



時系列データから所得の不平等の指数を推定する 新たな統計手法を考案

従来よりも安定的な推定に加え、将来予測も可能に

千葉大学社会科学研究院の小林弦矢准教授の研究グループ（千葉大学グローバルプロミネント研究基幹次世代研究インキュベータ「小地域」）は、ローレンツ曲線やジニ係数といった所得の不平等を表す指数を推定する新たな方法を考案しました。所得階級ごとの所得シェアについて蓄積されたグループデータを時系列データとして扱い、それを基に考案された状態空間モデル^{注1}を用いる本研究の方法では、従来の方法よりも不平等の指数を安定的に推定できる上に将来予測を行うことも可能になっています。

本研究成果は2021年2月2日に *Journal of Business & Economic Statistics* よりオンライン公開されました。

■ 研究の背景：ローレンツ曲線、ジニ係数とは

ローレンツ曲線は、ある母集団における所得の累積シェア率を縦軸に、人口の累積シェア率を横軸にとったグラフで、集団の所得の格差（or 不平等度）を視覚的にとらえるために用いられます。また、このローレンツ曲線から、不平等度をひとつの数値で表すジニ係数（0～1、0に近いほど格差が少ない）を算出することができ、この係数も、母集団内の所得の不平等を測るための非常に重要な経済指標の一つです。

図1左では3つのローレンツ曲線とジニ係数が示されています。ローレンツ曲線が45度線の際には全員の所得が同じである全く「平等」な場合を示し、「不平等1」と「不平等2」のローレンツ曲線は母集団において所得が不平等である場合を示しており、「不平等2」のほうが「不平等1」よりも格差の度合いが大きいことを示しています。

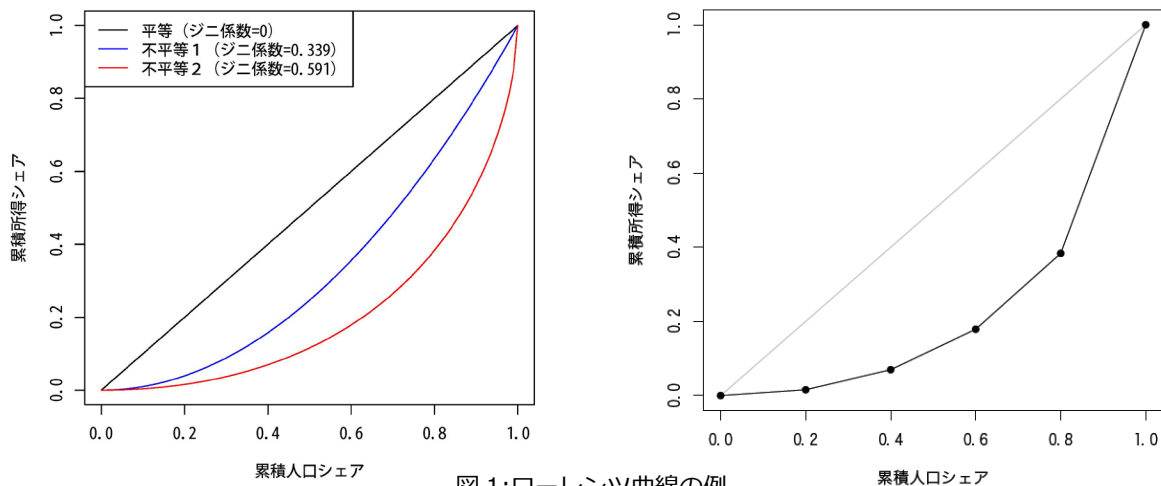


図1:ローレンツ曲線の例

このようなローレンツ曲線は、所得についての個票データが得られる場合には正確に推定することができますが、個人の特定を避けるなどといった観点から所得の個票データは利用することができません。代わりに、所得階級ごとの所得シェアに関する情報（グループデータ）をもとに、確率モデルをもとにした統計手法により推定されています（図1右）。

既存の研究では、例えば単一の地域や時点（国による毎年の調査など）ごとに独立に所得分布やローレンツ曲線を推定することを想定した統計手法のみが考えられてきましたが、推定結果がとても不安定で不確実性が大きいという問題がありました。それに対して本研究では、データの蓄積によってグループデータを時系列データとして扱うことで、ローレンツ曲線のモデルパラメータやジニ係数のような指標を安定的に推定できないかと考え、新しい統計モデルを考案し、さらに日本の家計調査データに当てはめて検証しました。

