

業務データを用いた給与収入格差分析*

北尾早霧 [†]	鈴木通雄 [‡]	山田知明 [§]
政策研究大学院大学	東北大学	明治大学

2025 年 8 月 12 日

* 本稿は 2024 年 9 月 26 日に開催されたカンファレンス「Economic Research using Administrative Data」での報告内容を基に、大幅な加筆・修正を行ったものである。上田淳二、川口大司、郡司大志、小林慶一郎、近藤絢子、富浦英一、平賀一希、深尾京司、福田隆巳、別所俊一郎、正木祐輔の各氏及びカンファレンス参加者、学術変革領域研究 (B)22B101 参加メンバー、DP 検討会参加者から有益なコメントを頂いた。ここに記して感謝の意を表したい。本研究は JSPS 科研費 J22H05008 の助成を受けたものである。

[†] Email: sagiri.kitao@gmail.com.

[‡] Email: michio.suzuki.a3@tohoku.ac.jp.

[§] Email: tyamada@meiji.ac.jp.

概要

本研究は、日本の地方自治体が保有する業務データ（約 550 万人、24 自治体）を用いて、所得格差の実態とその推移を分析する。従来のサーベイデータでは把握が困難であった高・低所得層を含む大規模な業務データパネルを活用し、給与収入の格差、時系列での変化、ライフサイクルにおける格差の特徴を詳細に分析する。2011～2023 年のデータから、コロナ禍の一時的な所得低下を挟みつつ、多くの自治体で格差縮小傾向が見られた。また、年齢とともに所得水準・ばらつきが増加し、地域差も大きいことが確認された。本研究は、日本における格差研究において行政データの活用可能性を示すものである。

Keywords: 地方自治体業務データ、所得格差、ライフサイクル

JEL Classification: J31, D31, C23

1 はじめに

近年、税務データのような行政活動における業務データ (administrative data) を用いた経済格差の分析が注目されている。これまでの経済格差研究では、政府や大学などの研究機関が収集している個票データをもちいた研究が中心であったが、そのようなサーベイデータを用いた格差研究には、サンプルサイズやセレクションバイアスなどの問題から、いくつかの側面で限界があることも指摘されていた。

本研究の目的は日本の地方自治体が保有・管理している業務データを用いて、日本の所得格差の推移を理解することにある¹。我々が利用するデータセットは、原則として、それぞれの自治体に居住する全ての人の収入及び所得に関する情報が含まれている。すなわち、ある自治体に住んでいて税金を払っている労働者だけでなく、子どもや引退した高齢者なども含めて、住民基本台帳に記録がある全ての個人が対象となる。従来のサーベイデータでは、調査方法によっては超富裕層や低所得層が多く脱落してしまい、結果として、経済格差問題でもっともフォーカスしたい分布の両テールの理解が不十分になる可能性が指摘されていた。業務データはこの問題を大きく改善することができる可能性がある。本論文では業務データに基づいた日本の経済格差の特徴を示し、その結果を様々な先行研究と比較することによって、従来の格差研究を補完するとともに、業務データを用いた格差研究の可能性を示す。

経済格差の測り方は非常に多様であるが、先行研究では主に、時間を通じた日本全体 (あるいは特定の属性を持った集団) の経済格差の推移と、各個人の様々な経済変数が年齢とともにどのように変化していくかというライフサイクル方向の変化から分析されてきた。我々は、時系列側面とライフサイクル側面の両方に関心があるが、サンプル期間の問題から本研究ではライフサイクル側面により注力している。これは、これから詳細に説明するように、他国の業務データでは数十年程度のサンプル期間があり時系列方向の変化、すなわち経済成長の影響や景気循環との関係性を把握するのに十分な期間がある一方、我々が利用する日本のデータでは制度面の制約から最長でも 2011 年から 2023 年までの期間に限られるためである。

業務データを用いた本研究の強みと弱みは以下のようにまとめられる。サーベイデータを用いた従来型の研究と比較した最大の強みは、サンプルサイズが極めて大きい点である。一部の自治体で

¹ 本論文は、東京大学政策評価研究教育センターが行っている「EBPM 推進のための自治体税務データ活用プロジェクト」及び科学研究費補助金 学術変革領域 (B)「税務データを中心とする自治体業務データの学術利用基盤整備と経済分析への活用」に基づいた研究である。プロジェクトの概要については下記の HP を参照してほしい。
https://web.iss.u-tokyo.ac.jp/jichitai_data/

は全住民のうち約 50% のデータ提供にとどまっている自治体が混在しているものの、多くの自治体では全住民が対象となっている。すなわち、ある特定の自治体については、サンプル調査ではなく全体像を把握することが可能である。また、自治体内で個人を識別する ID が存在していることからパネル化も可能であり、個人の給与収入の推移を正確に把握できる。加えて、サンプルサイズが十分に大きいことから三次のモーメントである歪度や四次のモーメントである尖度といった、従来のサンプルサイズでは困難であった高次のモーメントの推定も可能となっている。一方、日本における業務データを用いた研究には、従来の研究と比較して、弱みも存在する。本プロジェクトへの参加協力をしていただけた自治体が限られていることから、日本の国全体の格差動向を把握できるわけではない。また、サンプル期間が短いため、長期的な経済成長が格差に与えた影響や格差と景気循環との相関といった他国で行われている分析は困難である。加えて、本データセットは自治体が保有する住民税に関連する情報に基づいているため、給与所得以外の情報が限られている。例えば、資本所得やストックオプションなどの分析は難しい。他に、サーベイデータを用いた研究との比較では、消費や資産といった情報がない点も挙げられる。この点は業務データを用いた研究全般の問題点であるが、経済格差はフローの収入だけでなく消費や資産といった情報がないと正確な厚生分析ができない可能性がある。そのため、日本における経済格差の推移や特徴を正確に把握するためには、両者のメリットを活かした複合的な研究が必須である。

直近の日本の給与収入の推移について、税務データからわかってきたのは以下の通りである。2011 年から 2023 年にかけて、多くの自治体で平均所得は伸びているが、コロナ禍で一時的に平均給与収入は低下した。また、ほとんどの自治体で給与収入格差は低下している。そのため、一般的に言われているような、足元で経済格差が顕著に拡大しているといった傾向は見つからなかった。しかし、平均給与もジニ係数も自治体間のばらつきがかなり大きく、ジニ係数で測るともっとも高い自治体と低い自治体の給与収入ジニ係数は 0.1 程度の差がある。従来の格差研究ではサンプルサイズの問題からあまり分析が行われてこなかったが、日本における格差研究においては、自治体間の違いを考慮することの重要性が示唆される結果となった。

本論文の構成は以下の通りである。2 節では、業務データを用いた先行研究及び従来型のサーベイデータを用いた経済格差研究の動向をまとめる。3 節では、本研究で用いたデータセットの特性について説明する。4 節では、自治体間ごとの特性についてまとめている。5 節では、時系列方向から給与収入及び所得格差がどのように推移してきたのかを自治体ごとにまとめて、6 節では給与収入・所得格差のライフサイクル側面から分析を行う。7 節では、給与収入・所得格差以外の側面の特性について分析をこなう。最後に、8 節で本研究の発見をまとめた上で今後の研究の展望につ

いて議論する。

2 先行研究

■業務データと景気循環、所得格差： 近年、様々な国において業務データを用いた研究が盛んに行われている。米国の業務データを用いた初期の研究として、[Guvenen et al. \(2014a,b\)](#) がある。両研究は、米国の社会保障局 (the US Social Security Administration; SSA) から特別な許可を得て全米の 10% の労働者の給与データにアクセスしたうえで、1978 年から 2011 年における景気循環、特に景気後退が個人の収入・所得のばらつきや個人が直面する所得変動リスクにどのような影響を与えているかを分析した。[Storesletten et al. \(2004b\)](#) などが示したように、通説では、景気後退期には所得ショックの分散が増大する (counter-cyclical income shock) と考えられていた。この考え方が正しければ、一部の労働者は不況期に大幅な所得減少を経験する一方で、他の労働者は大きな所得増加を経験することが予測される。しかし、[Guvenen et al. \(2014a,b\)](#) は、所得ショックの分散は景気循環に対してあまり変化しないことを明らかにした。代わりに、所得ショックの 3 次のモーメントである歪度が強く循環的に変動することが確認された。これは、景気後退期には大幅な所得減少が頻繁に発生する一方、大きな所得増加の機会が減少することを意味する。このような発見は、サンプルサイズが十分に大きく、低所得層及び高所得層の情報をカバーした長期間のパネルデータがはって初めて可能となったといえる。その後、[Busch et al. \(2022\)](#) は、米国に加えて、ドイツ、スウェーデン、フランスの業務データに基づいて、同様の研究を行い、国ごとの景気循環と所得リスクの関係性に関する特性を明らかにしている。同じ SSA のデータセットを用いて、[Guvenen and Kaplan \(2017\)](#) は所得トップ層の不平等度の推移を分析しているほか、[Guvenen et al. \(2022b\)](#) はライフサイクル側面に注目し、1957 年から 2013 年までの長期間の個人の賃金格差について分析を行い、男性の場合は初期のキャリアにおける賃金の低下が、生涯賃金の停滞の主因であることなどを明らかにした。

