

## 我が国における税務データに基づく分析の意義<sup>\*1</sup>

國枝 繁樹<sup>\*2</sup>

### 要 約

本稿においては、我が国における税務データの学術利用の意義につき、高額所得者の所得分布の分析の例を中心に説明する。欧米での税務データの学術利用の現状やそれを利用した様々な実証研究、我が国の過去の高額所得者の所得分布の研究を述べた後、今回の国税庁が開始した税務データを利用した共同研究につき概要を紹介し、申告所得税データを利用した研究の成果の一例として、高額所得者の所得分布のパレート係数の推計を説明する。我が国の2020年の高額所得者の所得分布のパレート係数の推計値は、過去の水準よりも大幅に低い1.45程度となっているが、これは、超高額所得者への所得集中が、資本所得を中心に進んだことを示している。この点は、今回、国税庁との共同研究で申告所得税の税務データの利用が可能になったことにより、明らかになったことである。我が国における税務データの学術利用には、学術面のみならず、政策面においても重要な意義があると考えられる。

キーワード：所得分布，所得税，税務データ

JEL Classification：H21, H24

## I. はじめに

コンピューターの処理速度の向上およびビッグデータの利用可能性の拡大は、多くの分野におけるビッグデータ利用を促している。経済学においても、集計データではなく、個別の主体

\* 1 本論文は、國枝・米田(2023)「日本の所得税制に関する税務データに基づく分析の意義」(税務大学校ディスカッションペーパー)の國枝執筆部分に基づいている。同ディスカッションペーパーおよび本論文に関し、日本経済学会2023年度秋季大会企画セッションの参加者、日本財政学会第80回大会企画セッションにおける討論者の石田良氏(財務省)および参加者から貴重なコメントをいただいた。また、本研究は、「税務大学校との共同研究における国税庁保有行政記録情報利用に係るガイドライン」に基づき国税庁に利用申出を行い、2022年3月に承諾を受け、税務大学校と共同で実施した研究の成果である。税務データの整備等に関し、国税庁・税務大学校のスタッフによる多大な協力をいただいた。ここに記して、感謝申し上げる。しかしながら、本論文の見解は執筆者個人の責任において発表するものであり、財務省及び財務総合政策研究所、国税庁及び税務大学校の公式見解を示すものではない。なお、國枝は、本研究につきJSPS科研費JP23K01426の助成を受けている。

\* 2 中央大学法学部教授、税務大学校客員教授

のデータであるマイクロデータに基づく実証分析が一般的となってきている。特に、最近では、従来のサーベイデータに加え、行政データ（administrative data）に基づく実証研究が重視されてきている。財政学においても、多くの国において、個別の納税者の匿名データ（氏名、詳細な住所等は除いたデータ）が学術目的で利用可能となっており、そうしたデータに基づく実証分析が必須になってきている。残念ながら、これまで、我が国においては、国税庁統計年報のような所得階層別の集計データや高額所得者・納税者公示制度が存在したものの、個票を含む税務データの利用がほとんど許されておらず、我が国の財政学者は、国際的に評価される租税政策に関する実証研究を行うことが困難となっていた。

しかし、国税庁は、2021年6月に、我が国の税・財政政策の改善・充実等に資する統計的研究を実施する研究者を公募することを公表した。この研究においては、税務大学校職員との共同研究を前提に、国税庁の保有する行政記録情報（税務データ）を利用した分析等を行うことができる。筆者（國枝）の属する研究グループは、幸い、「我が国の所得税制に関する税務データに基づく分析」の共同研究のパートナーに選ばれ、2022年3月より共同研究を進めているところである<sup>1)</sup>。同研究は、我が国での申告所得税の税務データの最初の学術利用となるため、まずデータクリーニング等のデータの整備を、時間をかけて行っているが、いくつかの研究結果も得られつつある。本稿においては、高額所得者の所得分布の分析を一例として、我が国において税務データが利用できない状況での過去の研究とその限界を紹介した後、国税庁の提供する税務データの利用の成果の一例として、最近の高額所得者の所得分布の分析について、その一端を紹介する。高額所得者の所得分

布はパレート分布で近似でき、パレート係数が低いほど、超高額所得者への所得集中が進んでいることを意味しているが、パレート係数の精度の高い推計には、高額所得者の個別の税務データが必要である。我が国の過去の推計では、国税庁統計年報の集計データや高額所得者・納税者公示制度のデータ等が用いられたが、高額納税者公示制度は2005年に廃止されるなど、推計するためデータの材料としては限界がある。今回、国税庁の提供する申告所得税の個票データを利用し、合計所得のみならず、資本所得および労働所得についてもパレート係数を推計した。2020年の合計所得のパレート係数は1.45程度、資本所得のパレート係数は1.35程度および労働所得の1.95程度となっている。合計所得のパレート係数は、高額所得者・高額納税者公示制度に基づく過去の推計値と比較して大幅に低下している。これは、高額所得者の所得の大半が、資本所得であるため、資本所得の超高額所得者への所得集中を反映したものである。国際的にも、我が国の資本所得のパレート係数は、リーマンショック前の米国の資本所得のパレート係数よりも低くなっている。このように、高額所得者のデータを十分含まないサーベイデータでの分析と異なり、税務データを利用することで、我が国においても、超高額所得者への所得集中が進んでいることを明らかにすることができる。高額所得者の所得分布については、さらに分析を進めることとしているが、税務データの学術利用の意義を示す一例と考えられる。

以下、第Ⅱ節で、経済学での税務データの利用が求められるようになった背景を説明し、第Ⅲ節で、税務データを用いた財政学における重要な実証研究を紹介する。第Ⅳ節において、米国、英国および北欧諸国といった税務データの学術利用の先進事例につき紹介する。第Ⅴ節で、

1) 筆者（國枝）の属する研究グループは、筆者のほか、宮崎毅九州大学教授、大野太郎信州大学教授、宮崎憲治法政大学教授、郡司大志大東文化大学教授、平賀一希名古屋市立大学准教授、および栗田広暁尾道市立大学講師である。なお、本共同研究では、2021年11月30日時点の国税庁保有行政記録情報（申告所得税）個別データを利用している。

我が国におけるこれまでの税務データの利用の実際とその限界を、高額所得者の所得分布の分析の例に焦点を当て、紹介する。第Ⅶ節では、今回の国税庁との共同研究の開始について、その概要と利用可能となった税務データのうち、申告所得税の税務データについて簡単に説明す

る。第Ⅶ節では、国税庁の提供する税務データの利用の成果の一例として、最近の高額所得者の所得分布の分析について、その一端を紹介する。最後に、申告所得税データを利用した実証研究の今後の展望について述べて、本稿の結びとする。

## Ⅱ. 行政データの経済学での利用の背景

最近の経済学においては、実証研究の占めるウエイトが重くなってきている。2018年最後の英エコノミスト誌は、最近の経済学の動向と今後注目される若手経済学者を紹介する特集記事を掲載したが、1988年の同様の特集記事で取り上げた最も優れた若手経済学者の多くが理論経済学者であったのに、30年後の今回の記事で取り上げた若手経済学者の多くが実証研究の分野で活躍していることを指摘している。こうした背景には、コンピューターの処理速度の向上およびビッグデータの利用可能性の拡大がある。

これまで、我が国を含む先進国の実証研究の分野で多く用いられてきたのは、サーベイ（アンケート）調査に基づくマイクロデータであった。米国においては、Current Population Survey (CPS)、Panel Study of Income Dynamics (PSID)、Survey of Income and Program Participation (SIPP)等のサーベイ調査が広く用いられてきたし、我が国でも、特に統計法改正後、総務省統計局の行う全国家計構造調査（旧全国消費実態調査）、厚生労働省の行う国民生活基礎調査等の政府の行うサーベイ調査の個票データを用いた実証研究が行われている。

これに対し、欧米における最近の実証研究においては、サーベイ調査ではない行政データを用いる実証研究がさかんになりつつある。サーベイデータと比較して、行政データには次のような利点がある（Card et al. (2010)）。第一に、サー

ベイデータの多くは、国民の一部のみを対象とした調査だが、多くの行政データは全人口を対象としており、標本数がきわめて大きい。第二に、行政データは、同一の個人または家計の長い期間にわたるデータを含んでおり、政策の長期にわたる効果等を調べるのに適している。第三に、一般に行政データは、無回答、対象人員の減少および過少報告等の問題を抱えるサーベイ調査で入手できるデータよりも、質の高いデータを提供している。我が国においても、アンケートに基づく調査は、共働き家計の増加その他の理由で回収率が低下するなど、様々な問題が生じてきている。

実証研究に用いられる行政データとしては、医療保険、失業保険等の社会保障関連のデータも多いが、税務データは、高額所得者も含めた対象範囲の広さや後述する経済学における租税回避行動への関心の高まりもあって、非常に有用である。ただし、税務データの利用には、納税者のプライバシー保護というセンシティブな側面もあり、プライバシー保護のための高い制約があるのも事実である。

### Ⅲ．現代租税理論の展開と税務データを利用した実証研究

#### Ⅲ－１．労働所得税

##### Ⅲ－１－１．課税所得の弾力性の重要性

かつては、労働所得税の効果としては、労働供給への効果が重視されており、経済学者は、労働供給の賃金弾力性を推計するために、賃金水準と労働時間についての個票に基づく実証研究を行い、家計の主たる男性労働者の労働時間は、賃金の変化にあまり反応しないことが明らかとなった。これに対し、Lindsey (1987) は、賃金税の引下げに対し、労働者は、労働時間ではなく、労働密度 (intensity) を増加させることで対応しており、その結果、課税所得が増加すると指摘していたが、Feldstein (1995) は、財務省が提供したランダム・サンプリングにより選ばれた3,954人の納税者の1985年と1988年のパネルデータを用い、差分の差分法により、課税所得の弾力性につき、約1.5という非常に大きい推計値を示した。これに対し、Goolsbee (2000) らは、課税所得が労働所得税率の変化に反応しても、その大半は、労働供給や労働強度の変化ではなく、租税回避行動によるものであることを指摘した。しかし、Feldstein (1999) は、課税所得の変化が租税回避によりもたらされるとしても、租税回避も歪みを引き起こすため、労働所得税の厚生効果を知るためには、課税所得の弾力性を推計することが不可欠であることを指摘した。

このように、課税所得の弾力性が、財政学における最重要のトピックの一つとなったが、納税者の課税所得が、労働所得税率の変化により、どのように反応したかを見るためには、税務データが不可欠である。税務データ以外のデータに基づき税法に沿って課税所得を算定する方法もありうるが、租税回避による課税所得の変化を計測することはできない。このため、税務データを用いた課税所得の弾力性に関する実証

研究が多く行われており、Saez, Slemrod and Giertz (2012) は、米国の課税所得の弾力性につき0.12～0.4の間との結論を得ている。また、他国においても、課税所得の弾力性に関する実証研究が進んでいる。例えば、Kleven and Schultz (2014) は、デンマークの全住民の税務データを含む行政データを用いて、推計を行っている。

なお、課税所得の弾力性の研究が進む一方、最近では、税務データと労働関係のデータを連結し、労働供給の弾力性につき分析する研究も現れてきている。

##### Ⅲ－１－２．高額所得者の所得分布

ごく少数の者への所得集中の形での経済格差拡大が問題になる中、トップ1%や0.1%という高所得者の所得分布が問題となっている。こうした高所得者のデータは、一般のサーベイでは欠落していることが多く、利用できない。税務データは、高所得者を含むデータが完備しており、利用価値が高い。

Feenberg and Poterba (1993) は、米国の高額所得者の分布を、パレート分布で近似し、パレート係数 (不平等なほど小さくなる) が1980年代以降低下しており、不平等度が大きくなってきたことを指摘した。Saez (2001) は、米国財務省のPublic Use Fileを利用し、米国の高額所得者の分布がパレート分布に従い、パレート係数が約2であることを示した。同ファイルは、全数でなく、サンプルデータだが、高所得者を多く含んでいるので、高所得者の分析には好都合である。(税務データを用いた高額所得者の所得分布の推計については、後半で詳述する。)