■ 研究成果

本研究で開発した方法（以下、提案手法）では、観測されている所得階級毎の所得のシェア率状況の時系列データをもとに状態空間モデル^{注1)}を考案し、ローレンツ曲線のモデルパラメータやジニ係数の時系列変化を推定します。従来の、時系列データを用いず各時点において独立にローレンツ曲線を推定する手法よりも、複数時点のデータの情報を利用することで関心のある数量を安定的に推定できる上に、モデルの時系列構造から将来予測を行うことも可能になりました。（詳説①）

この提案手法を日本の家計調査データ（5階級）^{注2)}に適用し、2000年1月から2018年12月のデータをモデル推定に使用しました。さらに2019年1月から2020年8月までのデータについて予測をし、モデル比較のための検証を行い、本データにおける予測パフォーマンスのよさから適した確率分布やローレンツ曲線の関数型モデルの比較を行いました。（詳説②）

パラメータの推定結果については、提案手法のほうが、不確実性を大幅に減らすことができることが確認できました。図2は同データを基に算出した、あるローレンツ曲線の関数型モデルにおける曲線の形状をコントロールする2つのパラメータとそれに付随するジニ係数の推定値（事後平均、黒線）と不確実性（95%信用区間、灰色のエリア）で、左側が従来の手法で推定した場合、右側が提案手法で推定した場合は、図からわかるように、提案手法での結果のほうが推定値の推移がよりなめらかで不確実性が低くなっています。