[Guvenen et al. \(2021\)](#) は、公刊年は 2021 年であるが Working Paper の段階から業務データを用いた所得格差研究としてよく知られていた重要な研究で、米国の労働者数百万人の所得データを用いて、生涯にわたる所得動態の特徴を明らかにしている。特に、歪度や尖度といった、これまでにはサンプル数の問題から分析対象とはされず、注目されてこなかった高次のモーメントについて、時系列方面とライフサイクル方面の両方から分析をしている。一方、[Song et al. \(2019\)](#) は SSA が管理する機密データベースである Master Earnings Profile をもちいて、賃金格差の増大の 3 分の 2 が企業間の平均賃金の格差の拡大に起因していることを示した。

■業務データから GRID プロジェクトへ： その後、ミネソタ大学の Guvenen 教授はプリンストン大学の Violante 教授、スタンフォード大学の Pistaferri 教授と Global Repository of Income Dynamics (GRID) プロジェクトを立ち上げた。GRID プロジェクトの目的は、所得格差と所得動態に関する高品質な個人レベルのマイクロデータを、世界各国の業務データから構築し、そのデータベースを提供することにある。Quantitative Economics 誌は、2022 年に業務データに関する特集号を組み、各国における業務データの現状及びそこから明らかになった各国の労働者の賃金の推移を経済格差の特徴をまとめている。参加国は多岐にわたり、アルゼンチン (Blanco et al. (2022))、ブラジル (Engbom et al. (2022))、カナダ (Bowlus et al. (2022))、デンマーク (Leth-Petersen and Sæverud (2022))、フランス (Kramarz et al. (2022))、ドイツ (Drechsel-Grau et al. (2022))、イタリア (Hoffmann et al. (2022))、イギリス (Bell et al. (2022))、メキシコ (Puggioni et al. (2022))、ノルウェー (Halvorsen et al. (2022))、スペイン (Arellano et al. (2022))、スウェーデン (Friedrich et al. (2022))、アメリカ (McKinney et al. (2022)) が含まれている。各国の所得格差の動態の違いについては Guvenen et al. (2022a) が簡潔にまとめているが、国によってサンプルサイズや業務データがカバーする範囲、利用可能期間も様々である点も興味深い。例えば、英国は 1975 年から 2020 年までの 45 年間という長期間のデータセットになっているが、サンプルサイズはクロスセクションで 10 万人程度であるのに対して、米国やカナダではクロスセクションでサンプルサイズが数 100 万人におよぶ。

■日本における業務データの活用と国税データ： 日本においても、近年、業務データを用いた様々な実証研究が行われるようになってきた。近藤・深井 (2023) は、本研究と同じ地方自治体の業務データを用いて、既婚女性の就労行動について分析を行っている。Fukai and Kondo (2025) はチャイルド・ペナルティの観点から業務データを用いた研究を行っている。また、Kawaguchi and Toriyabe (2025) は賃金構造基本統計調査のような政府による賃金統計がどの程度の労働者をカバーしているかを評価するために、政令指定都市の賃金統計と税業務データの比較を行っており、政府による賃金統計は賃金分布の右裾も左裾も過小評価していることを指摘している。Fukuda (2025) は地方自治体の業務データの特性について詳細なレビューを行っている。

本研究を含めたこれらの研究は、住民税を徴税するために必要になる地方自治体が管理している業務データに基づいている。一方、所得税や資本課税などの国税に関する業務データについては政府がデータセットを整備しており、別のプロジェクトとして同時進行的に研究が進んでいる。国税を用いた研究としては、北村 (2025)、國枝 (2025)、宇南山他 (2025)、大野他 (2025)、國枝・米田

(2023)、Gunji et al. (2025) を参照して欲しい²。

■サーベイデータを用いた日本の格差研究： 政府統計をもちいた経済格差の研究については長期間の蓄積がある。例えば、Sudo et al. (2012) や Lise et al. (2014) は、総務省統計局が集めている家計調査と全国消費実態調査/全国家計構造調査を用いて日本の所得・消費・資産格差の実態を明らかにしている。Kitao and Yamada (2024a,b, 2025) はデータセットを延伸して、直近までの日本の経済格差の推移について考察を行っている³。

3 業務データ

3.1 国税と地方税：利用可能性と限界

労働者が労働所得を稼得した場合、金額に応じて税の支払い義務が発生するが、その徴税方法は大きく分けて2種類存在している。原則として、勤め人の場合、雇用主から給与を支払われる際に源泉徴収として、所得税や各種保険料を先に天引きされたうえで支給されている。勤め先以外から収入がない場合、場合によって年末調整などが必要になる可能性があるものの、原則として労働者から労働所得に関する追加的な報告義務や税の支払いは発生しない⁴。一方、個人事業主や勤め人でも副収入がある人、給与収入が高額の人などは確定申告を行い、自分の当該年の収入を申告する必要がある。

家計が支払う税は大きく分けて2種類存在している。それが、国税と地方税である。国税は、所得税や法人税などを指す。国税は国が徴収する税金であり、国税庁が管轄している。一方、地方税は住民税、事業税、固定資産税などが該当し、都道府県や市区町村が徴収する税金である⁵。そのため、個人がいくら収入を得ていくら税金を支払ったかという情報は、国と地方自治体の税務当局とで情報の共有が進められているものの、全ての個人の過去から現在にいたるまでの収入及び税に関する包括的なデータというものは存在しない^{6,7}。日本の税制度においては、源泉徴収票の税務署への提出義務は限定的であり、年間給与収入が500万円を超える個人に関するもののみが源泉徴収義務者（雇用主）から税務署に提出されることが法令で定められている。このため、確定申告を

² 同じく業務データである法人税データを用いた研究としては、Arata et al. (2023)、土居他 (2025)、宮川他 (2025) などがある。

³ そのほか代表的な所得・消費格差の研究として、Ohtake (2005)、Unayama (2023) 等を参照してほしい。

⁴ ただし、給与所得者であっても還付目的の確定申告を行っている人も存在する。

⁵ 消費税は2025年8月時点で10%であるが、標準化税の場合、地方消費税率2.2%を含んでいる

⁶ 例えば、日景 (2002) を参照。

⁷ 国税庁では独自に国税庁保有行政記録情報を用いた共同研究を行っている。詳細は下記のHPを参照：<https://www.nta.go.jp/about/organization/ntc/kyodokenkyu/>

行っている納税者を除いては、政府機関は個人の給与所得に関する包括的なデータベースを保持していない^{8,9}。

国税データに含まれるものとして、確定申告をした人たちの所得、例えば、高所得層（所得が2000万円超の人は確定申告が必須）、副収入が20万円以上ある人、還付目的で確定申告をした人、自営業者などの情報がある。多くの勤め人は源泉徴収だけなので、我々が知る限り、個人の情報と関連付けた給与収入及び所得に関する包括的なデータベースの形式とはなっていないようである。一方、主に住民税を徴税する必要性から、各地方自治体は個別にデータベースを作成して、住民の過去の給与収入・所得に関する情報を保管している。地方公共団体情報システムの標準化に関する法律（令和3年法律第40号）に基づいて現在進行系で情報システムの標準化が進められている。保管している期間は、国税については「国税庁標準文書保存期間基準」において7年と定められているものの、データベースへの入力の問題などもあり、実際に各自治体が提供可能な形式で保有しているデータは5年から10年以上とかなりのばらつきがある。業務データはプライバシー保護などの法律上の問題から、どこの国でも研究目的であっても利用が制限されている¹⁰。それに加えて、日本では国税と地方税という仕組みが存在することが、業務データを用いた経済格差の把握を困難にしている。一方で、日本の地方自治体は教育や医療・介護といった分野でデータの接続が可能なケースも多々あることから、研究分野によってはアドバンテージがある事例も存在している。