その後、高額所得者の所得分布については、他の国についても、税務データを用いた分析が

進んでいる (Atkinson, Piketty and Saez (2011))。Piketty (2014) は、これらのデータに基づき世界的ベストセラーとなった「21世紀の資本」を執筆し、また米国を中心に起こった Occupy the Wall Street 運動の「1%の高額所得者への過度の所得集中」との主張の根拠を提供することとなった。格差問題を正確なデータに基づき議論するに当たっても、税務データは不可欠になっている。

### Ⅲ-1-3. 最適労働所得税の導出

最適所得税理論は、Mirrlees (1971) によって初めて論じられ、その後、理論的な研究は行われてきたものの、効用関数の形状等に一定の仮定を置いた場合の最適所得税の税率構造をシミュレーションで導出することが主であり、現実の租税政策への影響は限定的であった。しかし、Diamond (1998) と Saez (2001) は、課税所得 (または労働供給) の弾力性と高額所得者の所得分布のパレート係数、さらには社会厚生関数の各人へのウエイトがわかれば、最適な

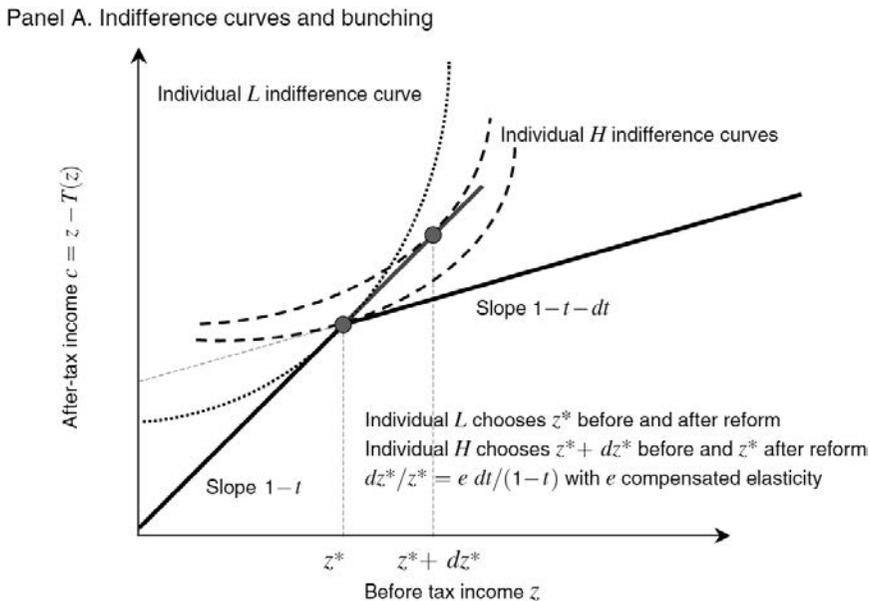
最高所得税率を導出することが可能であることを示した。上述のように、課税所得の弾力性および高額所得者の所得分布の推計には、税務データが必要であり、望ましい所得税の税率構造を推計するためにも、税務データが利用可能である必要がある。

### Ⅲ-1-4. バンチングに基づく弾力性の分析

所得税は、限界税率がブラケットごとに増加していくため、家計が直面する予算制約式は図1 (横軸が課税前の所得額、縦軸が課税後の所得額) の実線で描かれた予算制約式のように屈折する。屈折していなければ、図1中の個人Lと個人Hは、それぞれ異なる課税前の所得額となる点を選択していたはずだが、予算制約式が屈折しているため、両者とも屈折点を選択する。こうしたメカニズムのため、労働所得の分布は、理論的には、屈折点に対応する労働所得周辺に多くの家計が留まるはずである。これが、バンチングと呼ばれる現象である。

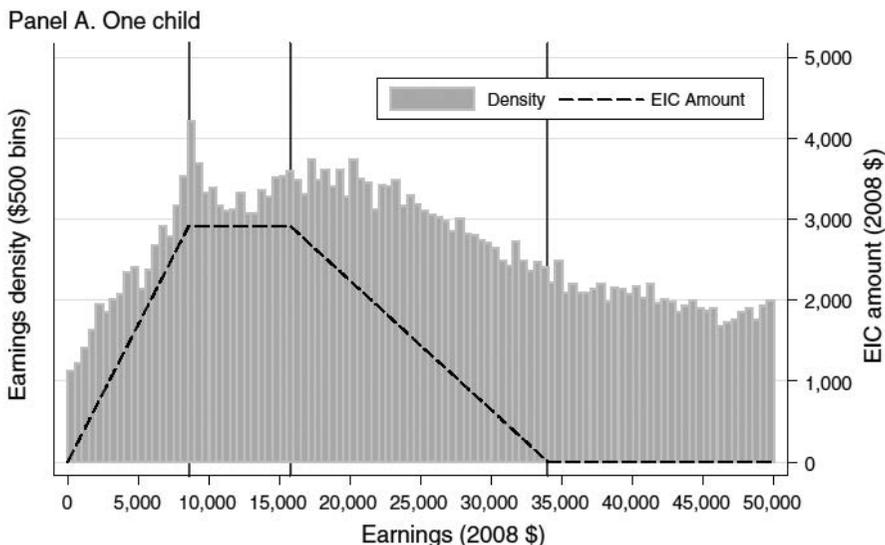
図2は、実際に米国の勤労所得税額控除に伴

図1 予算制約式の屈折点でのバンチングの発生



(出所) Saez (2010)

図2 米国の所得税におけるバンチング



(出所) Saez (2010)

う屈折点に納税者が集中している状況を示している (Saez (2010))。Saez (1999, 2010) は、このバンチングがどの程度、起きるかは課税所得の弾力性に比例することを指摘し、バンチングの状況を調べることで、課税所得の弾力性を推計できることを示した。

Chetty et al. (2011) は、デンマークの税務データと労働市場データを結合して、15~70歳の全人口の99.9%超をカバーしたデータを作成し、バンチングを利用して、労働供給の弾力性を推計した。

また、税制によっては、予算制約式がジャンプすることもありうる。これは、ノッチ (notch) と呼ばれ、ノッチの周辺にも家計が集中する。(これもバンチングの一種である。) Kleven and Waseem (2013) は、ノッチについても、バンチングの状況を調べることで、課税所得の弾力性を推計できることを示し、パキスタンの税務データを用い、課税所得の弾力性の推計を行った (図3)。

こうしたバンチングに基づく実証研究は、近年、注目を集めて、多くの最新の研究がなされているが、そのためには、個別の納税者の正確

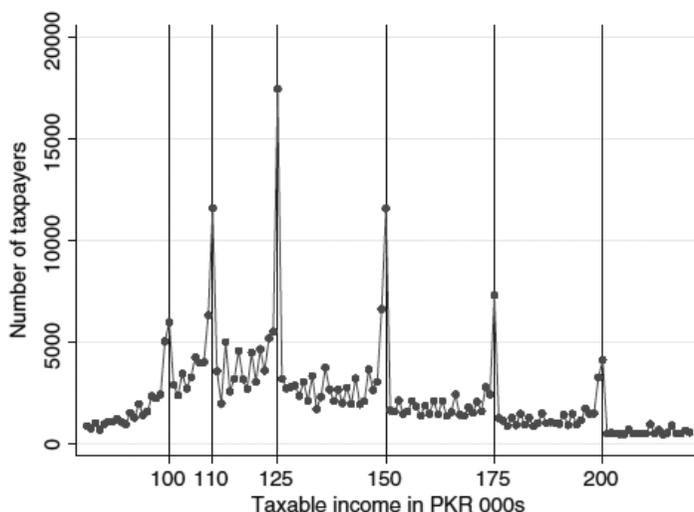
な分布状況のデータが不可欠である。

### Ⅲ-2. 資本所得税

資本所得税の消費・貯蓄の選択や資産選択の影響については、税引き後の収益率のわずかな差よりも、投資教育や行動経済学で重視されるデフォルト設定のあり方が重要であることが知られている。そうした研究は、米国の確定拠出年金を提供している個別企業の従業員のデータ等を用いて行われることが多いが、一国レベルで、金融所得税制の影響を分析した重要な文献として、Chetty et al. (2014) がある。同論文は、デンマークの全国民の税務データと労働市場データを結合して利用し、貯蓄優遇税制の効果につき分析した。この分析によれば、家計の中には、大多数のパッシブな貯蓄者と少数のアクティブな貯蓄者が存在する。パッシブな貯蓄者は、税引き後収益率の変化にはあまり反応しない。これに対し、アクティブな貯蓄者は、優遇税制に反応するが、既に貯蓄を有しており、これを優遇税制の対象に移転するだけなので、家計としての貯蓄は変わらない。従って、貯蓄優遇税制の影響が非常に限定的であることを示し

図3 パキスタンの所得税におけるバンチングの状況

## A First Six Notches



(出所) Kleven and Waseem (2013)

た。このように、異質の家計が存在する場合には、個別の家計のデータがなければ分析が難しく、金融所得税の効果を見るためにも、税務データが必要である。

最近では、一定の仮定の下、Saez and Stantcheva (2018) が、最適労働所得税と同様に、資本所得の分布のパレート係数を用いた最適資本所得税の公式を導出している。米国のリーマンショック直前の2007年の税務データに基づき、労働所得、資本所得および両者の合計である合計所得（図中では Total Income）のパレート係数を推計したのが、図4である。

Saez and Stantcheva (2018) の推計によると、労働所得のパレート係数は約1.6、資本所得のパレート係数は約1.38である。パレート係数は小さいほど、高額所得者に所得が集中し、格差が大きいことを意味する。従って、資本所得の方が、格差が大きい。資本所得と労働所得を合計した合計所得のパレート係数は、高額所得者の所得に占める資本所得の割合が高いため、資本所得のパレート係数に近い約1.4になる。このように、労働所得の格差以上に、資本

所得の格差が、経済格差の大きな要因となっており、Saez and Stantcheva (2018) の公式による最適資本所得税もより累進的なものとなる。

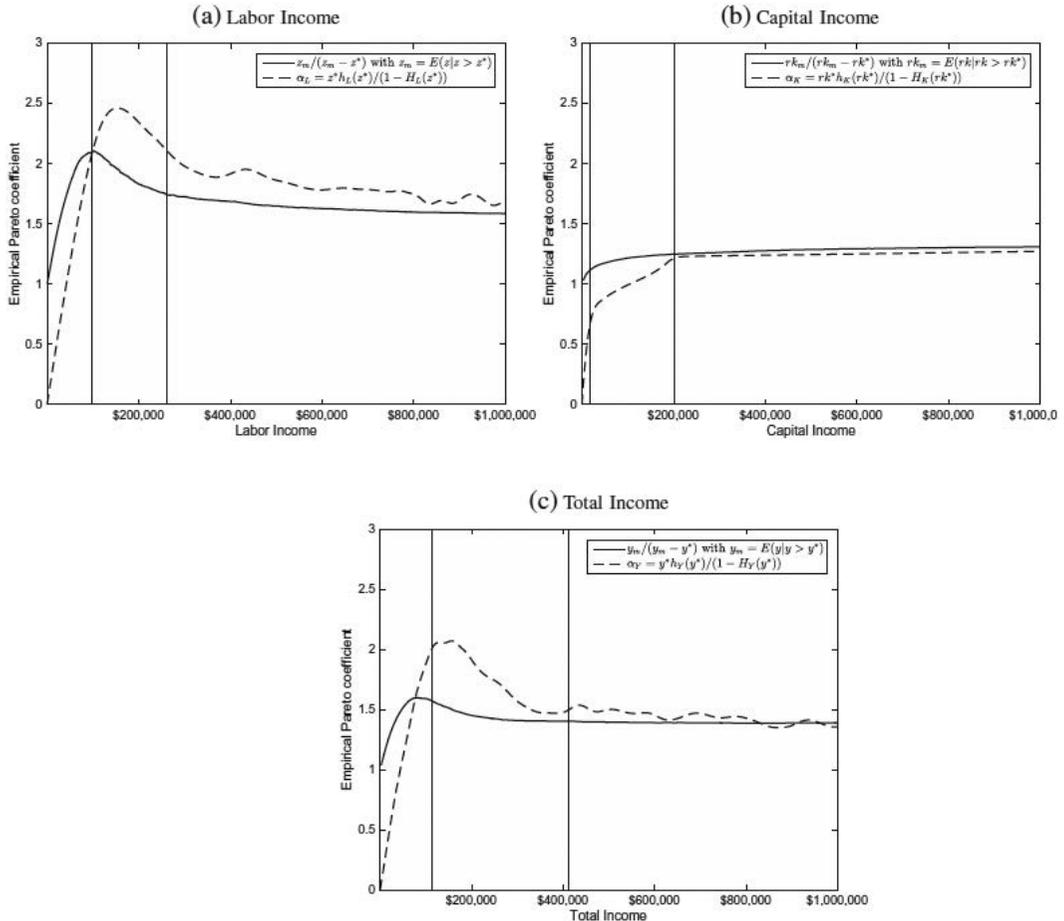
## Ⅲ-3. 資産課税

遺産税についても、税務データが分析に使われている。Kopczuk and Saez (2004) においては、資産分布の推計のために、内国歳入庁のSOIが整備した1916年以降（1982年からは各年）の遺産税の申告書のマイクロデータが、集計データと併せて利用されている。