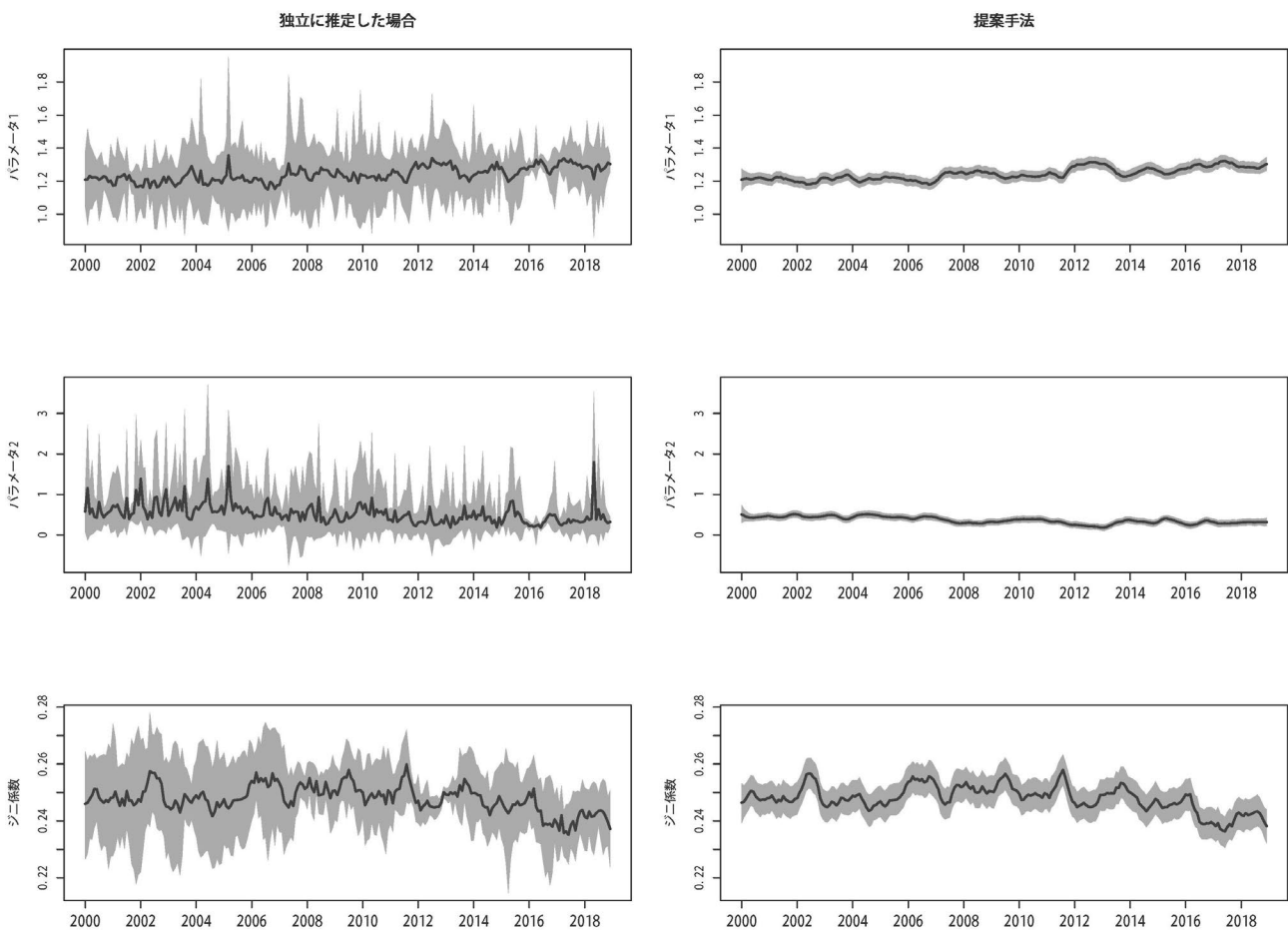


図2: 日本の家計調査データを基に算出したローレンツ曲線の2つのパラメータとジニ係数の推定値

図3はもっとも当てはまりのよかったローレンツ曲線の関数型モデルのもとでの、提案手法のもとでの各所得階級の所得シェアとジニ係数の推定結果です。各階級を見てみると、変動はあるもののおよそ2010

年以降、下位ふたつの階級の所得シェアが若干の増加傾向にあるのに対して、上位ふたつの階級のシェアが相対的に減少傾向にあると読み取れ、それに関連してジニ係数が少しずつ低下傾向となっています。これは2008年の金融危機あるいは2011年の東日本大震災以降の日本において、不平等の度合いが少しずつ小さくなる傾向にあるという結果を表しています。