本研究では、各地方自治体が保管する業務に関するデータセットを利用する。具体的には、総務省関係者経由あるいはCREPEのホームページを通じて、各地方自治体に対して、本研究を含むプロジェクトの趣旨を伝えたくて任意の参加をお願いした¹¹。そこで関心を示してくれた複数の自治体のデータセットを、研究レベルで利用できるように個人情報などを削除したうえで、利用した

⁸ 2013年にマイナンバーが導入されて以降、同一個人か否かの判断ができるようになったものの、中央政府が保有する2013年以前の業務データは、同一個人をつなげることができるようにはなっていない。そのため、他国の研究で行われているような長期間のパネルデータとして業務データを活用することは困難である。また、国税庁が管理する“税に関する”データベースなので、子どもや非課税の人といった個人は含まれず、世帯構成などを把握することもできない構造になっている。これは税務署が業務を行う目的で作成されたデータベースであることからくる限界である。また、利子所得については、個人が証券会社を経由して資本所得を申告する場合は個人の同定ができないため、やはり完全な把握は不可能である。

⁹ 源泉徴収票の提出義務について、詳細は下記のHPを参照：

<https://www.nta.go.jp/taxes/shiraberu/taxanswer/hotei/7411.htm>

ただし、2023年度（令和5年度）の税制改正によって、2027年（令和9年）1月1日以降に提出すべき給与所得及び公的年金等の源泉徴収票については、給与等の支払をする者が、市町村長に給与支払報告書等を提出した場合には、税務署長に給与所得等の源泉徴収票を提出したものとみなすこととされており、源泉徴収にかかる市町村に提出されるデータが、国税庁にも一括して提供されることになるとされている。詳細は下記のHPを参照：

https://www.mof.go.jp/tax_policy/tax_reform/outline/fy2023/05taikou_01.htm

¹⁰ 例えば、共同研究者にそのデータを取り扱う内部の人間が含まれている等、従来の個票データとは異なる仕組みが取られている場合がある。

¹¹ https://web.iss.u-tokyo.ac.jp/jichitai_data/data.html を参照。

ものである。ただし、所得項目など、研究に必要な最低限のデータが提供可能な自治体に限定されている。

3.2 自治体について

本研究で実際に使用するデータセットは、参加協力をしてくれた 24 自治体である。日本の市区町村の総数は 2025 年 8 月時点で 1,741 自治体あり、24 自治体は一見かなり少ないようにみえる。しかし、多くの市区町村はサイズが小さいため、人口でカバー率を考えると決して小さな割合ではない。我々のデータセットに含まれる自治体の人口サイズは、最小の自治体で 1 万数千人程度、最大の自治体は 100 万人を超えている。各自治体の人口を合計すると 550 万人超となる。日本の人口は 1 億 2500 万人程度なので、全人口の 4.5% 程度をカバーしていることになる。なお、我々のデータセットに区と村は含まれていない。これは何らかの理由でセレクションを行ったわけではなく、あくまで自発的な参加を受け付けた結果、区と村は存在しなかっただけである。なお、参加協力の際して、全サンプル（全住民）の提供は困難であるが、ランダムな 50% サンプルであれば提供可能であるという自治体が存在した¹²。そのため、100% サンプルの自治体と 50% サンプルの自治体が混在している。24 自治体のうち、5 自治体が 50% サンプルの提供で、残りの 19 自治体で 100% サンプルが利用できる。

サンプル期間は自治体によって異なり、最短で 5 年間、最長で 13 年間のデータが提供されている。そのため、参加しているすべての自治体からデータが提供された 2017-2021 年ではサンプルサイズが 550 万程度になるが、それ以外の年だとサンプルサイズは大幅に減る。

図 1 (a) は日本における全自治体の人口サイズをプロットしたものである。横軸がそれぞれの自治体の人口規模、縦軸がその人口規模の自治体の数である。多くの地方自治体の人口は 20 万人以下で、左に大きく偏っており、規模が小さい自治体が多数を占めるのがわかる。図 1 (b) は本プロジェクトで利用可能な参加自治体の人口規模をプロットしたものである。中規模の自治体が多いが、居住者数が 100 万人を超える自治体も 2 つ含まれている。一方、人口規模が 1 万人を下回る自治体は含まれていないため、小規模な町や村は含まれていないという特徴がある。

¹² トップコーディングや同一個人を追跡したパネル化、世帯データの作成の問題などから厳密に 50% というわけではなく、“約”50% のランダムサンプリングとなっている。

3.3 利用可能な変数と個人属性

本データセットは、本来、地方自治体の税業務に使用するデータであるため、利用可能な変数は基本的に税に関連するものに限られる。ただし、税データを各地方自治体から提供してもらう際に、住民基本台帳あるいはそれに準ずる全住民をカバーしたデータに基づいた形で住民のデータとの紐づけを行っている¹³。もし純粋に税に関わる個人だけに基づくのであれば、子どもや高齢者、専業主婦などの住民税に直接関係しない層のデータは存在しないことになる。しかし、それでは世帯所得などが把握できなくなることから、住民基本台帳に紐づけを行っている。これにより、住民税を支払っていない人たちも我々の業務データ上に含まれる形になり、そういった個人の給与収入・所得の値はゼロとなっている。なお、当然のことではあるが、個人名や住所といった個人を特定化するような情報は一切含まれていない。

業務データに入っていて我々が利用可能であった個人属性は、調査年、生年(年齢)、性別、世帯主からみた関係性(続柄)、住んでいる市区町村名、個人ID、世帯ID、調査年の1月1日に当該自治体に居住していたか否か、である¹⁴。

経済に関連する変数としては、課税前の勤労所得(給与収入)と合計所得金額、総所得金額、総所得金額等の3種類の所得がある。給与以外の収入は下記の項目が利用可能である：営業等所得、不動産所得、金利、配当所得、雑所得：その他、雑所得：公的年金。ただし、これらで資本所得や財産所得がすべて把握されているわけではなく、多くの漏れが存在する可能性が高いため、本研究では基本的に「給与収入」を主に扱っている¹⁵。同様に、税控除関連の項目も利用可能であるが、一部の自治体で控除関連の変数の利用に制限があることから、分析には使用していない¹⁶。すべての変数はコアCPI(2020年基準)で実質化している。

図2はデータを提供してくれた自治体が属している都道府県の地図である。日本全国からバランスを考えて自治体を選定したわけではないが、結果的にある程度、日本全体に散らばっている。ただし、九州が多めになっている。

¹³ 日常的な税業務においても世帯情報等が必要になることから、本プロジェクトと関係なく、ある程度、税務データの中に住民基本台帳データが入っている場合が多い。

¹⁴ 続柄は住民票に記載されたものである。学生などが進学で住民票を動かさずに親元を離れて住んでいる場合などについては居住実態とズレが生じる可能性があるが、そのようなケースは多くはない。本研究で用いているデータは調査年の1月1日に当該自治体に居住していた人たちのみを用いている。

¹⁵ 例えば、証券会社を通じて資産運用していて「特定口座・源泉徴収あり」を選択している個人の場合、金額次第ではあるが、原則として確定申告が不要になる。業務データに記載されている金利や配当所得は確定申告で申告している金額に限られ、証券会社経由で申告された資本所得は含まれていない。

¹⁶ 所得控除に関連する変数として、下記の項目が利用可能である：配偶者控除、配偶者特別控除、扶養控除、医療費控除、社会保険料控除、基礎控除、控除合計。

3.4 サンプルセクション

前節で説明をしたように、本研究では主として給与収入にフォーカスしている。給与収入なので、子どもや高齢者など、働いていない世代の値は多くの場合ゼロになる。そのため、人口分布は、給与収入格差の分析に大きな影響を与える。また、給与収入は男女差もかなりの程度ある。様々な属性が混在する社会状況をみるのも経済格差分析の在り方の一つではあるが、そのような分析では何が原因で経済格差が変化しているのか理解しにくい。そのため、本研究では下記の通りのサンプルセクションを行った。まず、年齢・性別が欠損値あるいは不明の個人は落とした。58,000 程度が該当するが、全自治体の全期間をプールした際の観測値の総数は約 3500 万なので大きな影響はない。サンプルセクションに際して、[Guvenen et al. \(2021\)](#) 及び GRID プロジェクトを参考に下記の 3 ステップに従った。

1. サンプル A：年齢が 25 歳以上 59 歳以下に限定する。
2. サンプル B：サンプル A+ 男性に限定する。
3. サンプル C：サンプル B+ フルタイム労働者とみなせる個人に限定する。