また、遺産税の効果の分析にも、税務データが利用され、例えば、Gordon, Joulfaian and Poterba (2016)（著者のうち、David Joulfaian は、Office of Tax Analysis の有名なエコノミスト）は、内国歳入庁のSOIで用意した遺産税関連の申告書類に関するデータを用いて、2010年に遺産税の申告とキャピタルゲイン税の課税ベースのキャリーオーバーのどちらかを選択できる状態であった際に、納税者がどのような選択を行ったかを分析している。

また、北欧諸国やスイス等では、資産に直接、

図4 米国の労働所得、資本所得および合計所得の分布のパレート係数



(出所) Saez and Stantcheva (2018), Fig.4

課税を行う富裕税 (wealth tax) が、過去に課税されていた、あるいは現在も課税されており、富裕税の税務データを利用すれば、資産分布自体が直接把握できる。こうしたデータを利用した実証研究も最近、数多く行われている。(詳細は、國枝 (2023) を参照されたい。)

### III-4. 法人税

法人税の分野では、大企業であれば、一律の法人税率が適用される一方、財務諸表が公開されているため、法人税の納税申告を利用した税務マイクロデータを実証研究に用いる例は少なかった。国際比較を行う場合には、Amadeus、

Orbit や Compustat といった先進国の企業財務データベースが広く利用されている。

それでも、個人税分野でのマイクロデータ利用拡大の影響も受け、最近では、いくつかの国の研究者により、税務マイクロデータを利用した実証分析が開始されてきている。オックスフォード大学の Devereux 教授やその周辺の研究者のグループは、英国歳入関税庁 (HMRC) の Datalab の提供する匿名法人マイクロデータを用い、法人課税所得の弾力性 (Devereux, Liu and Lorenz (2014)), 法人税の負債政策への影響 (Devereux, Maffini and Xing (2018)), 移転価格を通じた租税回避行動 (Liu, Schmidt-

Eisenlohr and Guo (2017)) 等の研究を行っている。Devereux, Liu and Lorenz (2014) が、かつての英国の法人税率が利潤額 10,000 ポンドで変わっていたことを利用し、個人所得税におけるバンチングを利用した課税所得の弾力性の推計と同様の手法で、法人税の課税所得の弾力性の推計を行っている。また、Devereux, Maffini and Xing (2018) は、財務諸表に基づく分析と税務マイクロデータに基づく分析を比較し、財務諸表に基づく分析は租税特別措置等を十分反映していないため、観測誤差が大きく、税務マイクロデータを用いた分析の方が優れているとしている。

ドイツについても、Dwenger と Steiner がドイツ連邦統計局のリサーチ・データ・センターから

入手した匿名の法人税マイクロデータを用いて、課税所得の弾力性 (Dwenger and Steiner (2012)) や法人税の負債政策への影響 (Dwenger and Steiner (2014)) の実証研究が行われている。連邦統計局は、法人税のマイクロデータを 3 年ごとに公表している。ドイツの法人税の全対象企業のマイクロデータで、匿名化されているが、法人税計算のための重要な情報が記載されており、産業・地方・法的な形態等もわかるとされる (Dwenger and Steiner (2014))。

法人税の分野においても、こうした税務マイクロデータに基づく分析の優位性から、今後も税務マイクロデータに基づく実証分析が拡大していくものと考えられる。

## IV. 各国における税務データの学術利用

本節では、税務データの学術利用が進んでいる米国、英国および北欧諸国の税務データの学術利用の概要について説明する。各国とも、個人情報保護という制約の下、税務データの学術利用を実現するため、様々な方策を講じている。

### IV-1. 米国における税務データの利用<sup>2)</sup>

#### IV-1-1. Office of Tax Analysis による税務データの利用

米国において、個票の税務データを用いて分析を行っている主体としては、まず米国財務省の Office of Tax Analysis (OTA) がある。(以下は、OTA へのインタビュー等に基づく。) OTA は、財務省の一部局だが、経済学博士号を有する専門家を集め、租税政策の経済分析を行う機関である。OTA は、税務データに関する守秘義務を定める内国歳入法第 6103 条にお

いて、個票の税務データへのアクセスを認められている。

内国歳入庁 (IRS) には、個人・法人の確定申告情報等を含めたデータベースの Compliance Data Warehouse (CDW) が存在している。このデータベースより、個人税の部分については、所得階層別のサンプリングを行い、INSOLE (Individual and Sole proprietor) ファイルと呼ばれるファイルが作成される。サンプリングの方法については、内国歳入庁の統計部局である Statistics of Income セクション (SOI) が公表する。これらのファイルは、SOI の他、議会の Joint Committee on Taxation のスタッフおよび財務省の OTA のエコノミストのみがマイクロシミュレーションモデル作成やその他の租税政策関連の目的で、閲覧が許されている。CDW のデータを用いた研究の成果の公表には、ディレクター

2) 米国の税務データ利用の説明の一部は、財務総合政策研究所のスタッフとともに、米国財務省の Office of Tax Analysis 等を訪問した際のインタビューの内容を反映したものである。インタビューに応じてくれた Office of Tax Analysis のスタッフや同訪問に便宜を図ってくれた関係者に感謝したい。

の審査が必要である。原則として、個別のデータは、合計された形でしか公表されず、個票数が10未満の情報は公表できない。最近の学術誌では、発表論文のデータへのアクセスを提供し、他の研究者による再現可能性を確保することが求められるが、OTAの研究者のCDWを利用した研究には、そうしたルールは適用できない。それでも、OTAの外部的評価の高さゆえ、OTAの研究者の論文は、有力学術誌に掲載されている。

また、外部研究者がコンピューター・プログラムを内国歳入庁のSOIの研究者に提供し、そのプログラムをSOIの研究者がCDWのデータを利用して実行し、その結果のみを外部研究者に渡すという形での分析も実施例がある。Card et al. (2010)（同ペーパーの注2）によれば、上述のKopczuk and Saez (2004)の遺産課税の分析は、コンピューター・プログラムをSOIのBarry Johnson氏に渡し、実行してもらい、その結果を受け取るという形でなされている。

#### IV-1-2. Public Use File

INSOLE ファイルに加え、外部研究者が利用できるように、内国歳入庁のStatistics of Income (SOI) セクションよりPublic Use File (PUF) が提供されている。

PUFは、INSOLE ファイルから作成されるが、内国歳入庁の守秘義務を守るため、納税者の特定を避けるよう、サンプリングやデータの加工を行う<sup>3)</sup>。納税者の氏名・住所・社会保障番号のデータは除外される。サンプリングにおいては、サンプリングされなかった多くのデータは除外される。例えば、PUFには、INSOLE

ファイルの高額所得者のデータの10%のみが含まれる。また、極端な値を取るデータについては、特定化を避けるため、個別データからは除外される。データの一部は、統計学で「ぼかし(blurring)」と呼ばれる報告値を平均値に置き換える手法等で、修正される<sup>4)</sup>。PUFが、内国歳入庁の守秘義務に合致しているかは、SOIおよびSOIの統計コンサルティング先であるMathematica社により、毎年確認される。また、PUFの利用者や研究者達を加えたワーキンググループによりレビューを毎年行い、必要な修正を行っている。(例えば、2012年の修正については、Bryant et al. (2014)に詳細に説明されている。) PUFはSOIより有料で入手可能で、2024年11月の時点で2012～2015年のデータが提供されている。PUFは、財政学者による実証分析やTax Policy Center等の税制関連のシンクタンクの政策分析に幅広く用いられ、同データに基づく有力な論文が数多く書かれている。

#### IV-1-3. 合成データ (synthetic data) 導入の検討

PUFは貴重なデータを提供しているが、最近のネット上の検索機能の向上等により、納税者の特定化のリスクも増加している。SOIも、そうしたリスクに対応すべく、上記のとおり、PUFの修正を図っているが、より抜本的な改革として、synthetic data (合成データまたは模造データ)の導入の可能性についても検討が行われている。合成データとは、「機密保持へのアプローチとして実際のデータを公開する代わりに、一つ、あるいは複数の母集団モデルから発生さ

3) 各国における公的統計マイクロデータの匿名化措置については、伊藤(2018)を参照されたい。

4) ぼかし(blurring)とは、「ぼかしとは、報告値を平均値に置き換えることである。ぼかしを実施するための方法は数多くある。平均値を求めるためのレコードのグループを、他の複数の変数に基づくマッチング、あるいは当該変数でソートすることにより形成できる。(平均値を求める)1つのグループに含まれるレコード数は、固定的あるいはランダムのもいずれも可能である。グループの平均値は、グループ内の全メンバー、あるいは(移動平均のような場合には)“中位”メンバーに割り当てることができる。各変数について異なるグルーピングにより、複数の変数にぼかしを行なうことができる。」(統計データ機密保護に関する国連欧州経済委員会/EU統計局合同ワークショップ『統計データ開示抑制に関する用語集 改訂版、2005年8月』(翻訳:独立行政法人統計センター研究センター、2006年6月))とされる。

せた合成データを公開すること」(統計データ機密保護に関する国連欧州経済委員会・EU 統計局合同ワークショップ (2006)) とされる<sup>5)</sup>。2007年時点でのSOIワーキングペーパー (Vartivarian, Czajka and Weber (2007)) においては、税務データへの合成データの適用について、一般的な技術的課題に加え、①控除前の所得、控除額および控除後の課税所得のように、会計上、一定の関係があるデータが多いが、合成データ作成の過程でその関係が失われるおそれがあること、②税率が非線形であることに伴う問題、③データを用いた分析を行う上、重要な変数がどれになるのか必ずしも明確でないこと等の問題があることを指摘しつつも、従来のPUFが抱える特定化リスクの問題を考えると、合成データの検討を行う必要があるとしている。同様の問題意識は、PUFの利用者の間でも認識されており、Tax Policy Centerでも積極的な検討が行われている<sup>6)</sup>。また、SOIの委託研究で、Saez and Zucman (2018) は、所得分配の分析に利用可能な合成データの作成方法について提案し、実際に、INSOLEファイルに基づき、2013年から2016年の合成データを作成している。

2024年末時点での最新の情報については、Burman et al. (2024) が説明を行っている。同論文によれば、新たに税務データの学術利用に次の4つのレベルを設けることが検討されている。第1のレベルは、これまでも公開されてきた各種の集計表やレポートで、誰でもアクセスできる。第2のレベルが、個人所得税の合成データによるPUFで、従来のPUFに代わるものである。このデータには、SOIに要望すれば、誰でもアクセスが可能となる。その際の合成データの作成の方法については、Bowen et al. (2022) で紹介されているような手法が想定