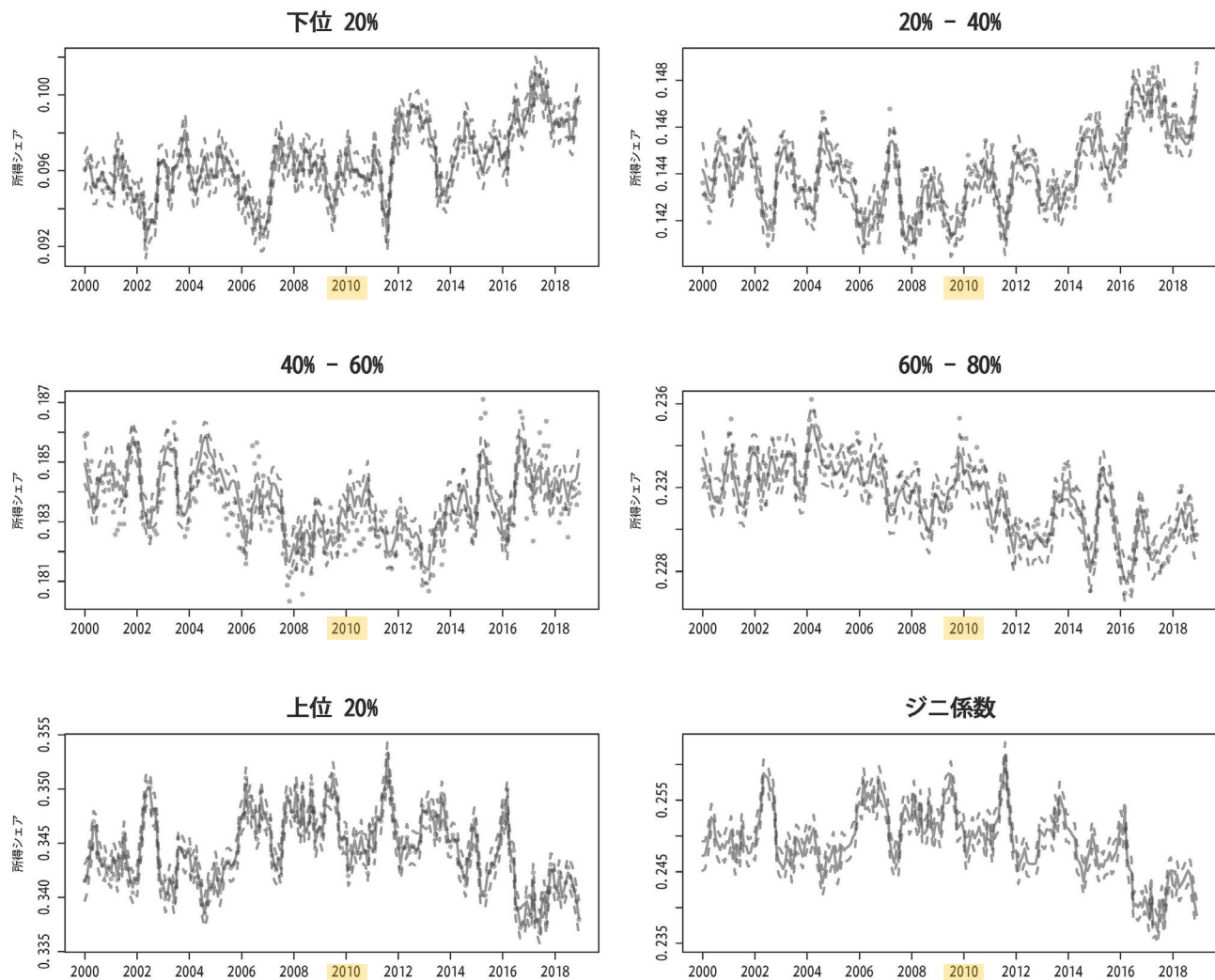


図 3: 提案手法のもとで得た、2000 年 1 月から 2018 年 12 月にかけての 5 つの所得階級の所得シェアとジニ係数の推定結果

■ 今後の展望

データの蓄積によって同じ調査についてより多くの情報が利用できるような状況においては、各時点のデータのみを用いて独立にモデルを推定するよりも、不確実性を減らしたり将来の予測を行ったりすることが可能になることが分かりました。今回はディリクレ分布などの確率分布を用いましたが、今後は分布の特定化によらない分析手法の開発に取り組みたいです。

■ 詳説

① 本研究で考案された統計手法の概要

グループデータとして所得階級 $k=1, \dots, K$ の所得シェア q_{tk} が時点 $t=1, \dots, T$ まで観測されるとします。本研究では所得シェアの期待値 (平均) が隣り合う所得階級の閾値でのローレンツ曲線の値の差に等しいと仮定します: $E[q_{tk}] = L(p_{tk}|\theta_t) - L(p_{t,k-1}|\theta_t)$ 。ここで $L(p_{tk}|\theta_t)$ は人口シェア p_{tk} におけるローレンツ曲線で θ_t はローレンツ曲線のパラメータ (観測されない, 時間によって変化する) を示します。ここでの特徴はローレンツ曲線のパラメータが每期変化することを想定しています。このような所

得シェアに関する仮定に対応する確率分布としてディリクレ分布^{注3)}を用い、時変パラメータを時系列構造に従うとする状態空間モデルを考案しました。時間方向にデータの情報を借り合うことでモデルパラメータの推定が安定化し、時系列構造により予測ができるようになりました。