まず、サンプル A によって、勤労期に限定する。これによって、引退世代（＝ゼロ給与収入）の増加に伴う格差拡大の可能性を排除する。また、サンプル B では男女間格差の問題を捨象して、男性にフォーカスする。当然、依然として存在する男女間給与格差は大きな社会問題であるが、女性は男性と比較して働き方が多様であり、それが格差指標の推移の解釈を困難にすることから、サンプル B では男性に限定する。

サンプル C について、フルタイム労働者か否かを直接判定するための変数は我々のデータセットには含まれていない。そこで、労働時間が年間 2000 時間であると仮定をして時間給を計算して、時給が最低賃金以上になる人としてフルタイム労働者を定義した。日本では最低賃金は都道府県ごとに設定されている。我々のデータセットでは市町村レベルでの居住地の情報は存在しているものの、勤め先の所在地は不明なため、厳密に都道府県別の最低賃金をあてはめることはできない。そのため、各年について全都道府県でもっとも低い最低賃金水準を用いて年 2000 時間働いた際の最低年収を算出したうえで、その水準を超える人達をフルタイム労働者と定義した。サンプル C の定義の仕方から、失業者やパートタイム労働者に加えて、事業所得が中心の個人事業主（自営業）の多くも結果的に落としている点に注意して欲しい。

表 1 は年ごとのサンプルサイズと参加自治体数の推移である。サンプル全体としては、2011 年

から 2023 年までの期間が存在している。しかし、例えば 2011 年のデータが残っていた自治体はわずか 2 自治体しかなく、サンプル数もわずか 6 万人程度となる。2012 年以降、参加自治体数は徐々に増えていき、2017 年のデータが利用可能な自治体数は 22 でサンプルサイズも 500 万人を超える。しかし、直近のデータを提供してくれている自治体も 2022 年は 7、2023 年は 3 しかないため、サンプルサイズは大きく減少する。まとめると、2017 年から 2021 年の期間はサンプルサイズも十分に大きく、日本の様々な地域を包括的に含んだデータセットと言っても差し支えない一方、その前後の期間では各自治体固有の状況に左右されて、結果が日本全体の傾向とは言えない可能性が高い。

サンプルセレクションとしては、全サンプルから 25 歳から 59 歳の勤労期に限定することで、自治体によって異なるが 49% から 64% 程度のサンプルが脱落する。自治体間でばらつきがあるのは、自治体によって人口構造と高齢化の度合いが異なるためである。男性に限定することでサンプルは更におよそ半減し、サンプル C の時点では四分の一未満になる。しかし、それでも多い年では 100 万人近いサンプルが残っているのは、他のサーベイデータにはない大きな強みである。

3.5 トップコーディング

自治体によって、プライバシー保護の観点から、提供データにトップコーディングを施している場合がある。トップコーディングは、際立って高収入の個人について、特定を避ける目的で、ある水準を超えた場合にその水準にすべての給与収入を丸めるという形で処理している。そのため、給与収入及び所得の右側の裾野が切断された形となる。

表 2 は、サンプル C を用いて、各自治体の給与収入のトップ 1%、0.1%、0.01% のそれぞれの閾値をまとめた表である。全自治体をプールしたデータセットに基づく、給与収入でトップ 1% 以内に入るためには、個人で 1910 万円以上稼ぐ必要がある。また、トップ 0.1%、0.01% に入るためにはそれぞれ 4700 万円、6080 万円以上となる。しかし、一部の自治体、例えば自治体 F ではトップコーディングを施していることから、トップ 0.1% とトップ 0.01% が同じ値となっている。これは超富裕層の給与収入を閾値である 4320 万円に丸めているためである。同様の処理が施された自治体が 9 自治体存在している。そのため、実際の閾値はより高い位置に存在すると考えられる。

4 自治体間のばらつき

はじめに、自治体ごとの特徴を簡単にまとめる。表 3 は、2019 年の自治体ごとの給与収入の平均及び様々な経済格差指標を作成したものである。表 1 で見たとおり、サンプルサイズとしてはより直近の 2020 年あるいは 2021 年も十分なサンプルサイズを確保できているが、コロナ禍の影響を考慮して、コロナ禍直前の年を選定した。我々が利用可能な自治体は 24 あるため、それぞれにアルファベットを割り当てて、匿名化をしている。全ての数値はサンプル C を用いて計算した。

自治体間の平均給与収入のばらつきは非常に大きい。給与収入が一番低い自治体では 400 万円台前半である一方、平均が 700 万円を超える自治体も存在しており、差は 270 万円以上ある。大都市圏近郊であるか否かや、通勤圏に大企業が存在しているかなどで平均給与収入は大きく異なっている。なお、25 歳から 59 歳限定であるため、自治体間の平均年齢の差に大きな違いはなく、概ね 40 歳代前半となっている。

各自治体の給与収入のジニ係数も、低いところでは 0.21 程度であるのに対して、高い自治体では 0.3 を超すなど大きな違いがある。99 パーセンタイルあるいは 90 パーセンタイルと 50 パーセンタイル (中央値) の比を取った $p99/p50$ 、 $p90/p50$ も同様に自治体間でばらつきを見せている。 $p99/p50$ や $p90/p50$ が中央値と高所得層の間の格差なのに対して、 $p1/p50$ や $p10/p50$ は低所得層との差を表している。 $p90/p50$ と $p50/p10$ の相関係数は 0.38 程度で、給与収入分布の右側である富裕層の裾野が大きい自治体と左側の裾野が大きい自治体が混在していて、格差が大きい自治体も必ずしも両端が等しく広がっているわけではない。

Kitao and Yamada (2024a, 2025) は、家計調査及び全国消費実態調査/全国家計構造調査をもちいて、日本の世帯間経済格差を分析している。家計調査や全国消費実態調査/全国家計構造調査は調査単位が世帯なので、個人所得を記録している業務データと単純比較はできない点に注意する必要がある。Kitao and Yamada (2024a) は、家計調査を用いて、家族の人数を調整した勤労世帯の等価労働所得のジニ係数を計算している。それによると、日本経済のジニ係数は 1980 年代前半は 0.24 程度であったのが、1980 年代から 2000 年代にかけて上昇し、現在は 0.28~0.29 程度になっている。業務データに基づく各自治体のジニ係数は概ねこのレンジに入っており、家計調査の結果と税務データには整合性があることが確認できる。また、同じく Kitao and Yamada (2024a) の補論に基づく、2019 年の労働所得の $P50/P10$ はおおよそ 2.0、 $P90/P50$ は 1.9 程度である。どちらも業務データよりも家計調査のほうがわずかながら数値が高くなっている。一般的に家計調査や全国消費実態調査/全国家計構造調査はいわゆる中間層の把握に優れている一方で、分布の両裾

野である貧困層と富裕層に関しては抜け落ちている可能性が指摘されている。しかし、業務データと比較した場合、この指摘は必ずしも当てはまらない。富裕層の格差指標について業務データのほうが小さくなっている理由の一つとして、表 3 はそれぞれの自治体内の格差なので、自治体間の格差が反映されない分、低めになっている可能性が挙げられる。一方で、貧困層については、業務データではサンプル C を使っていることからフルタイム労働者とみなされない人たちが脱落している事が理由の一つとして考えられる。ただし、Kitao and Yamada (2024a) でも給与所得格差は勤労世帯に限定していることから、実際に日本のフルタイム労働者に限定すれば、格差の指標は表 3 の値の範囲におさまると考えられる。

5 時系列方向からみた経済格差：2011-2023 年

次に、時系列方向から、日本の各自治体の平均給与収入及び格差がどのように推移しているかを確認していこう。表 4 はサンプルサイズ及び参加自治体数が十分に大きい 2017 年から 2021 年について、すべての観察値をプールして作成した平均給与所得の推移をまとめたものである。サンプル A である勤労期間の個人の平均給与はおよそ 300 万円台前半である。ただし、この数値は男性と女性が混在しており、例えば 2020 年であれば男性の給与収入の平均が 470.1 万円なのに対して女性は 211.8 万円と男女で倍以上の差がある¹⁷。サンプル A 及びサンプル B では、自営業・個人事業主や無職、専業主婦など給与収入がゼロ扱いになっている人たちも含まれており、それが平均値を押し下げており、特に専業主婦が含まれる女性の平均値が低めに算出される。そこでフルタイム男性労働者と考えられるサンプル C に限定をすると、平均給与収入は 580 万円から 600 万円程度になる。全体を通じて平均収入はわずかながら成長しているが、女性の平均給与収入を除いて、2020 年のコロナ禍においてわずかに収入が低下している様子が確認できる。