されている。従来の手法に基づくPUFは、2015年分までで、2016年分以降は、合成データ形式のPUFに移行することが予定されている。第3のレベルとしては、validation server (検証サーバー) の設置が検討されている。これは、SOIにより審査済の研究者に、関係ないノイズを付加するなどの方法で納税者のアイデンティティーがわからない形の処理をした上で、自動システムを通じた税務データへのアクセスを認めるものである。第4のレベルとしては、validation serverで十分な分析が実施できない場合に、SOIにより審査済で、信頼された数人の研究者に限り、税務データへの直接アクセスを認めることが考えられている。第2および第3のレベルについては、未だ解決すべき課題も多いが、SOIおよびTax Policy Center等の民間シンクタンクは、精力的に検討を進めている。

## IV-2. 英国歳入関税庁 (HMRC) の Datalab

### IV-2-1. Datalab による税務データの提供

英国歳入関税庁のDatalabは、許可を受けた研究者に、秘密保持を保証した環境下で匿名化したHMRCのデータへのアクセスを認めている<sup>7)</sup>。(以下の説明は、HMRCのホームページでの説明による。) その目的は、HMRC、研究者のコミュニティおよび公衆の利益になる高品質の分析を生むことにある。

Datalabは、歳入関税庁の秘密保持政策に従っている。研究者が許可を得るためには、所属機関からの提案の後、短期のトレーニングコースを受ける必要がある。また、歳入関税コミッショナー法は、税務データへのアクセスの提供が、HMRCの機能に貢献することを求めており、研究提案はHMRCの機能に貢献する

5) 合成(模造)データについては、星野(2010)を参照されたい。

6) Tax Policy Centerでのインタビューによる。

7) 英国における行政データ一般の提供については、伊藤(2016)が紹介している。HMRCにおける税務データの取扱いについても、行政データ全般の取扱いを踏まえたものとなっていると考えられる。(ただし、同論文で言及しているAdministrative Data Research Network (ADRN)は、2018年7月で終了し、現在は、Administrative Data Research Partnershipに移行している。)

ものでなければならない。

許可を得ることができる研究者は、英国の学術機関または政府機関、さらに独立研究機関のようなNPOからの研究者である。2018年4月からは、営利目的の研究機関からの提案も受け付けるが、それは政府部門からの委託研究の場合に限られる。外国人研究者は、ケースバイケースで受け入れる可能性があるが、一般的には、上記の機関による契約・保証、雇用、または同機関での研究の必要がある。

HMRCは、研究提案がHMRCの機能への貢献、HMRCにとっての長期的利益、研究計画とその想定される影響、研究の実現可能性などの基準に基づき評価を行うが、現在の研究上の課題にとって重要なプロジェクトを支持する可能性が高い。

HMRCで行われる研究は、全て独立した研究であり、その結果は、公表されるか、あるいは政府部門で用いられるものでなければならぬ。

#### Ⅳ-2-2. 提供されるデータ

Datalabが提供するデータには、次のものが含まれる。

(各税の関連データ)

法人税：パネルデータ

(自営業者等の)自己申告

個人所得税のサーベイ：Public Use Tape  
(英国データサービスが保有)

キャピタルゲイン税：納税者のサンプルからの資産ごとのデータ

社会保険料（英国では歳入関税庁が徴収）：10%の納税者サンプルからのいくつかの項目についてのデータ

Enterprise Investment Scheme

研究開発税額控除

不動産印紙税

税額控除パネルデータ

付加価値税

輸出入（関税部門からのデータ）

(税務執行に関するデータ)

税務執行に対する中小企業の認識サーベイ  
税務執行に関する大企業の意見のサーベイ  
HMRCの利用者のサーベイ

※これらに含まれないデータについても、提案は受け取る。また、研究者が自ら保有するデータのマッチングについては、ケースバイケースで対応する。ただし、HMRCのデータの特定化が必要になるので、HMRCが自ら行う。

#### Ⅳ-2-3. 実際の利用機関・データ

実際に実施された（またはされている）研究の研究機関は、オックスフォード大学、LSEほかの英国の大学、財政研究所（Institute of Fiscal Studies）、中央政府の他機関、ウェールズ政府等である。研究内容は、法人税、個人所得税、不動産印紙税その他の多岐にわたっている。上述のオックスフォード大学のDevereux教授やその周辺の研究者のグループの法人税の実証研究も、Datalabの提供したデータを用いている。

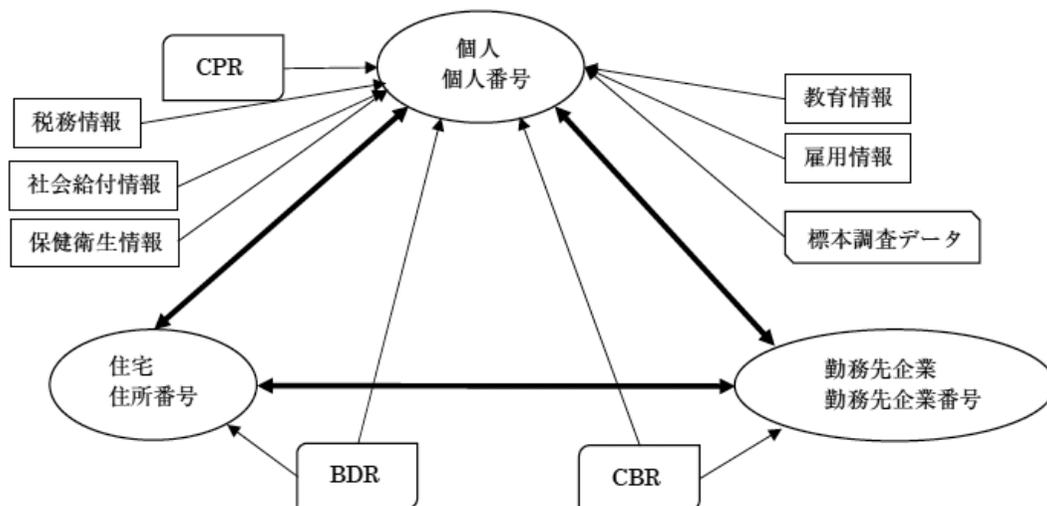
なお、Datalabの承認した研究プロジェクトのリストは、公開されている。

#### Ⅳ-3. 北欧諸国の税務データ利用

税務データのみならず、行政データが幅広く利用されている国の例としては、デンマーク等の北欧諸国があげられる。我が国においては、総務省統計局が国勢調査等の国勢の基本となる調査を、調査票を用いて行っているが、北欧諸国においては、行政記録データに基づく公的統計の作成が行われており、レジスター型センサスと呼ばれている。レジスター型センサスを最初に導入したのが、デンマークである。（以下、図5にあるデンマークレジスター制度の説明は、伊藤（2017）による。）

デンマークにおいても、調査票ベースのセンサスが行われていたが、1968年に中央人口レジスター（CPR）が設立され、1970年には税

図5 デンマークにおける統計情報システム



(出所) 伊藤 (2017)

務レジスターが整備されている。その他のレジスターも順次、整備が進められ、1981年よりレジスターベースのセンサス統計が作成されるようになった。個人番号制度が古くから整備されている北欧諸国では、各種のデータの連結も比較的容易である。上述した Chetty et al. (2014) や Kleven and Schultz (2014) の実証研究においては、所得税レジスター、人口レジスターおよび IDA (Danish Integrated Database for Labor Market Research) からのデータを結合して、サンプルが作成されている。

個票データ (非識別データ) は、個票データのアクセスが認証された大学等の機関において提供される。研究者は、利用申請を行い、誓約書署名の上、利用が可能となる。個票データの分析結果については、一部の結果をランダムに選んだ上で、職員が事後的にマニュアルでチェックを行っていると言われる。

現在の北欧諸国の行政データの学術利用は、世界で最も進んでいると考えられている。労働所得税と労働供給の関係等を研究する際には、労働関係のデータが必要なので、申告書等の記載事項に限られた税務データだけでは分析が難しいが、デンマークにおいては、Chetty et

al.(2014) や Kleven and Schultz (2014) の実証研究のように、税務データと労働関係のデータ (IDA) の結合により、そうした分析が容易に実行できる。

デンマークの充実した行政データの提供体制は、最先端の経済分析を可能とする。このため、ハーバード大学の Chetty 教授等のトップクラスの研究者が、現地の研究者と協力しつつ、デンマークのデータを使った最先端の研究を行っている。こうした状況は、我が国から見れば、税務データを含む行政データの利用がはるかに進んでいる米国の研究者にも懸念を抱かせており、Card et al. (2010) は、このままでは、最先端の研究が、米国から欧州に移ってしまい、現在の米国の経済学界におけるリーダーシップが失われ、また、米国の重要な政策課題についての分析が行われなくなってしまうと指摘している。

ただし、北欧諸国における税務データ利用については、税務データの守秘義務についての認識が、他の国とは異なっていることにも留意する必要がある。デンマークは別だが、北欧諸国のうち、スウェーデンやノルウェーのような国においては、伝統的に税務データが公開情報と

認識されてきた。例えば、Perez-Truglia（2020）によれば、ノルウェーでは、19世紀より税務データは公開情報であり、税務当局に他者の税務情報を照会することが可能であった。2001年には、新聞社が納税者の税務情報のウェブサイトを作成したことから、誰でも他者の税務情報に容易にアクセスできるようになった。その後、匿名でのアクセスができなくなるなどの制限が課されたが、税務データが完全に非公開になることはなかった。こうした北欧諸国の税務データの守秘義務の考え方に関連して、馬場（2016）は、スウェーデンの例につき、個人情報保護より、出版や表現の自由の確保や公共情報へのアクセス権を上位に位置付けていると指摘する。スウェーデンの国税庁（SKV）は、課税に関するほとんどの情報は「秘密にすべき」

情報であるが、所得課税に関する決定は公共情報であると強調している。このように、北欧諸国が、所得税データに関する守秘義務につき、我が国や英米とはかなり異なる認識でいることには留意する必要がある。

#### IV-4. 途上国での税務データの利用

最近では、先進国のみならず、途上国においても、税務データの利用が注目されている。Bachas and Jensen（2023）は、途上国における租税政策のための行政データの利用につき、セネガルやホンジュラスの事例も紹介しつつ、概観している。また、Brockmeyer（2019）は、税務データの利用についての実践的なガイドを提供している。

## V. 我が国でのこれまでの税務データを利用した研究とその限界： 高額所得者の所得分布の推計の例

過去においては、我が国の税務データの個票は、実証研究での利用はできなかった。税務データに関する統計としては、国税庁統計年報に掲載されている所得階級別の集計データがあり、過去においては、高額所得者公示制度・高額納税者公示制度で公示された税務データもあった。我が国の研究者は、これらのデータに基づき分析を行い、いくつかの成果を得てきたが、深刻な限界もあった。本節では、そのうち、高額所得者の所得分布の分析、特にパレート係数の推計に絞って、過去の研究とその限界について述べることにする。

国税庁統計年報のような所得階級別の集計データに基づき、パレート係数を推計する方法としては、いわゆるランクサイズ回帰（サンプルの大きい順に並べて順位（ランク）を付し、サンプルの規模の対数値を、順位の対数値に回帰させ、その係数を推計する手法）による推計

がある。ランクサイズ回帰においては、所得階級ごとの集計データしかない場合も、所得階級ごとの納税者数がわかれば、所得階級に対応する所得（対数）と順位に対応する最高所得階級からの累積納税者数（対数値）を利用することで、精度は劣るものの、パレート係数の推計が可能となる。

我が国においては、古くは、汐見他（1941（初版は1933））が戦前の税務データに基づくパレート係数の推計を示している。汐見他（1941）の方法を用いて、戦争直前までのパレート係数を推計したOno and Watanabe（1976）では、表1のように推計が示されている。

別の方法としては、Piketty（2003）やPiketty and Saez（2003）で用いられた方式（Blanchet, Piketty and Fournier（2022）では、“constant Pareto coefficient”と呼ばれている。）等の方法がある<sup>8)</sup>。この方法では、パレート分布の下限値

以上の標本の期待値を、下限値で除した値が一定となることに着目し、その値からパレート係数を推計する方法である。我が国の所得分布に関しては、Moriguchi and Saez (2008) が、国税庁統計年報より、我が国の高額所得者の所得が、我が国全体の所得に占めるシェアを推計しているが、その過程において、高額所得者の所得がパレート分布で近似できるとの仮定に基づき、Piketty and Saez (2003) の方法で、スーパーリッチ層の所得分布の推計を行っている。

