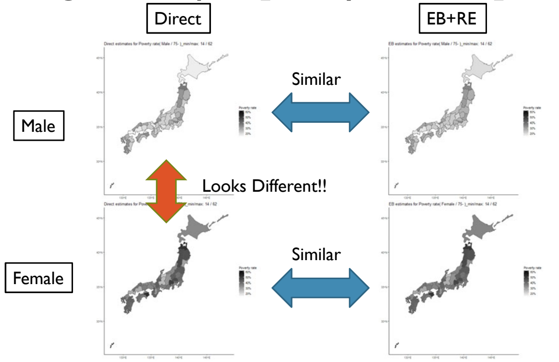
Female



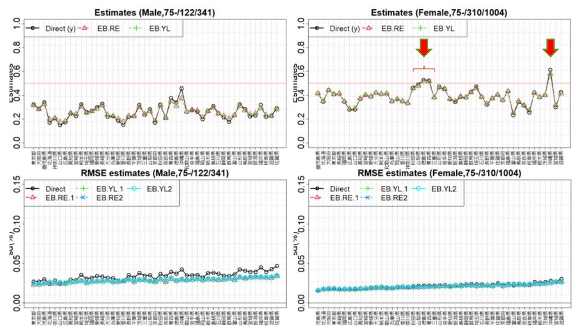
Age Group 4 [45~54 years old]



Age Group 7 [75~ years old]



Age Group 7 [75~ years old]



まとめ

- 都道府県別の貧困率推定のため、公的ミクロデータとエリアレベルモデルの適用を検討
- エリアレベルモデルでの推定法について簡単な改良を加える
- 都道府県別の貧困マップを作製

- 今後の課題
  - 別のデータの活用を検討
  - 乗率など、適用可能性について検討
  - EBPMに向けて役立つ資料にするには…

- 謝辞
- 本研究は2019年度日本経済研究センター研究奨励金助成を受けたものである。
  2019年度「貧困対策・格差是正に向けた実態把握資料のための貧困地図作成」
- Methodology部分は科研費18K12758 (乗率考慮なし), 22K01426(乗率考慮)の助成を受けたものである。
- オンサイト施設利用で大変お世話になりました
  - 統計センター、統計データ利活用センター、岡山大学オンサイト施設、ROIS-DSオンサイト施設

主な参考文献

- Casas-Cordero, C., Encina, J. and Lahiri, P. (2015), *Poverty Mapping for the Chilean Comunas, In Analysis of Poverty Data by Small Area Estimation*, (Edited by Monica Pratesi,). Wiley Series in Survey Methodology.
- Efron, B., & Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, *70*(350), 311-319.
- Fay III, R. E., & 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*(366a), 269-277.
- Hirose, M. Y., Ghosh, M., & Ghosh, T. (2023). Arc-sin transformation for binomial sample proportions in small area estimation. *Statistica Sinica*, 33, 705-727.
- Yoshimori, M., & Lahiri, P. (2014). A new adjusted maximum likelihood method for the Fay–Herriot small area model. *Journal of Multivariate Analysis*, *124*, 281-294.
- 戸室健作. (2013). 近年における都道府県別貧困率の推移について—ワーキングプアを中心に. 山形大学紀要 (社会科学), 43(2), 35-92.
- 戸室健作. (2016). 都道府県別の貧困率, ワーキングプア率, 子どもの貧困率, 捕捉率の検討. 山形大学人文学部研究年報, 13, 33-53.

ご静聴
ありがとうございました