表 5 は、同じく 2017 年から 2021 年について、給与収入のジニ係数をまとめたものであるサンプル A においてはジニ係数は 0.5 を超えて極めて大きな値となる。これは、勤労期間に限定しているとはいえ、男女間格差やフルタイム労働者とパートタイム労働者のばらつきの影響を受けている上、自営業者などの収入を過小に評価して計算しているためである。実際、労働参加率が高い男性労働者のジニ係数は 0.42 から 0.43 程度であるのに対して、より働き方に多様性がある女性の給与収入は 0.55 程度と高い値になる。フルタイム労働者に限定したサンプル C ではジニ係数は 0.28～0.29 程度となり、家計調査や全国消費実態調査から算出された切ろう所得格差と比較可能な水準に

¹⁷ サンプル A の男性限定はサンプル B と同じである。

なる。わずか5年間の期間で確実なことは言えないが、我々が利用可能な期間を通じて、給与収入のジニ係数はごくわずかながら低下傾向をしてしており、少なくとも足元で格差が拡大している様子はうかがえない。この点は [Kitao and Yamada \(2024a\)](#) と整合的な結果である。

表6と表7は、同じく2017年から2021年の平均及びジニ係数を、給与収入ではなく合計所得金額を用いてまとめた結果である。給与収入が税引前の値であるのに対して、合計所得金額は様々な税額控除を差し引いた後の金額であるため、平均は低下する。ただし、合計所得金額には給与からの勤労所得以外の様々な所得も含まれている点に注意してほしい。給与収入の代わりに合計所得金額を用いても、時間を通じたトレンドに大きな違いはない。

では、平均給与所得あるいはジニ係数は自治体ごとにどのような特性があるのだろうか。図3は、自治体の人口サイズごとに平均給与収入と給与収入のジニ係数をプロットしたものである。人口サイズと平均給与収入には緩やかな正の関係性がみられるものの、その傾向は必ずしも強くない。大都市周辺のベッドタウンなどでは、人口規模は比較的小さいものの平均給与収入が高い自治体も見られる。一方、給与収入ジニ係数と人口規模にも正の関係性がみられ、サイズが小さい地方自治体は格差が小さく比較的同質的であるのに対して、大きな都市では給与収入のばらつきも大きくなる傾向が見て取れる。

図4～図7は、それぞれの自治体について時系列方向で給与収入の平均および各種格差変数がどのように推移したかをプロットしたものである。全てを一つの図にまとめるのは困難なので、人口サイズが10万人未満と比較的小規模の自治体、10万人以上30万人未満の中規模自治体、30万人以上の大規模な自治体に分割した。平均給与所得に関しては、自治体間でもかなりのばらつきがあるものの、全体的に緩やかな右上がりの傾向が見て取れる。ただし、2020年のコロナ禍の時期に多くの自治体の平均給与収入は一度低下をしている。自治体の人口規模が小さいほうが平均値のばらつきが大きく、人口規模が10万人未満といっても、地方の小さい市と大都市圏近郊のベッドタウン、あるいは大きな産業を抱える自治体でばらつきが大きくなっている。また、自治体によってサンプル期間がかなり異なる。

一方、自治体間のジニ係数についてもばらつきがある。こちらは全体的な傾向として右下がりの傾向を示す自治体が多く、2010年代後半に給与収入格差は緩やかに縮小傾向であったことを示唆している。この点は他のデータセットでも確認されており、[Kitao and Yamada \(2024a\)](#) は家計調査に基づいて2010年代後半にわずかながら労働所得格差の縮小傾向を報告している。コロナ禍が給与収入格差に与えた影響は判断が難しく、明確な影響を与えていないようにみえる。ただし、同時期に消費格差がそれほど低下していないことも報告されており、この時期に家計間の経済格差が

どのようなになっているかについてはさらなる研究が必要である。また、平均同様にジニ係数で測っても、10 万人未満の自治体の間では動きにばらつきがある一方で、人口規模が 30 万人以上の自治体に対しては似たような動きが観察された。これはサンプル数の違いがあることに加えて、小規模自治体のほうがそれぞれの自治体の特徴に応じた影響が反映されやすい一方、多くの人が集まる自治体は自治体間で似たような動きになるのではないかと考えられる。

6 ライフサイクル側面

最後に、給与収入のライフサイクル側面について、分析をしていこう。図 8 はすべての自治体のデータを全期間プールしたうえで、各年齢ごと平均給与収入、給与収入の対数分散及びジニ係数を計算したものである。図 8(a) が示すように、平均給与収入は年齢とともに増加していき、50 歳代前半をピークとして、緩やかに低下する。この山型の平均プロファイルは賃金構造基本統計調査など、他のデータを用いても観察される特徴である。一方、図 8(b) と図 8(c) が示すように、給与収入のばらつきも年齢とともに増加していく。例えば、対数分散で測った場合、25 歳時の対数分散はおよそ 0.1 であったのが、定年退職直前には 0.35 程度まで増加する。25 歳時の給与収入の格差は先行研究と比較して低めであるが、これは最低賃金で低収入層を落としている事が理由の一つとして考えられる。例えば、最低賃金で落とさずに全ての個人を用いて年齢ごとの給与収入のばらつきを計算すると、若年層の水準が高くなり、J カーブのような形状となる。年齢によって労働所得の格差が拡大する傾向は、[Ohtake and Saito \(1998\)](#) による研究以降、よく知られた特徴である。

一般的に年齢によって給与収入は増加するが、その伸び方は必ずしも一様ではない。図 9 は、自治体の規模ごとに給与収入の平均プロファイルをプロットしたものである。給与収入の伸びは自治体によってばらつきがあるが、これは各自治体の産業などの特徴を反映しているものと考えられる。また、大規模自治体にフォーカスした図 9(c) によると、25 歳時の給与収入は自治体間でばらつきが非常に小さい。ここでも大規模自治体のほうが自治体間の動きが似通っているように見える。大都市近郊は様々な大企業の本社や大規模支店などが存在していて、労働市場が競争的な結果として給与水準が横並びとなり、ばらつきが小さくなっている可能性が考えられるが、この点についてはさらなる分析が必要である。

図 10 は、同じく自治体の規模ごとに給与収入のジニ係数のライフサイクルプロファイルのプロットしたものである。こちらも年齢とともに増加するパターンは変わっておらず、いずれの自治体でも年齢とともに収入格差は拡大する。増え方はほぼ線形に増えていっているが、大規模自治体ではやや凹関数の傾向がみてとれる。加齢とともに給与収入の格差の伸びが徐々に低下する現象

は、米国経済のデータを用いた [Storesletten et al. \(2004a\)](#) でも指摘されている。

これまでの図は何もコントロールしていない給与収入の平均及び対数分散であった。前述の [Storesletten et al. \(2004a\)](#) を含む多くの先行研究では、労働所得を観察可能な変数でコントロールすることによって、説明できない部分の動きに注目をしている。そこで、年齢とともにどの程度平均給与収入が増えるのか及び残差がどのような性質を持っているのかについて、[Aguiar and Hurst \(2013\)](#) に基づいた推定を行った。推定式は

$$\ln Y_{it} = \beta_0 + \beta_{\text{age}} D_{it}^{\text{age}} + \beta_c D_{it}^{\text{cohort}} + \beta_{\text{year}} D^{\text{year}} + \beta_{\text{city}} D^{\text{city}} + \epsilon_{it}$$

とする。ただし、 $\ln Y_{it}$ は給与所得の対数、 D_{it}^{age} 、 D_{it}^{cohort} 、 $\beta_{\text{year}} D^{\text{year}}$ 、 D^{city} はそれぞれ、年齢、コーホート、年および居住地ダミーである。

図 11(a) は加齢効果をあらわす β_{age} をプロットしたものである。平均給与収入は 25 歳から年齢とともに増え、54 歳時点で約 60% 上昇する。一方、図 11(b) は残差 ϵ_{it} の対数分散をプロットしたものである。残差は観察された変数でコントロール後に残った所得のばらつきであり、固有リスク (idiosyncratic risk) であると考えられる。残差の対数分散も年齢とともに増えていくことから、固有リスクには持続的ショック (persistent shock) 部分が存在する事が示唆される¹⁸。