しかし、国税庁統計年報の高額所得者の所得階級は十分細かく区分されておらず、税務データの個票を用いたパレート推計と比較すると、推計の精度は劣る。また、最適所得税の導出には、Gruber and Saez (2002) の分析のように、個票データを用いた課税所得の弾性値の推計値が必要だが、所得階層別の集計値しか存在しない状況では、課税所得の弾性値の推計も困難であった。

過去には、我が国で利用可能であった個票データとしては、高額所得者公示制度・高額納税者公示制度により公示されたデータが存在した。高額所得者公示制度は、1950年に導入され、高額所得者の氏名・住所・所得額が公示されていたが、1983年より納税額を公示する高額納税者公示制度に変更された。しかし、プライバ

シー保護の観点等から、2004年分の公示を最後に廃止されることとなった。高額納税者公示制度に変更される直前の1982年分では、所得1,000万円超の44万人超（申告納税者の6.7%に該当）の納税者が公示対象となった。税額1,000万円を対象とする高額納税者公示制度の下では、公示対象者数が大きく減ったが、バブル期には、17万人超の納税者が公示対象となっている。

高額所得者公示制度の個票データを用いてパレート係数を推計した分析としては、溝口(1987)がある。1962年から高額所得者公示制度の最後の年となる1982年までの20年間のパレート係数を推計しているが、その5年ごとの上位3,000人のパレート係数の推計は、表2のとおりである。

また、高額納税者公示制度になってからの分析としては、Fujiwara et al. (2003)がある(図6)。同論文では、高額納税者公示制度のデータを用い、そのうちのトップ1%と下位10%を除き、ランクサイズ回帰により、推計を行っている。

図6においては、パレート係数の推計値は平均すれば2を中心に変動しているが、1989～1991年のバブル発生時に大幅に低下し、バブル崩壊後に2を超える水準に戻っている。他方、同時期を含むパレート係数のNirei and Souma

表1 戦前のパレート係数の推計

年	1905	1910	1915	1920	1925	1930	1935	1940
パレート係数	1.87	1.90	1.81	1.70	1.72	1.66	1.65	1.57

(注) 1940年は、1938-39年の平均。

(出所) Ono and Watanabe (1976), Table 6

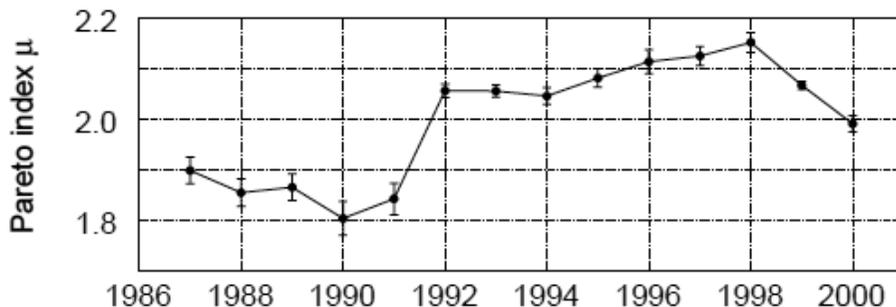
表2 1962～1982年の高額所得者公示に基づくパレート係数の推計

年	1962	1965	1970	1975	1980
パレート係数の推計値	2.224	2.236	2.197	2.176	2.528

(出所) 溝口(1987)

8) その他には、Blanchet, Piketty and Fournier (2022) が“mean-split histogram”と呼ぶ方法がある。Blanchet, Piketty and Fournier (2022) は、これらの既存の方法の問題点を指摘した上、一般化パレート分布に基づく推計方法を推奨している。

図6 1987～2000年の高額納税者公示に基づくパレート係数の推計



(出所) Fujiwara et al. (2003)

(2007)の推計値は、より大きな変動があったとし、バブル崩壊前後には、1.5台まで低下したと見ている(Nirei and Souma (2007)Figure. 5)。いずれにせよ、両推計におけるパレート係数の変動は、バブルにより所得格差が大きく拡大した後、バブル崩壊により所得格差が縮小したことに対応していると考えられる。

2000年代については、國枝(2012)が、高額納税者公示制度が終了する直前の2003年のデータに基づき、パレート係数を推計している<sup>9)</sup>。パレート分布の下限の設定を変えて、推計を行った結果が図7である。

パレート係数は、パレート分布の下限が達するまでは低下傾向にあるが、1億円を超えたあたりから約2.1で安定する。この水準は、溝口(1987)の1970年代後半のパレート係数の推計値よりも低く、当時より超高額所得者への所得集中が進んでいることを意味するが、米国のFeenberg and Poterba (1993)の推計値と比較すると、米国の1980年代前半の水準に留まっている<sup>10)</sup>。

國枝(2012)は、パレート係数の推計値に加え、労働の弾力性に関する先行研究の推計値を用い、我が国における最適最高税率を導出している。また、Miyazaki and Ishida (2022)は、1986年から1989年の間の高額所得者公示の

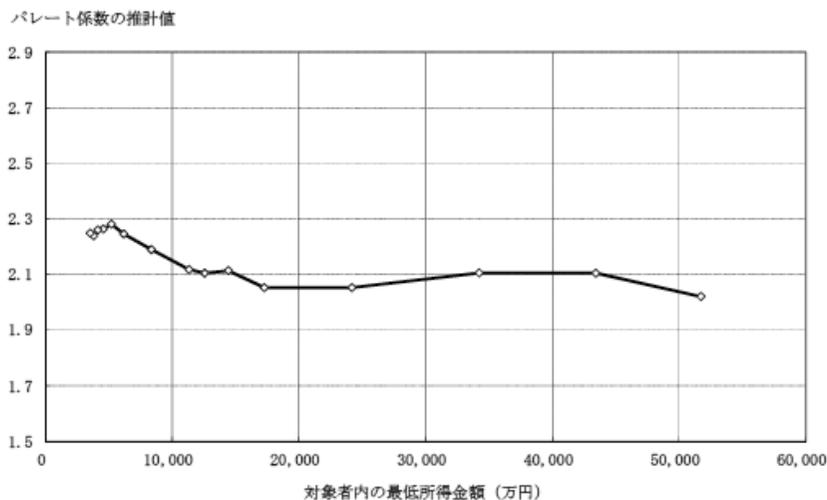
データを用いて、当時のパレート係数に加え、課税所得の弾力性を推計し、最適所得税の観点から最高税率引上げの余地があることを明らかにした。従来の我が国における課税所得の弾力性の推計は、サーベイデータである全国消費実態調査の個票データを用いた北村・宮崎(2010)等に限られたが、課税所得そのもののデータを用いていないため、限界があった。Miyazaki and Ishida (2022)の高額所得者番付のデータを用いた課税所得の弾力性の推計値は、0.055～0.074で、欧米諸国での推計値と比較して小さな値となっている。これは、我が国における最適な最高限界税率がかなり高い税率となることを示唆しており、政策的にも重要な意味を持つ。

高額所得税公示制度・高額納税者公示制度を利用した推計は、集計データと異なり、最上位の高額納税者の個別の税務データを利用できる点で、非常に有益であった。反面、実際の税務データの利用と比較すると、次のような制約もある。①氏名・所得・住所のデータしか公示されないため、所得の内訳やその他の要因がわからず、Saez and Stantcheva (2018)のように、資本所得・労働所得別の最適所得税を分析したり、所得の変動の要因分析を行ったりすることが難しい。②高額納税者公示制度に移行してからは、納税額しかわからないため、一定の仮定

9) 國枝(2012)は、ランクサイズ回帰ではなく、最尤法を前提に不変推定量(蓑谷(1998))を推計している。

10) Feenberg and Poterba (1993)の1980～84年のパレート係数の平均は、2.14である。

図7 我が国の2003年の所得税の分布のパレート係数の推計



(出所) 國枝 (2012)

を置いて、所得額を推計するしかないが、現実には所得の種類により納税額と所得額の関係は異なるため、不完全な推計とならざるをえない。

しかし、2005年以降、個人情報保護法制定の影響もあり、国税庁が高額納税者公示制度を終了してしまったため、高額納税者公示制度に基づく推計も困難になった。2000年代以降、

多くの先進国では、税務データの学術利用が一般的となり、そうしたデータを用いる分析が最先端の研究となる中、我が国の研究者は、そうした分析を行うことができず、我が国においても、税務データの学術利用が強く求められる状況となっていた<sup>11)</sup>。

## VI. 税務大専攻との共同研究の開始と提供される申告所得税データ

国税庁は、2021年6月に、我が国の税・財政政策の改善・充実等に資する統計的研究を実施する研究者を公募することを公表した。この研究においては、税務大専攻職員との共同研究を前提に、国税庁の保有する行政記録情報を利

用した分析等を行うことができる。税務データは、プライバシーの保護が強く求められるものであることから、税務データを利用する場合は、税務大専攻の任期付き職員に任用され、国家公務員法等の守秘義務が課される。共同研究の

11) なお、地方自治体の税務データについては、東京大学のEBPM推進のための自治体税務データ活用プロジェクトが、2021年より経済分析での活用を進めている。ただし、住民税データ中の所得等のデータの相当部分は、国税の所得税の確定申告等で申告された所得等のデータがそのまま、利用されているものであり、所得税データが、事実上、国税での守秘義務ルールと異なる方法で利用されている形になっている。国税当局との取扱いとの整合性を確保する観点から、地方自治体での税務データの取扱いのあり方につき検討することが望まれる。

テーマについては、これまでテーマ1「所得税及び復興特別所得税の確定申告書」第一表（AおよびB）及び第三表を用いた定量的な分析、テーマ2「法人税申告書 別表一（一）」（白色申告及び青色申告）を用いた定量的な分析の2つのテーマについて公募がなされた。筆者（國枝）の属する研究グループは、幸い、「我が国の所得税制に関する税務データに基づく分析」の共同研究のパートナーに選ばれ、2022年3月より共同研究を進めているところである。

提供される税務データのうち、申告所得税データについては、「所得税及び復興特別所得税の確定申告書」（以下、「確定申告書」）第一表および第三表に記載されている各項目のデータが利用可能である。（ただし、納税者氏名等は、プライバシー保護の観点からアクセス不可。）我が国の所得税制は、所得を性質に応じて分類し、それぞれの所得金額の計算方法につき定めているため、確定申告書第一表には、まず、事業所得（営業等所得・農業所得）、不動産所得、利子所得（源泉分離課税分を除く）、配当所得（源泉分離課税分を除く）、給与所得、雑所得（公的年金等の雑所得、業務に係る雑所得およびその他の雑所得）、総合課税の対象となる土地・建物等以外の短期・長期の譲渡所得および一時所得の収入金額等および所得金額等の各項目がある。

他方、申告所得の中には、分離課税されるものがあるが、確定申告書第三表において、そうした所得に関する項目が記載されている。具体的には、分離課税の対象となる土地・建物等の短期譲渡所得（一般分・軽減分）および長期譲渡所得（一般分・特定分・軽減分）、一般株式等の譲渡所得、上場株式等の譲渡所得、上場株式等の配当所得等所得および先物取引に係る雑所得等が含まれる。また、山林所得および退職所得も、総合所得と異なる税務上の取り扱いの対象とされ、分離して課税されるため、確定申告書第三表に記載がなされる。第三表に記載される各所得についても、確定申告書第三表では、収入金額と所得金額が記載されている。確定申

告書第一表に記載される総合所得の各所得と第三表に記載される申告分離課税の各所得の合計は、「合計所得金額」と呼ばれる。（ただし、総合課税分の長期の譲渡益および一時所得については、1/2の金額のみに課税がなされる。）その後、繰越損失控除を引いた金額は、総合課税分のみの場合は「総所得金額」、申告分離課税分も含めた場合は「総所得金額等」と呼ばれる。

