

個票データを用いた自治体レベルの所得分布推定

一橋大学経済研究所「ミクロデータから見た我が国の社会・経済の実像」 2022年3月15日

岡部智人（青山学院大学）・野際大介（福島大学）・白川清美（立正大学）

所得格差の研究

- **T. Piketty 著『21世紀の資本』(2014)：所得格差についての歴史統計学的研究**
 - 税務統計を利用した確度の高い推計。
 - 特に所得分布の**1%シェア**（超富裕層）の広がりに注目。
 - 世界**30**か国についてデータベースを整備（**World Wealth and Income Database**）
- **Moriguchi & Saez (2008)、森口(2017)：日本の所得格差研究**
 - 日本においては富裕層の超富裕層化は観察されず、低所得層の貧困化が顕著になってきていると指摘。
- 所得分布情報は政府・自治体による福祉政策の立案に特に重要と思われる。

所得格差の研究（続）

- 自治体の所得分布はどうなっているのか？
 - 日本の各自治体別の課税対象所得総額は公表されている（『市町村税課税状況等の調』）
 - 家計あたり所得の「平均値」は推計可能。
その他の分布特性（中央値、分散、四分位数等）はほとんど不明
- 税務統計は秘匿性が高く、データ公開はあまり期待できない。
- 各自治体で分布を把握しているとしても、自治体間の比較は困難。
- 問い：自治体レベルの所得分布をなんとかして推定できないか？

自治体レベルの所得分布に関する公的統計

- 米国：
 - **American Community Survey (ACS) Public Use Microdata Samples (PUMS) by U.S. Census Bureau**
 - **1969年～**
 - **366 metropolitan areas** および **3,113 countries** をカバー
 - 自治体レベルの個人・家族所得（税引き前）の調査を目的とする唯一のデータソース
- 英国：
 - **Office for National Statistics (ONS)**
 - **1997年～**
 - **139 の metropolitan areas/ countries** をカバー
 - 家計の可処分所得を公表
- 豪州：
 - **Bureau of Statistics**
 - **2005年～（約10年おき）**

自治体レベルの所得分布に関する公的統計（続）

- 小規模自治体の推定を行うには、**tax return**（確定申告）のマイクロデータを利用できるのが一番よいが、これを公表しているのはアメリカだけで、他の**OECD**国でも**1970～1980年**のデータを公表しているのみ。
- **スペイン**：
 - 統計局で自治体レベルの所得分布推定値を公表
 - **Tax return**の標本を使い、地域毎の重みを改定して推計
 - **2015年～**
 - 約**9,000**自治体をカバー
 - https://www.ine.es/en/experimental/atlas/experimental_atlas_en.htm

日本で利用可能な公的統計

- (データ甲) 全国消費実態調査
 - 個票データ
 - 一部自治体のみしかカバーされていない(欠測・標本数不足)
- (データ乙) 自治体別各種統計(市町村のすがた)
 - 集積データ
 - 各自治体の特性を表す変数
 - 全自治体をカバー
- (データ丙) 国勢調査
 - 個票データ
 - 所得の情報はなし
 - 全自治体をカバー
 - 全消の翌年に調査される(1年の調査時期遅れあり)

全国消費実態調査

● 2004年（平成16年）全国消費実態調査

- 982市区町村（自治体区分は2019年時点、区は東京23区のみ）
- 約50%の自治体をカバー（ただし、標本数が不十分な自治体多数）

都道府県	自治体名	全世帯数	標本世帯数	所得の中央値 (百万円)
千葉県	浦安市	66,908	34	644
東京都	国立市	33,195	16	611
北海道	根室市	12,154	24	583
北海道	稚内市	17,783	21	566
福井県	高浜町	4,008	25	564
茨城県	下妻市	14,759	23	540
愛知県	高浜市	14,537	25	518
静岡県	掛川市	38,630	47	516
岐阜県	各務原市	49,033	47	510
宮城県	角田市	10,267	24	506

都道府県	自治体名	全世帯数	標本世帯数	所得の中央値 (百万円)
宮崎県	えびの市	9,119	25	119
沖縄県	沖縄市	44,500	47	143
鹿児島県	垂水市	7,659	23	159
沖縄県	宮古島市	20,527	47	160
沖縄県	東村	709	24	162
大分県	豊後高田市	9,678	24	169
高知県	仁淀川町	3,368	22	170
愛媛県	大洲市	18,967	5	171
北海道	夕張市	6,242	25	172
沖縄県	名護市	22,142	30	174

関連する先行研究

● 統計学：

- 小地域推定問題 (**Sugasawa & Kubokawa, 2020**)
- その他、空間統計的アプローチによる地域間所得間格差推定などもあり（ただし、関心は自治体内での所得分布に非ず）。