7 追加分析

これまで勤労期にあたる男性の給与収入にフォーカスしてきた。最後にサンプル C に該当しない人たちも含めたサンプルの特徴について確認する。

7.1 給与収入以外の所得

図 12 は、給与収入がゼロの人たちがどのような収入源を持っているかをプロットしたものである。図 12 (a) が示すように、給与収入がゼロである人たちにとって、所得の主な源泉は営業等所得である。ただし、その割合は高い時期でも 20% 程度であり、給与収入がゼロの人たちは給与収入以外もすべてゼロである可能性が高い。営業等所得以外がプラスの値をとっている割合は 5% にも満たない。一方、図 12 (b) は、それぞれの所得の値が正值であるサンプルに限定した平均値である。営業等所得がおよそ 200 万円台で、農業所得を得ている人たちも平均的には 200 万円から 300 万円程度の所得を稼いでいる。ただし、農業所得を含めて、営業等所得以外が正值の割合自体

¹⁸ 推計方法の詳細は、[Storesletten et al. \(2004a\)](#) や [Blundell et al. \(2008\)](#)などを参照。

が極めて低い点に注意する必要がある。我々が利用可能な自治体は市及び町であり小規模な村が含まれていないことから、農業所得が多くなりそうないわゆる農村は含まれていないことも、結果に影響を与えている可能性がある。

7.2 ゼロ所得リスク

図 13 は、次年に給与収入がゼロになる確率を年齢階層及び給与収入階層（10 段階）ごとにまとめたものである。コロナ禍の影響を確認するために、2018 年から 2019 年にかけての 2 年間（左列）と、2019 年から 2020 年にかけての 2 年間（右列）を用いて計算し、男性と女性にわけている。給与収入水準がもっとも低い層（0-10%）に属する若者の場合、次年度におよそ 4 分の 1 から 5 分の 1 が給与収入がゼロになる。年齢とともに次年に給与収入がゼロになる確率は下がっていき、30 歳代以降はかなりフラットになる。また、全体的に、コロナの影響から 2020 年に給与収入がゼロになった人たちは、ごくわずかながら増えているように見える。

男性と比べて、女性は年齢が上昇してもなお、次年に給与収入がゼロになる確率が男性よりも高い。特に、下位 10% に入る女性は次年にゼロになる確率が 15% を超えて高いままである。ただし、我々のデータからはこれが失職の影響なのか、自発的に職を辞したり何らかの所得調整をした結果なのかは判断できない点に注意してほしい。いずれにしても、女性労働者は職に留まる可能性が男性よりも低い傾向が観察された。

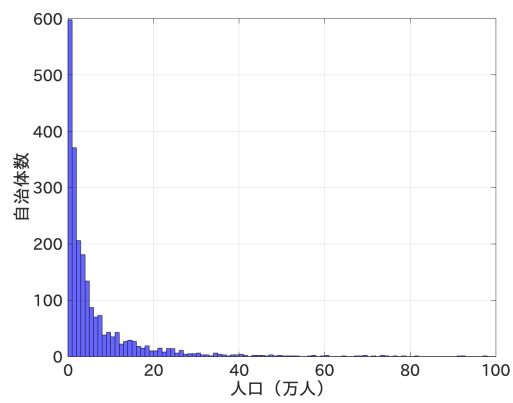
7.3 労働参加率

最後に我々のデータセットでどの程度の人が働いているのかを確認する。図 14 は、年齢ごとに全サンプルに占める給与収入がプラスの値の人たちの割合である。労働参加率の算出には全サンプルを用いて、全サンプルに占める給与収入が (a) ゼロ以上の割合および (b) 100 万円以上の割合をそれぞれプロットした。男性については、賃金構造基本統計調査などでもみられるように、勤労期には全体的に高い水準にある。一方、女性については、昔ながらのいわゆる M 字型の労働参加率の影響が残っているように見える。特に、給与収入が 100 万円以上の人たちに限定をすると、30 歳代以降の女性の労働参加率は明確に下がる。ただし、前節でも言及したように、これが解雇などに伴う失職なのか、自発的な労働調整の結果なのかは業務データからでは判断がつかないため、その解釈には注意する必要がある。

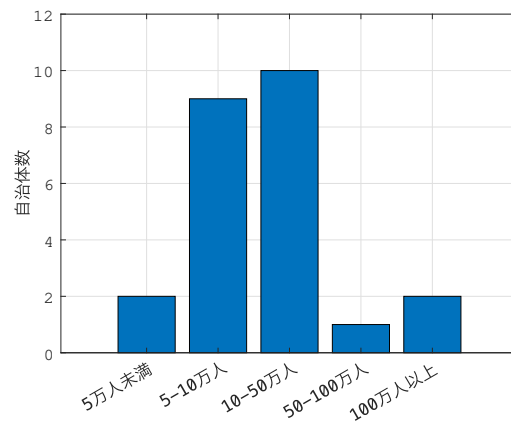
8 まとめ

本研究は、日本の地方自治体が保有する業務データ（24 自治体、約 550 万人）を用いて、2011 年から 2023 年にかけての所得格差の実態とその推移を分析した。従来のサーベイデータでは把握が困難であった高・低所得層を含む大規模データを活用することで、給与収入の格差、時系列での変化、ライフサイクルにおける格差の特徴を詳細に把握することが可能となった。主要な発見として、コロナ禍の一時的な所得低下を挟みつつも、多くの自治体で格差縮小傾向が見られること、年齢とともに所得水準・ばらつきが増加すること、自治体間で格差指標に大きなばらつきがあることが確認された。

業務データを用いた本研究は、日本における格差研究において行政データの活用可能性を示すとともに、従来のサーベイデータの限界を補完する重要な意義を持つ。一方で、データを提供した自治体が限られているため日本全体の代表性には課題があり、サンプル期間の制約から長期的な経済成長や景気循環との関係性の分析は困難である。今後は、より多くの自治体からのデータ協力を得ることで代表性を高めるとともに、国税データとの連携や長期間のパネルデータ構築により、日本の所得格差に関するより包括的な理解を深めることが求められる。



(a) 日本の全自治体



(b) 参加自治体

図 1: 日本および参加自治体の規模

注：市区町村の人口はいずれも令和 6 年 1 月 1 日時点での住民基本台帳人口に基づいている。

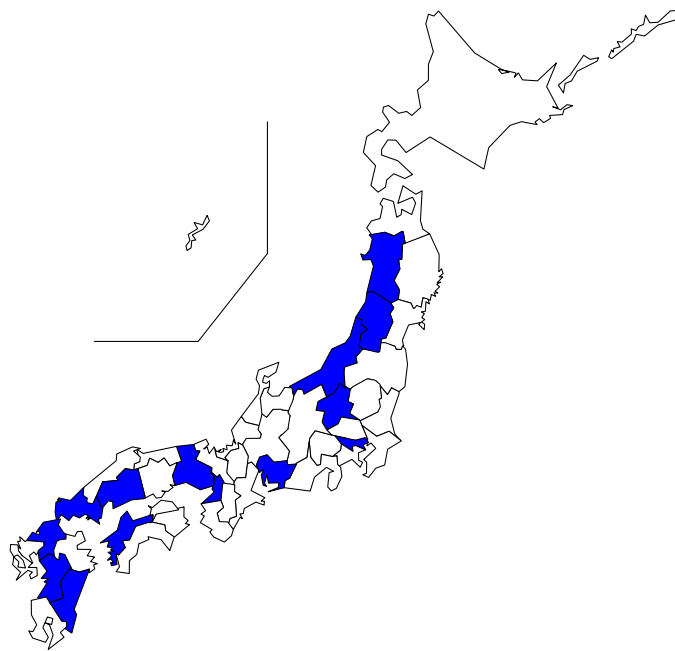
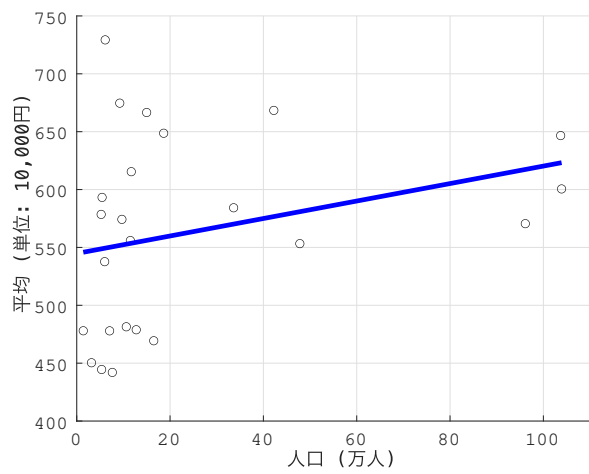
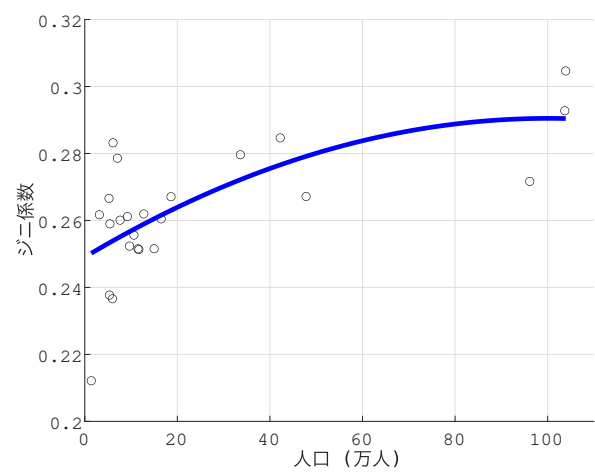