税額計算のためには、さらに各種の所得控除を差し引く必要があるが、確定申告書第一表には、所得から差し引かれる金額として、各種の所得控除（社会保険料控除、小規模企業共済等掛金控除、生命保険料控除、地震保険料控除、寡婦・ひとり親控除、勤労学生・障害者控除、配偶者（特別）控除、扶養控除、基礎控除、雑損控除、医療費控除および寄附金控除）が記載されている。

次には、税金の計算の過程となり、申告分離課税分がない場合、所得控除を差し引いた金額（「課税総所得金額」）に、税率を適用して、税額を計算する。申告分離課税分がある場合は、総合所得分に加え、分離課税の課税分も第三表で計算した上で、第一表に所得額・税額を転記する。さらに、各種の税額控除が差し引かれる。具体的には、配当所得に対する法人段階・個人段階の二重課税を調整するための配当所得税額控除、住宅借入金等特別控除、政党等寄附金等特別控除、住宅耐震改修特別控除等の税額控除の額を差し引く。さらに災害減免法の適用がある場合は、災害減免額が記載される。加えて、災害減免額を差し引いた基準所得税額に、2.1%を乗じた額が、復興特別所得税額となる。最後に、外国政府に対し既に納付した税額分を外国税額控除等として差し引く。申告所得に係る源泉徴収税額がある場合は、その額を差し引いた額が申告納税額となる。確定申告書第一表には、さらに若干のその他の記載事項がある。

以上のように、今回の国税庁との共同研究で提供される税務データは、申告所得税の各所得につき、税額計算を行うための詳細な項目が含まれており、課税所得の弾力性等の推計が可能

になる。また、サンプル調査ではなく、ある一年間を通した税務データが利用できることから、一年に2千万人を超える膨大な数の納税者をカバーすることができる。その中には、サーベイデータではおそらくあまり含まれていない超高額所得者も含まれている。また、申告内容についても、サーベイデータと異なり、真実でない申告を行った場合には、加算税や悪質な場合には罰則が科されるため、一定の正確性が担保されており、信頼性の高いデータとなっている。上述のように、我が国においては、税務データの個票の学術利用が認められておらず、財政学における実証研究の深刻な制約となってきたが、今回の税務データの学術利用の開始により、多くの成果が期待される。

ただし、申告所得税の税務データには、いくつかの限界があることにも留意する必要がある。一つは、対象が確定申告書の提出を伴う申告所得税のデータに限られる点である。我が国の課税形態には、申告書の提出が必要になる申告総合課税や申告分離課税の他、源泉徴収により納税手続が完結する源泉分離課税が存在し、利子所得や配当所得等に広く適用されている。このため、利子課税のほとんどについては、申告不要であり、今回の税務データには含まれない。(確定申告書第一表に含まれる総合課税の対象の「利子所得」は、国外で支払われる預金等の利子など国内で源泉徴収されない所得等に限られる。) また、我が国には、給与所得の精巧な年末調整制度が導入されているため、給与以外の所得のない多くの給与所得者は、源泉徴収のみで納税手続が完結しており、確定申告を行っていない。ただし、現在、給与収入が2,000万円を超える給与所得者については、申告が義務付けられているため、給与収入2,000万円を超える給与所得者の税務データは、今回提供されるデータに含まれており、高額所得者の所得分布の分析等を行う場合には問題がない。

もう一つの留意点としては、今回のデータが、税務行政のための行政記録のため、税務行政上、必要ないデータが含まれていない点がある。例

えば、確定申告書第一表・第三表には、家族構成に関するデータは含まれておらず、要因分析等を行う場合に、一定の制約が存在する。税務データの利用に当たっては、これらの点に留意する必要がある。

筆者も含めた研究グループでは、これらの留意点に注意しつつ、①パレート係数の推計等の高額所得者を含む納税者の所得分布の分析、②Gruber and Saez (2002)・Weber (2014)の方法およびバンチング・ノッチに基づく方法による課税所得の弾力性(ETI)と粗所得の弾力性(EGI)の推計、③所得税の再分配効果(特に控除の再分配効果)の分析、④①~③の分析結果を踏まえた我が国の最適な所得税制の考察、⑤マクロ経済学の観点から課税所得や粗所得の弾力性の推計値を用いた異質な個人が存在する動学的一般均衡モデルによるラッファーカーブ、平均限界税率および税収弾性値の推定を行っている。

なお、申告所得税の税務データの研究での利用は、我が国では、我々の研究グループが最初なので、まずは、税務データの整備を進める必要があった。その際、重複データ等を除外する通常のデータクリーニングに加え、税務データが、元々、税務行政を目的としたものであることから、税務行政上、無視できる場合でも、経済分析に用いる際には、チェックが必要となる項目が存在し、そうした問題への対処も必要になる。このため、税務大学のスタッフとも相談しつつ、時間をかけて、データの整備を進めている<sup>12)</sup>。

## Ⅶ. 申告所得税データ利用の成果の一例： 最近の高額所得者の所得分布の分析

このように、税務データの整備に時間を要しているものの、国税庁統計年報等の集計データでは把握できなかった点についても、次第に明らかになってきている。国税庁統計年報では得ることができない情報の代表的な事例としては、高額所得者の詳細な分布がある。上述したように、超高額所得者への所得集中の水準を測るために、高額所得者の所得分布は一般にパレート分布で近似できることから、そのパレート係数を推計する方法が一般的だが、高額納税者公示制度が廃止されてから、我が国では、その推計が難しくなっていた<sup>13)</sup>。また、高額納税者公示制度においては、所得の内訳等は明らかでなく、所得分類ごとの分布も不明であった。そうした点についても、今回初めて利用できるようになった申告所得税の税務データを用いれば、推計が可能になる。

その一例を示す観点から、我が国の高額所得

者の所得分布につき、資本所得・労働所得別のパレート係数の推計を以下、示すこととする。

（我が国の高額所得者の本格的な分析は、Kunieda (2025) を参照されたい。）Saez and Stantcheva (2018) の推計値と比較するため、確定申告書の申告所得を Saez and Stantcheva (2018) の資本所得の定義に沿って、資本所得と労働所得に分類する。具体的には、確定申告書第三表の短期譲渡所得（一般分・軽減分）、長期譲渡所得（一般分・特定分・経課分）、一般株式等の譲渡所得、上場株式等の譲渡所得、上場株式等の配当等、先物取引所得、山林所得に加え、確定申告書第一表の事業所得（営業等・農業）、不動産所得、利子所得、配当所得及び総合譲渡・一時所得を合計する。その他の所得（給与所得、雑所得、退職所得）を労働所得とする<sup>14)</sup>。

パレート係数を推計する方法は、上述のよう

---

12) なお、2025年度より、国税庁は、別途、研究者の利便性向上のため、申告情報に匿名化処理を行ったデータ（匿名データ）の提供を開始した。国税庁の「匿名データの利用の手引き」によれば、提供されるデータは、所得税の確定申告書の第一表・第三表のデータに対し、匿名化処理を行ったデータで、2014～2018の課税年分につき提供が行われる。ただし、本論文で用いているデータと異なり、この匿名データは、申告件数全体の1%を抽出したもので、各項目の99.5%タイルの値をもとにトップコーディングまたはボトムコーディングが実施されたものとなる。また、課税年分ごとのデータとなっており、個人別の時系列の分析はできないとされている。サンプリングされた申告情報という点では、この匿名データは、本論文での米国の税務データの学術利用における Public Use File (PUF) と似た性格を有するものと考えられよう。国税庁による匿名データの提供の詳細については、国税庁のホームページに掲示されている「匿名データの利用の手引き」等を参照されたい。

13) なお、最近の高額所得者の所得分布の研究では、後述の(1)式で示されるパレート関数（「第1種パレート関数」とも呼ばれる。）よりも、第2種パレート関数や一般化パレート関数（generalized Pareto function）を用いた分析（Jenkins (2017), Blanchet, Piketty and Fournier (2022)), Charpentier and Flachaire (2022))も行われているが、本論文では、過去の我が国での先行研究と比較するため、パレート関数（第1種パレート関数）に基づく分析に絞って論じている。（高額所得者の分析に用いられる分布関数のサーベイとしては、Hlasny (2021) を参照されたい。）

14) Saez and Stantcheva (2018) も認めているように、営業所得・農業所得等は、実際には、納税者の保有する資本や農地からの収益と、納税者自身の労働に対する報酬（労働所得）が混在した所得である。しかし、これらの所得を資本所得と労働所得に分割することも困難なため、Saez and Stantcheva (2018) では、本文での述べたような分類を行っており、本論文も比較のため、同じ定義で、資本所得と労働所得を分類している。

にいくつか存在する。経済学・ファイナンスの分野においては、いわゆるランクサイズ回帰（サンプルの大きい順に並べて順位（ランク）を付し、サンプルの規模の対数値を、順位の対数値に回帰させ、その係数を推計する手法）に基づく分析が多く見られる（Gabaix（2009））。しかし、ランクサイズ回帰に基づく推計には、Bauke（2007）や Clauset, Shalizi and Newman（2009）により、精度が低いことが指摘されている。また、ランクサイズ回帰では、順位と規模の両方の対数値での散布図を示し、その傾きが一定になることに基づき、パレート分布を含むべき乗則（power law）が成立しているとみなすことが多いが、Clauset, Shalizi and Newman（2009）は、対数正規分布や指数分布の場合にも、散布図が視覚的には同様に見えることを指摘している。所得分布は、一般に、低・中所得層の所得分布は対数正規分布に従い、高所得者層の所得分布はパレート分布に従うと考えられているため、この指摘は重要である。ランクサイズに代わる方法として、Bauke（2007）や Clauset, Shalizi and Newman（2009）は、最尤法に基づく推計を推奨している。連続型のパレート分布の確率密度関数は、(1)式で示される。

$$f(x) = \frac{\alpha x_{min}^{\alpha}}{x^{\alpha+1}} \quad (1)$$

（ここで、 $\alpha$  はパレート係数、 $x_{min}$  は分布の下限。）

この場合の最尤法に基づく推定量は、(2)式となる<sup>15)</sup>。（蓑谷（1998）、Clauset, Shalizi and Newman（2009））

$$\hat{\alpha} = n \left[ \sum_{i=1}^n \ln \left( \frac{X_i}{X_{min}} \right) \right]^{-1} \quad (2)$$

所得階層別の集計値しか利用可能でない場合は、ランクサイズ回帰による推計に拠ることは致し方ないとしても、租税回避等の可能性を無

視すれば、ほぼ全数調査の個票が含まれる税務データが利用可能な場合には、最尤法に基づく推計が適当と考えられる。今回の研究においても、まず、分布の下限を5,000万円から10億円の間で、5,000万円単位で設定して、2020年度の合計所得、資本所得および労働所得について、パレート係数の最尤推定量の推計を行った<sup>16)</sup>。推計結果は、図8である。

まず合計所得のパレート係数は、2億円を超えたあたりから、1.45程度で安定している。これに対し、資本所得のパレート係数は、2億円を超えたあたりから、1.35程度となる。他方、労働所得のパレート係数は、2億円を超えたあたりから、1.95程度で変動している。合計所得のパレート係数が、資本所得のパレート係数に近いのは、高額所得者の所得の大半が、資本所得だからである。

本研究の利用できる税務データは、2014年から2020年の7年分であり、パレート係数の推移を見ることによって、所得格差の動向について知ることもできる。各年の合計所得・資本所得・労働所得の2億円を下限と設定した場合のパレート分布の最尤法による推計値の推移を示したのが、図9である。