● 政治科学：

- 全国レベル世論調査を使った州レベルの事後層別化推定。
- 州別の投票結果を確定数（集積レベルの真値）として、世論調査の回答を州別に分解する。
- **Lax & Phillips (2009, AJPS)** 等

研究方法

- 階層ベイズモデルを用いたデータフュージョン的アプローチ
- 階層ベイズモデル：マルチレベルモデルとも呼ばれる
- データフュージョン：対象者の異なる複数のデータを統合して疑似的なシングルソースのデータを構成すること（星野 2009）
 - 欠測情報を相互に補完してデータを再構築する試み

	データ甲（全消）	データ乙（市町村別統計）	データ丙（国勢調査）
変数群A（等価年間所得等）	○有 （ただし一部欠測）	×欠測	×欠測
変数群B （自治体レベルの集積情報）	×欠測	○有	×欠測
共変量項目	○有（全対象に共通して 得られている変数）	×欠測	○有（全対象に共通して 得られている変数）

研究方法

1. フィッティング：データ **甲** と **乙** でモデルパラメーター $\{\beta, \sigma\}$ を推定する
 2. 予測：データ **丙** に当てはめて、自治体毎の所得分布を推定する
 3. 外的妥当性：データ **丁**（例えば、市町村税課税状況等の調）で分布の当てはまり具合を確認する
- ・ とりあえず、米国のデータ（**metropolitan area/ country** 別分布がわかる）を使って本推定方法の有効性を検討したい！

推定モデル

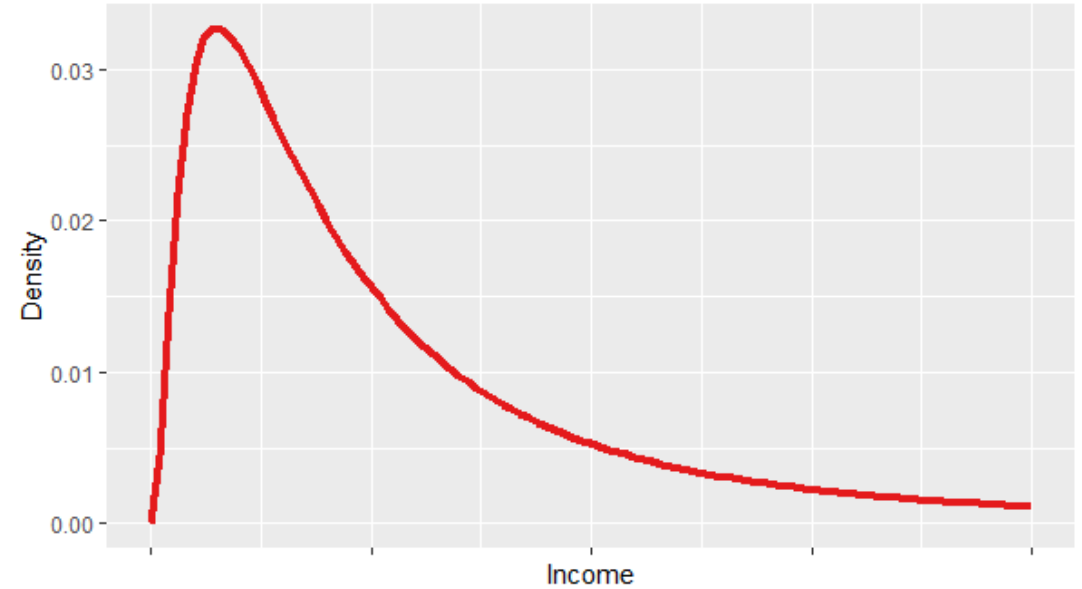
- **Unit notations**

i : household, j : city/municipality,

k : prefecture/state

$$y_{ijk} \sim \text{Lognormal}(\mu_{ijk}, \sigma_k)$$

$$\mu_{ijk} = \mathbf{X}_i\beta_1 + \mathbf{X}_j\beta_2 + \mathbf{X}_k\beta_3$$



- **X_i : household-specificな変数** (世帯主の性別・職業・年齢、配偶者の性別・職業・年齢、扶養者人数等)

- **X_j : city-specificな変数** (自治体の労働人口割合、産業構造、失業率等)

- **X_k : prefecture-specificな変数** (都道府県のダミー変数等)

今後の計画（検討中）

- 米国のデータを使って推定方法の有効性をシミュレーションしたい
- **Current Population Survey**：個人所得の標本（モデルフィッティングに利用）
- **Census PUMSU**：所得情報は使わず、世帯・個人の属性情報のみを使用（推計に使用）
- 自治体レベルでどの程度、所得分布が推定できるかを評価