図 2: データが提供されている自治体が属する都道府県



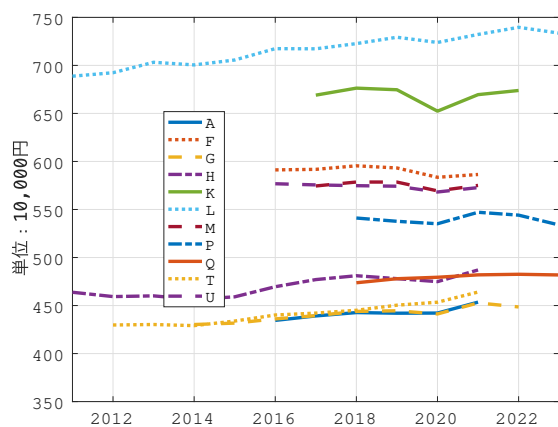
(a) 平均給与収入と人口サイズ



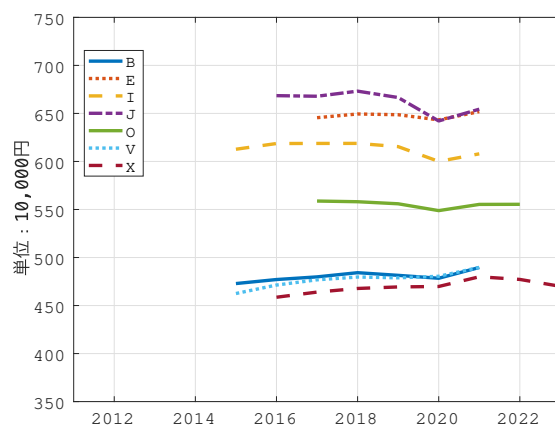
(b) 給与収入のジニ係数と人口サイズ

図 3: 市町村の人口サイズと平均給与収入、格差の関係

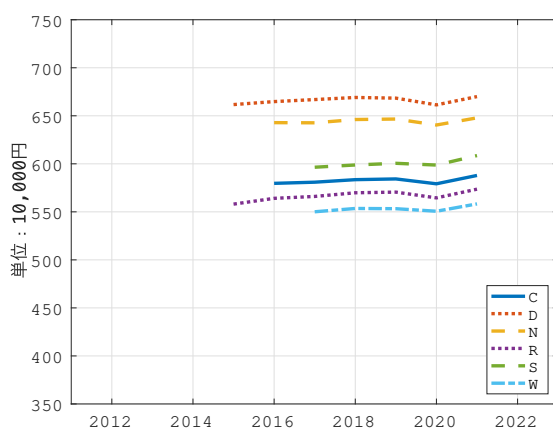
注：サンプル C を使用。



(a) 自治体サイズ：10 万人未満



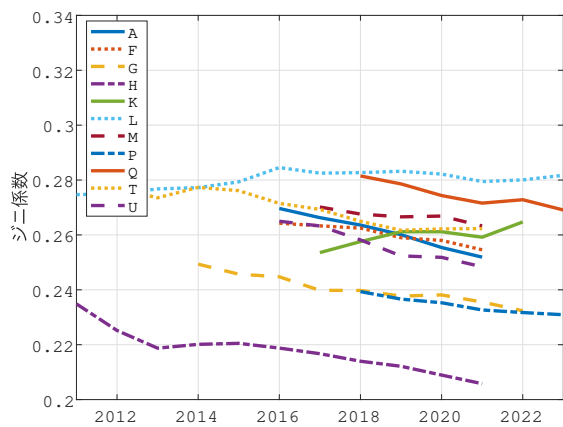
(b) 自治体サイズ：10 万以上 30 万人未満



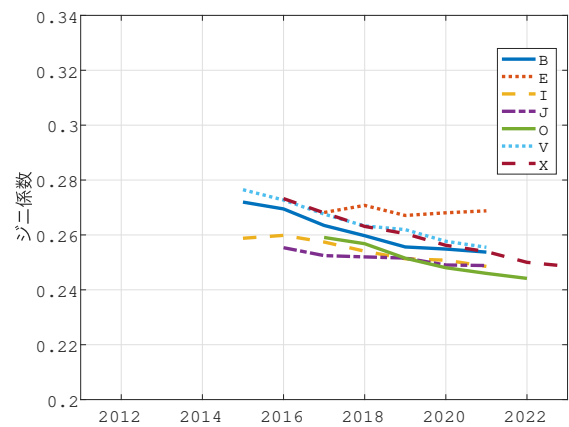
(c) 自治体サイズ：30 万人以上

図 4: 自治体ごとの給与収入の平均値

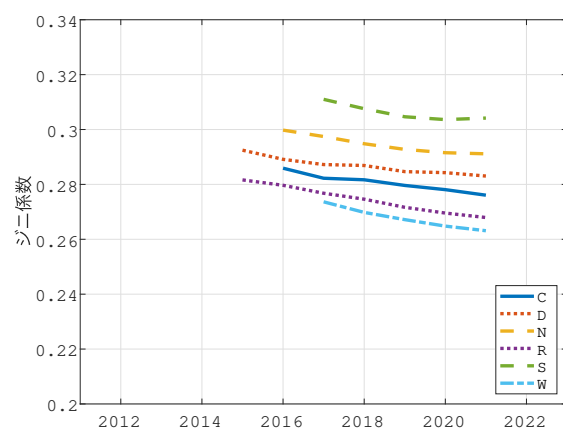
注：サンプル C を使用。



(a) 自治体サイズ：10 万人未満



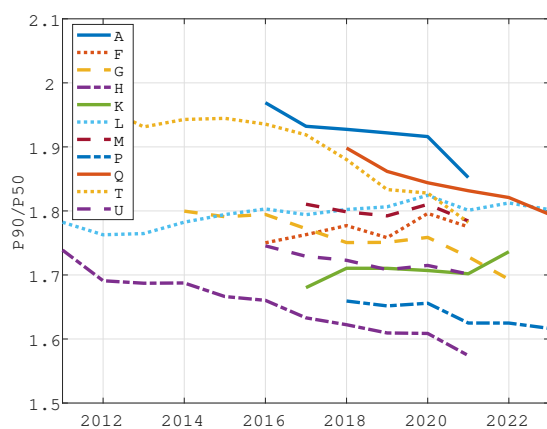
(b) 自治体サイズ：10 万以上 30 万人未満



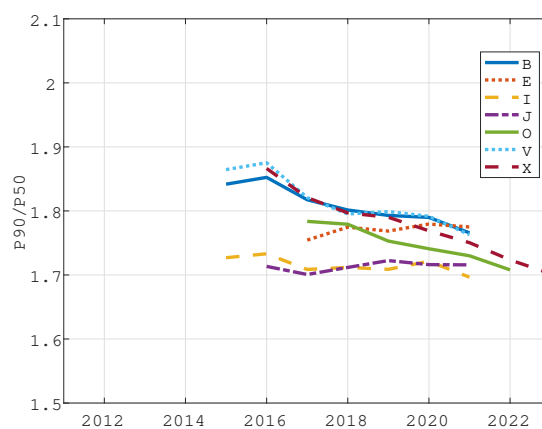
(c) 自治体サイズ：30 万人以上

図 5: 自治体ごとの給与収入のジニ係数

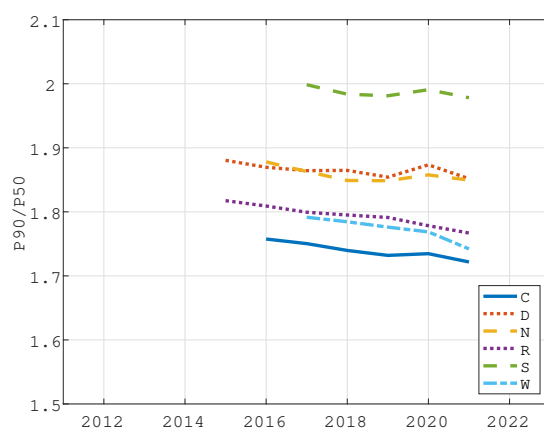
注：サンプル C を使用。



(a) 自治体サイズ：10 万人未満