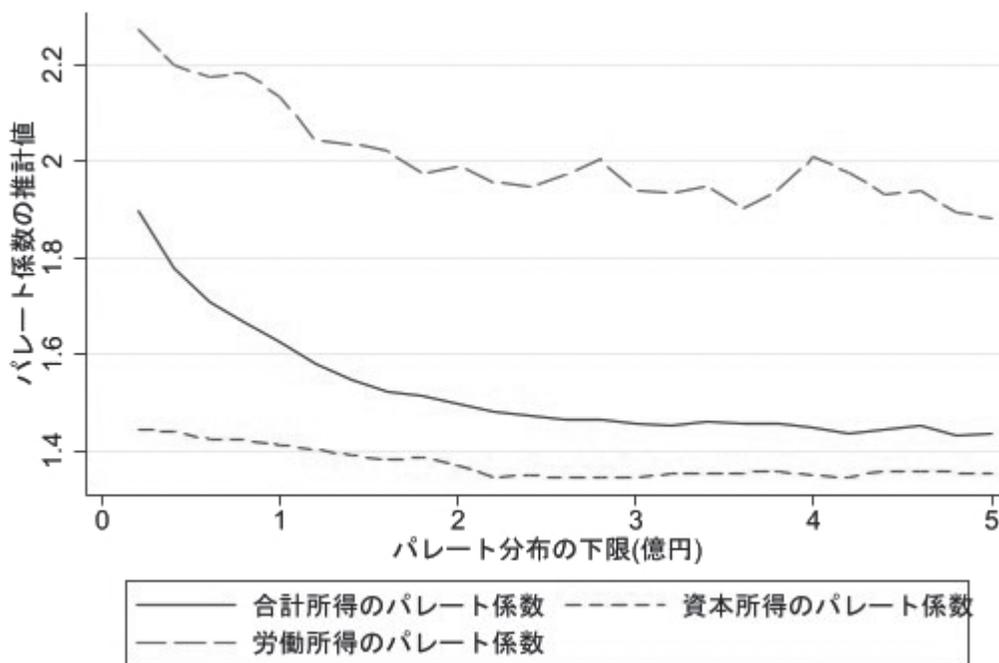
同図にあるように、合計所得のパレート係数は、1.5程度を中心に推移しているが、2014年～2020年の間に低下の傾向が見られる。資本所得のパレート係数も、1.35程度を中心に推移しており、同期間に低下の傾向が見られる。他方、労働所得のパレート係数は、2程度を中心に推移しており、はっきりしたトレンドは見られないように思われる。パレート係数の低下は、超高額所得者への所得集中を意味しており、今回の税務データの分析により、2014年～2020年の間に、我が国においては、資本所得を中心に、徐々に超高額所得者への所得集中が進んでいることが明らかになった。

また、今回の合計所得のパレート係数の推計値

15) Hill estimator とも呼ばれる。（Hill（1975））

16) このような図は、“Hill plot”とも呼ばれ、パレート分布の下限値を求める際に用いられる。

図8 2020年の各種所得のパレート係数の推計値



(出所) 筆者作成

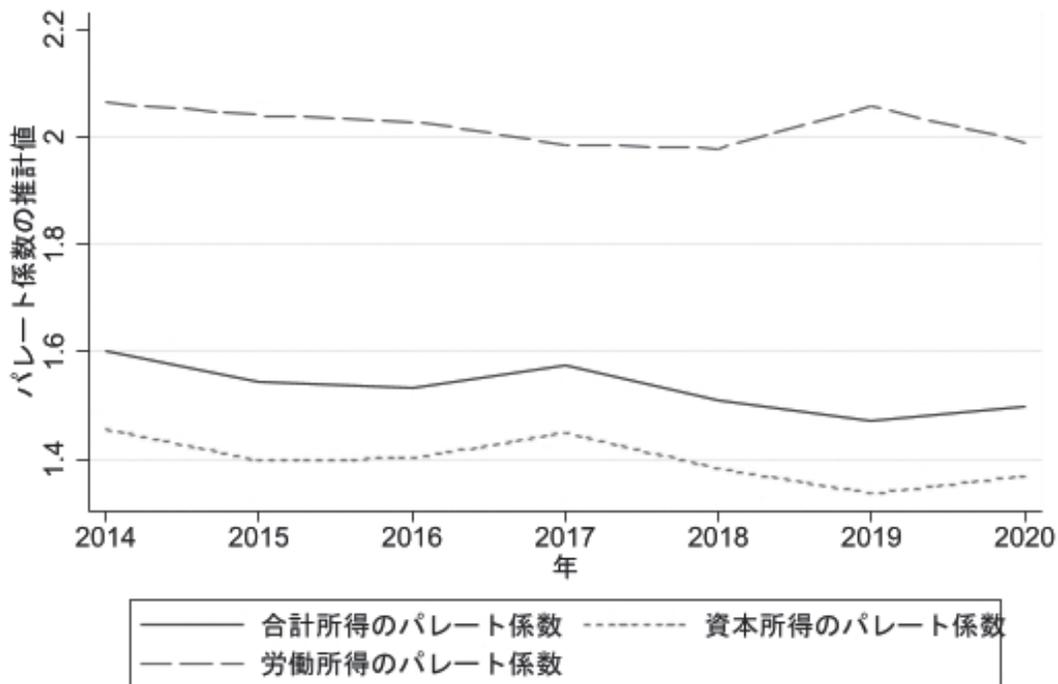
(15)は、高額納税者公示制度中止直前の2003年の高額納税者公示データに基づく國枝(2012)のパレート係数の推計値2.1と比較すると大幅に低くなっている。1.5というパレート係数は、バブル期と比較しても、Fujiwara et al. (2003)のバブル期の推計値よりも低く、Nirei and Souma (2007)のバブル崩壊前後(1989年~1991年)の推計値並みに低い水準である。さらに歴史を遡っても、Ono and Watanabe (1976)の推計に基づく農地解放や財閥解体前の太平洋戦争直前のパレート係数よりも低い数値である。先行研究とは対象データや推計手法が異なるため、単純な比較ではできないものの、パレート係数の推計値の推移からは、我が国において、2000年代前半以降、高額所得者への所得集中が大幅に進んだと考えることができる。パレート係数の水準からは、2020年時点の超高額所得者への所得集中の程度は、バブル期並みになっている可能性があると言える。

国際的な観点からは、本研究のこれらの我が

国の推計値は、Saez and Stantcheva (2018)の推計値(合計所得約1.4、資本所得約1.38及び労働所得約1.6)と比較すると、資本所得のパレート係数は、我が国の方が小さく、合計所得のパレート係数は、我が国の方が若干大きい。他方、労働所得のパレート係数は、日米で明確な差があり、我が国のパレート係数の方が、米国のパレート係数よりもかなり大きい。これは、米国企業の経営者等の報酬が一般社員よりも非常に高いのに対し、日本企業の経営者等の報酬がある程度抑制されていることが背景にあると思われる。日米の給与体系の違い等から、我が国の所得格差は小さいとの指摘がなされることも少なくないが、実際には、我が国の超高額所得層の所得の大半は、資本所得であり、資本所得まで勘案して、所得格差を考える必要がある。

なお、今回のパレート係数の推計は、株式譲渡益や土地譲渡益を含む「合計所得」を中心に行っているが、高額所得者への所得集中の分析

図9 2億円を下限とした場合のパレート係数の推計値の推移（2014～2020）



(出所) 筆者作成

において、株式譲渡益や土地譲渡益を含むべきかについては、議論がある。例えば、Moriguchi and Saez (2008) は、戦前からの長期のトレンドを見る観点から、戦前には課税対象とされず、戦後も様々な税制上の特別措置の対象となっていたこと、実現された譲渡益は変動が大きいこと等の理由から、株式譲渡益や土地譲渡益を除外した所得に着目して分析を行っている。また、同論文では、高額所得者の所得が、国民全体の所得に占めるシェアを推計することが主題となっているが、その場合、国税庁統計年報ではカバーされない課税最低限を下回る家計の所得データも必要となり、SNA（国民経済計算）の所得データを利用せざるをえなくなる。国民経済の生み出した付加価値を計算しようとするSNAでは、資産の売買に伴う譲渡益は、所得とみなされておらず、譲渡益を含んだ所得では、比較が難しいという事情もあるものと考えられる。

しかし、最適資本所得課税を含む最適な税制を検討する場合、現在の高額所得者の所得の大半が譲渡益を含む資本所得であることに鑑みれば、株式譲渡益や土地譲渡益を除いて、資本所得課税を考えてもあまり意味がない。このため、Saez and Stantcheva (2018) においては、実現した譲渡益を含む全ての資本所得につき、パレート係数を推計し、最適資本所得税率を導出している。また、過去においては、株式会社の自社株式取得は事実上制限されていたが、現在では、自己株式取得により株式譲渡益の形で利益還元を行うことは、配当と並び、株式会社の主要な利益還元の方法であり、株式譲渡益を除外し、配当のみを勘案して、資本所得を捉えることには問題が多い。さらに、過去の高額所得者・納税者公示制度では、譲渡益を含んだ所得あるいはそれに対する納税額が公表されており、過去のパレート推計との比較を行うためにも、株式譲渡益・土地譲渡益を含んだ所得概念

での分析が望ましい。こうした点に鑑み、我が国における最適な税制の考察を最終的な目的とする本研究では、株式譲渡益・土地譲渡益を含んだ合同所得および資本所得のパレート係数につき推計している。（ただし、上述のとおり、総合課税対象分の長期の譲渡益および一時所得

については、所得税法上の規定に従い、1/2のみを加算している。）

以上の点も含め、Kunieda（2025）は、所得税データを用いて、より詳細な我が国の高額所得者の分析を行っており、様々な興味深い点を指摘している。

## Ⅷ. 結びに代えて：今後の申告所得税の分析

本稿においては、税務データの学術利用の意義につき、高額所得者の所得分布の分析の例を中心に説明した。欧米での税務データの学術利用の現状やそれを利用した様々な実証研究、我が国の過去の高額所得者の所得分布の研究を述べた後、今回の国税庁が開始した税務データを利用した共同研究につき概要を説明した。その後、申告所得税データを利用した研究の成果の一例として、高額所得者の所得分布のパレート係数の推計を説明した。我が国の2020年の高額所得者の所得分布のパレート係数の推計値は、過去の水準よりも大幅に低い1.45程度となっているが、これは、超高額所得者への所得集中が、資本所得を中心に進んだことを示している。この点は、今回、国税庁との共同研究で申告所得税の税務データの利用が可能になったことにより、明らかになったことである。我が国における税務データの学術利用には、学術面

のみならず、政策面においても重要な意義があると考えられる。

高額所得者の所得分布の分析も含め、我々の研究グループによる申告所得税の税務データを利用した分析は、現在、進行中である。今後、本稿で紹介した高額所得者を含む所得分布の分析に加え、課税所得の弾力性（ETI）と粗所得の弾力性（EGI）の推計、所得税の再分配効果（特に控除の再分配効果）の分析、我が国の最適な所得税制の考察、課税所得や粗所得の弾力性の推計値を用いた異質な個人が存在する動学的一般均衡モデルによるラフファーカープ、平均限界税率および税収弾力性の推定等のトピックについても、引き続き研究を行っていく。得られた研究成果については、学会その他の機会を通じ、公表してまいりたい。本稿を通じ、我が国における税務データの学術利用の意義につき広く理解を得られれば幸いである。

## 参 考 文 献

伊藤伸介（2016）「政府統計における個票データの提供と秘密保護について—イギリスを例に—」、『経済学論纂（中央大学）』、第56巻第5・6合併号、1-19頁  
伊藤伸介（2017）「公的統計における行政記録データの利活用について—デンマーク、オランダとイギリスの現状—」『経済学論纂（中

央大学）』第58巻第1号、1-17頁  
伊藤伸介（2018）「公的統計マイクロデータの利活用における匿名化措置のあり方について」、『日本統計学会誌』、日本統計学会、第47巻第2号、77-101頁  
大竹文雄・小原美紀（2009）「所得格差」樋口美雄編、『バブル/デフレ期の日本経済と経