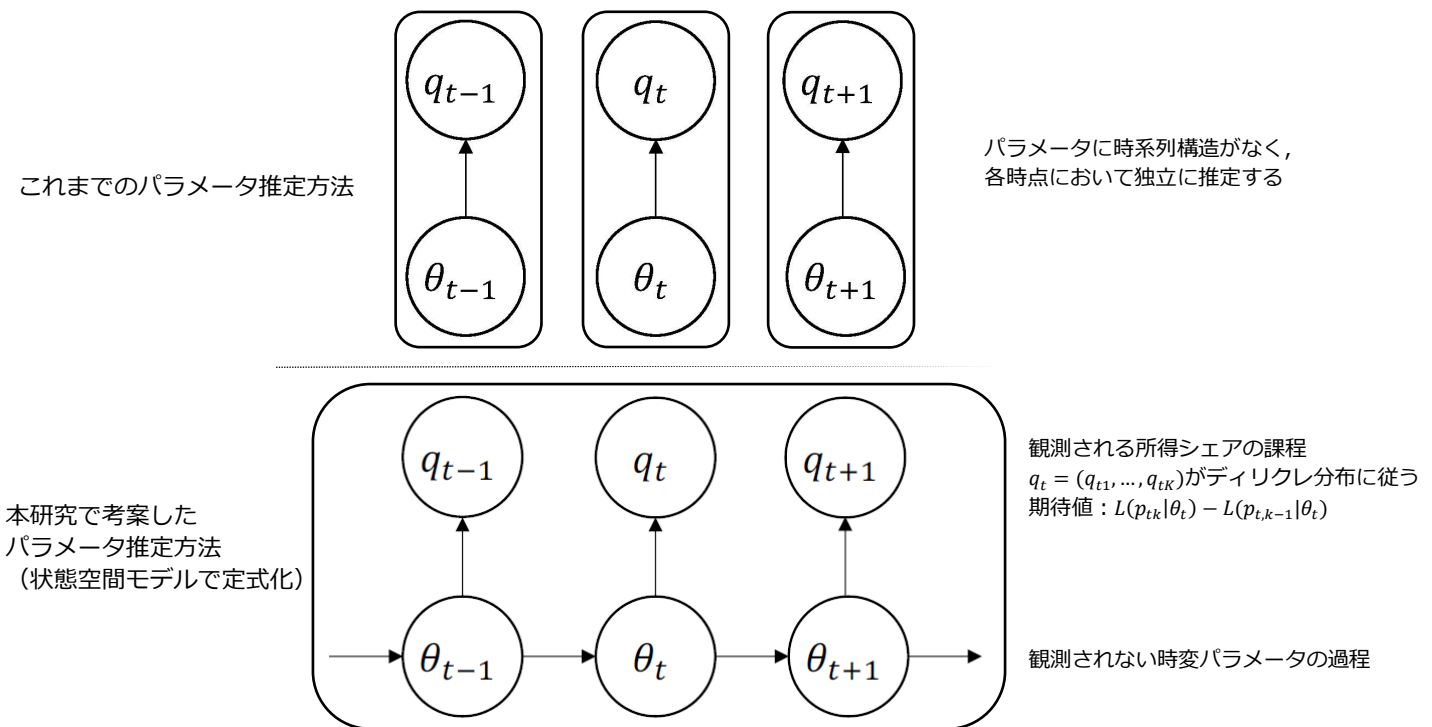


図4：本研究での状態空間モデルの概要。観測されないローレンツ曲線のパラメータが時系列構造に従っており、各時点におけるパラメータの値のもとで所得シェアがディリクレ分布に基づいて観測される。

②日本の家計調査データを用いての最適なモデルの検証

ディリクレ分布は要素間の共分散構造が制約的であるため、柔軟性を求めるために一般化ディリクレ分布のひとつ（以下 FD 分布）に基づいたローレンツ曲線推定のための状態空間モデルも考案しました。日本の家計調査データを用いた検証において、複数のローレンツ曲線のモデルをディリクレ分布および FD 分布の状態空間モデルに適用したところ、このデータにおいては予測パフォーマンスのよさからより柔軟な FD 分布と一般的で柔軟な形状を表現できるローレンツ曲線を用いたほうがよいということもわかりました。

■ 用語説明

注1) 状態空間モデル：時系列データを解析するためのモデルの枠組みのひとつ（図4参照）。

注2) 日本の家計調査データ：<https://www.stat.go.jp/data/kakei/index.html>

注3) ディリクレ分布：シェア率などといった、0 から 1 までの値をとり、合計値が 1 となるような確率ベクトルに対する確率分布のひとつ。

■ 論文情報

著者：Kobayashi, G., Yamauchi, Y., Kakamu, K., Kawakubo, Y. Sugasawa, S.(2021+).

論文タイトル：Bayesian approach to Lorenz curve using time series grouped data.

掲載誌：Journal of Business & Economic Statistics

DOI：<https://doi.org/10.1080/07350015.2021.1883438>

本件に関するお問い合わせ・取材のお申し込み

グローバルプロミネント研究基幹次世代研究インキュベータ「小地域」

TEL: 043-290-2438 メール：gp_sae@chiba-u.jp