- 濟政策 6：労働所得と所得分配』慶応義塾大学出版会, 253-285 頁
- 北村行伸・宮崎毅 (2010) 「日本における限界税率の課税所得弾力性と最適所得税率：全国消費実態調査の個票データによる分析」Global COE Hi-Stat Discussion Paper Series No. 150, 一橋大学経済研究所。
- 國枝繁樹 (2007) 「最適所得税理論と日本の所得税制」『租税研究』第 690 号, 69-82 頁。
- 國枝繁樹 (2012) 「新しい最適所得税理論と日本の所得税制」日本経済研究 第 67 号, 21-38 頁
- 國枝繁樹 (2023) 「経済格差と資産課税」, 証券税制研究会編『日本の家計の資産形成—私的年金の役割と税制のあり方』, 中央経済社
- 國枝繁樹・米田泰隆 (2023) 「日本の所得税制に関する税務データに基づく分析の意義」(230100-01ST), 税務大学校ディスカッションペーパー (230100-01ST)
- 小林良行 (2012) 「公的統計マイクロデータ提供の現状と展望——一橋大学での取り組みをもとに——」, 日本統計学会誌, 第 41 巻第 2 号, 401-420 頁
- 汐見三郎・宗藤圭三・毛利英於菟・武田長太郎 (1941) 『改訂 國民所得の分配』, 有斐閣
- 統計データ機密保護に関する国連欧州経済委員会/EU 統計局合同ワークセッション (2006) 『統計データ開示抑制に関する用語集 改訂版, 2005 年 8 月』(翻訳：独立行政法人統計センター研究センター, 2006 年 6 月)
- 馬場義久 (2016) 「スウェーデンの納税者番号制度」, 『納税環境の整備』, 日税研論集, 第 67 巻, 103-154 頁
- 藤田晴 (1992) 『所得税の基礎理論』, 中央経済社
- 星野伸明 (2010) 「公的統計マイクロデータ提供制度の課題」, 日本統計学会誌, 第 40 巻, 第 1 号, 23-45 頁
- 溝口敏行 (1987) 「日本の高額所得者の分布」, 『経済研究』第 38 巻第 2 号, 130-138 頁
- 箕谷千鳳彦 (1998) 『すぐに役立つ統計分布』, 東京図書
- Atkinson, A., T. Piketty, and Saez (2011), “Top Incomes in the Long Run of History,” *Journal of Economic Literature* Vol. 49, No 1, pp. 3-71
- Bachas, P., and A. Jensen (2023), “Administrative Data and Methodologies for Tax Policy,” Tax for Growth, International Growth Centre, London School of Economics and Political Science
- Bauke, H. (2007), “Parameter Estimation for Power-law Tail Distributions by Maximum Likelihood Methods,” *European Physical Journal B*, Vol. 58, pp. 167-173
- Blanchet, T., T. Piketty and J. Fournier (2022), “Generalized Pareto Curves: Theory and Applications,” *Review of Income and Wealth*, Vol. 68, No. 1, pp. 263-288
- Blockmeyer, A. (2019), “Working with Administrative Tax Data: A How-to-get-started Guide,” *MTI Practice Notes*, March 2019, Number 7, pp. 1-7, World Bank Group
- Bryant, Victoria L. John L. Czajka, Georgia Ivsin, and Jim Nunns (2014), “Design Changes to the SOI Public Use File (PUF)”, Prepared for the “New Resources for Microdata-Based Tax Analysis” Session, 2014 Annual Conference on Taxation, National Tax Association Santa Fe, New Mexico. (Available online at <https://www.irs.gov/pub/irs-soi/14 rfpufredesignrecommen.pdf>)
- Bowen, C., V. Bryant, L. Burman, S. Khitatrakun, R. McClelland, L. Mucciolo, M. Pickens, and A. Williams (2022), “Synthetic Individual Income Tax Data: Promises and Challenges,” *National Tax Journal*, Vol. 75, No. 4, pp. 767-789
- Burman, L., B. Johnson, V. Bryant, G. MacDonald, and R. McClelland (2024), “Protecting Privacy and Expanding Access in a Modern Administrative Tax Data System,” *National Tax Journal*, Vo. 77, No. 4, pp. 927-947
- Card, D., R. Chetty, M. Feldstein, and E. Saez

- (2010), "Expanding Access to Administrative Data for Research in the United State," mimeo for National Science Foundation 10-069 call for white papers on "Future Research in the Social, Behavioral & Economic Sciences."
- Charpentier, A., and E. Flachaire (2022), "Pareto Models for Top Incomes and Wealth," *Journal of Economic Inequality*, Vol. 20, pp. 1-25
- Chetty R, J. Friedman, T.Olsen, and L. Pistaferri L. (2011), "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics*, Vol. 126, pp. 749-804
- Chetty, R., J. Friedman, S. Leth-Petersen, T. Nielsen and T. Olsen (2014), "Active vs. Passive Decisions and Crowd-Out in Retirement Saving Accounts: Evidence from Denmark," *Quarterly Journal of Economics*, Vol. 129, No. 3, pp. 1141-1220
- Clauset, A., C.R. Shalizi, and M.E.J. Newman (2009), "Power-Law Distributions in Empirical Data," *SIAM Review*, Vol. 51, No. 4, pp. 661-703, Society for Industrial and Applied Mathematics
- Devereux, M., L. Liu, and S. Loretz (2014), "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records," *American Economic Journal: Economic Policy*, Vol. 6, No. 2, pp. 19-53
- Devereux, M., G. Maffini, and J. Xing (2018), "Corporate Tax Incentives and Capital Structure: New Evidence from UK Firm-level Tax Returns," *Journal of Banking & Finance*, Vol. 88, No. C, pp. 250-266
- Diamond, P. (1998), "Optimal Income Taxation: An Example with U-shaped Pattern of Optimal Marginal Income Tax Rates," *American Economics Review*, Vol. 88, No. 1, pp. 83-95
- Dwenger, N., and V. Steiner (2012), "Profit Taxation and the Elasticity of the Corporate Income Tax Base: Evidence from German Corporate Tax Return Data," *National Tax Journal*, Vol. 65, No. 1, pp. 117-150
- Dwenger, N., and V. Steiner (2014), "Financial Leverage and Corporate Taxation: Evidence from German Corporate Tax Return Data," *International Tax and Public Finance*, Vol. 21, No. 1, pp. 1-28
- Feenberg, D., and J. Poterba (1993), "Income Inequality and the Incomes of Very-high Income Taxpayers," J. Poterba (ed.), *Tax Policy and the Economy*, Vol. 7, pp. 145-177.
- Feldstein, M. (1995), "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economics*, Vol. 103, No. 3, pp. 551-72
- Feldstein, M. (1999), "Tax Avoidance and the Deadweight Loss of the Income Tax," *Review of Economics and Statistics*, Vol. 81, No. 4, pp. 674-680.
- Fujiwara, H., and Y. Ogawa (2018), "Estimating Compensation of Employed Based on Taxation data," *Japanese Economic Review*, Vol. 69, No. 4, pp. 394-413
- Fujiwara, Y., W. Souma, H. Aoyama, T. Kaizoji, and M. Aoki (2003), "Growth and fluctuations of personal income," *Physica A* 321, pp. 598-604.
- Gabaix, X. (2009), "Power Laws in Economics and Finance," *Annual Review of Economics*, No. 1, pp. 255-293.
- Gabaix, X., and R. Ibragimov (2011), "Rank — 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business & Economic Statistics*, Vol. 29, No. 1, pp. 24-39
- Goolsbee, A. (2000), "What Happens When You Tax the Rich? Evidence from Executive Compensation," *Journal of Political Economy*,

- Vol. 108, No. 2, pp. 352-378
- Gordon, R., D. Joulfaian, and J. Poterba (2016), "Revenue and Incentive Effects of Basis Step-Up at Death: Lessons from the 2010 "Voluntary" Estate Tax Regime," *American Economic Review*, Vol. 106, No. 5, pp. 662-667
- Gruber, J. and E. Saez (2002) "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics* Vol. 84, pp. 1-32.
- Hill, B. (1975), "A Simple General Approach to Inference about the Tail of Distribution," *Annals of Statistics*, Vol. 3, No. 5, pp. 1163-1174
- Hlasny, V. (2021), "Parametric Representation of the Top of Income Distributions: Options, Historical Evidence, and Model Selection," *Journal of Economic Surveys*," Vol. 35, pp. 1217-1256
- Jenkins, S. (2017), "Pareto Models, Top Incomes and Recent Trends in UK Income Inequality," *Economica*, Vol. 84, pp. 261-289
- Kleven, H., and E. Schultz (2014), "Estimating Taxable Income Responses Using Danish Tax Reforms," *American Economic Journal: Economic Policy*, Vol. 6, No. 4, pp. 271-301
- Kleven H, and Waseem M. (2013), "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *Quarterly Journal of Economics*, Vol. 128, pp. 669-723
- Kopczuk, W., and E. Saez (2004), "Top Wealth Shares in the United States, 1916-2000: Evidence from Estate Tax Returns," *National Tax Journal*, Vol. 57, No. 2, Part 2, pp. 445-487
- Kunieda, S. (2025), "A Tax-data based Analysis of Japanese High-income Earners," Discussion paper, 250100-01ST National Tax College of Japan
- Lindsey, L., (1987), "Individual Taxpayer Response to Tax Cuts: 1982-1984," *Journal of Public Economics*, 33, pp. 173-206
- Liu, L., T. Schmidt-Eisenlohr and D. Guo (2017), "International Transfer Pricing and Tax Avoidance : Evidence from Linked Trade-Tax Statistics in the UK," International Finance Discussion Papers 1214, Board of Governors of the Federal Reserve System
- Mirrlees, J.A. (1971), "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, Vol. 38, No. 114, pp. 175-208
- Miyazaki, T., and R. Ishida (2022), "Estimating the Elasticity of Taxable Income: Evidence from Top Japanese Taxpayers," *Japan and World Economy*, Vol. 61, pp. 101-116.
- Moriguchi, C., and E. Saez (2008), "The Evolution of Income Concentration in Japan, 1886-2005: Evidence from Income Tax Statistics," in A.B. Atkinson and T. Piketty eds. *Top Incomes; A Global Perspective*, Oxford University Press, pp. 76-170
- Nirei, M. and W. Souma (2007), "A Two Factor Model Of Income Distribution Dynamics," *Review of Income and Wealth*, Vol. 53, No. 3, pp. 440-459.
- Ono, A. and T. Watanabe (1976), "Changes in Income Inequality in the Japanese Economy", in H. Patrick (ed.), *Japanese Industrialization and its Social Consequences*, Berkeley and Los Angeles: University of California Press.
- Perez-Truglia, R. (2020), "The Effects of Income Transparency on Well-Being: Evidence from a Natural Experiment," *American Economic Review*, Vol. 110, No. 4, pp. 1019-1054
- Piketty, T. (2014) (translated by A. Goldhammer), *Capital in the Twenty-First Century*, Belknap Press
- Piketty, T., and E. Saez (2003), "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, Vol. 118, No. 1, pp. 1-39.
- Saez E. (1999), "Do Taxpayers Bunch at Kink

- Points?,” NBER Working Paper No. 7366
- Saez, E. (2001), “Using Elasticities to Derive Optimal Tax Rates,” *Review of Economic Studies*, Vol. 68, No. 1, pp. 205-229
- Saez, E. (2004), “Reported Incomes and Marginal Income Tax Rates, 1960-2000: Evidence and Policy Implications,” J. Poterba, ed. *Tax Policy and the Economy*, Vol. 18, pp. 117-174.
- Saez E. (2010), “Do Taxpayers Bunch at Kink Points?,” *American Economic Journal: Economic Policy* Vol. 1, No. 2, pp. 180-212
- Saez, E., J. Slemrod, and S. Giertz (2012), “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: Critical Review,” *Journal of Economics Literature*, Vol. 50, No. 1, pp. 3-50
- Saez, E., and S. Stantcheva (2018), “A Simple Theory of Optimal Capital Taxation,” *Journal of Public Economics*, Vol. 162, pp. 120-142
- Saez, E., and G. Zucman (2018), “Creating Homogeneous Synthetic Individual Tax Files for Distributional Analysis,” mimeo, UC Berkeley
- Vartivarian, S., J.L. Czajka, and M. Weber (2007), “Measuring Disclosure Risk and an Examination of the Possibilities of Using Synthetic Data in the Individual Income Tax Return Public Use File,” 2007 SOI Paper Series, Statistics of Income, Internal Revenue Service
- Weber, Caroline (2014) “Toward Obtaining a Consistent Estimate of the Elasticity of Taxable Income Using Difference-in-Differences,” *Journal of Public Economics* Vol. 117, pp. 90-103