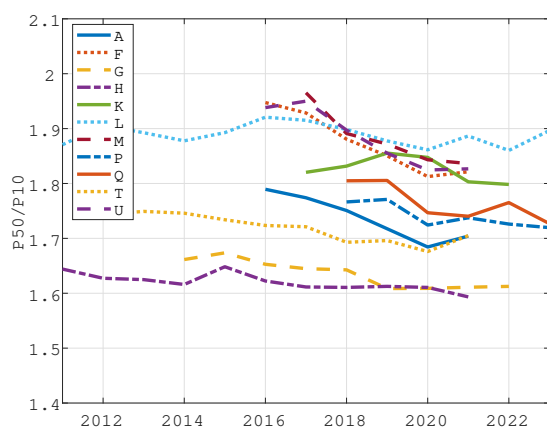
(b) 自治体サイズ：10 万以上 30 万人未満



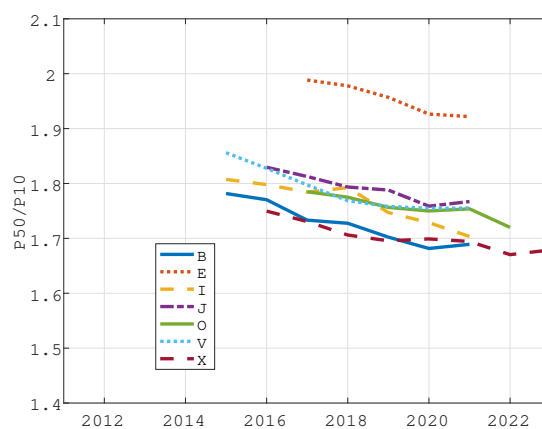
(c) 自治体サイズ：30 万人以上

図 6: 自治体ごとの給与収入の P90/P50 比

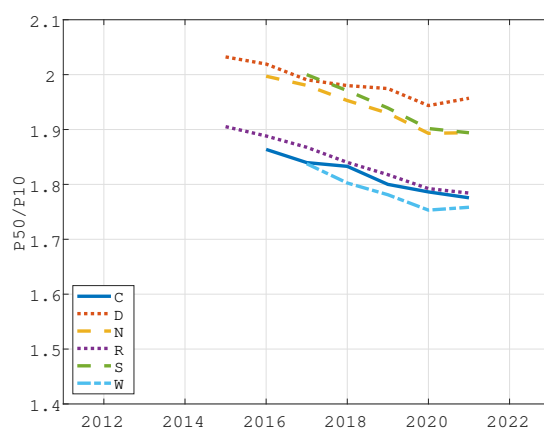
注：サンプル C を使用。



(a) 自治体サイズ：10 万人未満



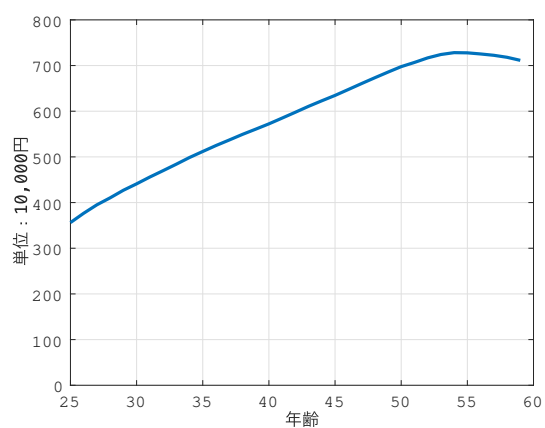
(b) 自治体サイズ：10 万以上 30 万人未満



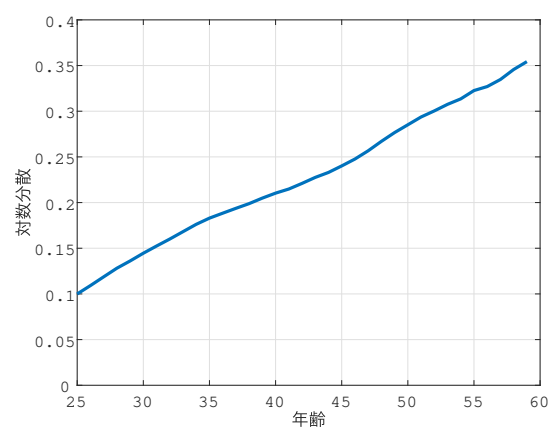
(c) 自治体サイズ：30 万人以上

図 7: 自治体ごとの給与収入の P50/P10 比

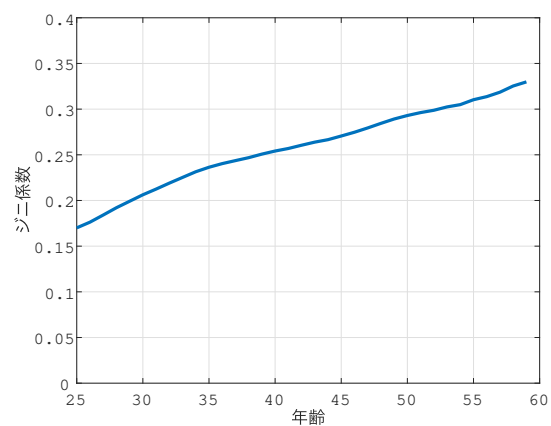
注：サンプル C を使用。



(a) 平均給与収入



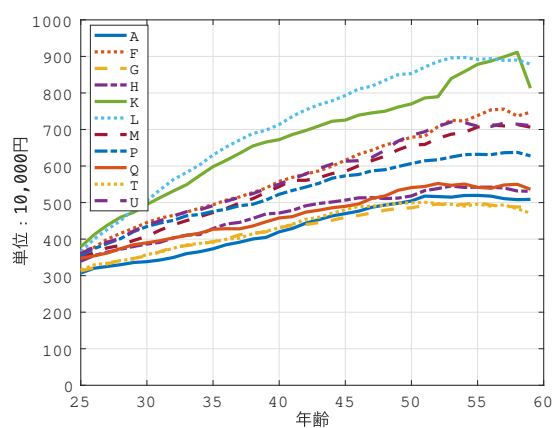
(b) 給与収入の対数分散



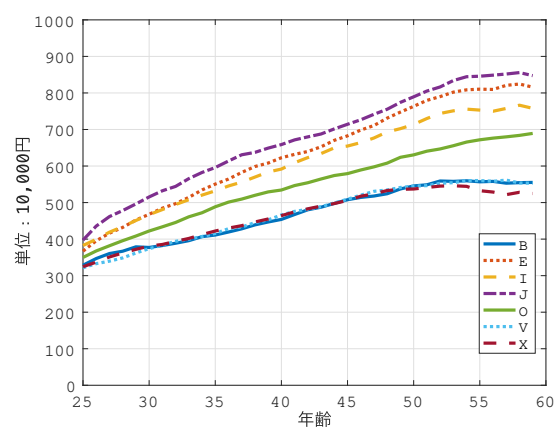
(c) 給与収入のジニ係数

図 8: ライフサイクル側面からみた自治体ごとの給与収入

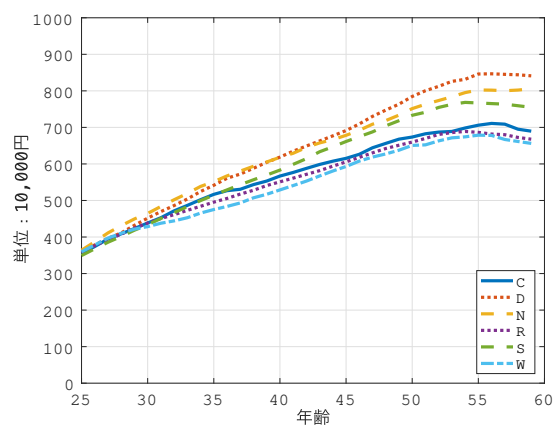
注: 全ての年の全ての自治体のデータをプールしたサンプル C を使用。



(a) 自治体サイズ：10 万人未満



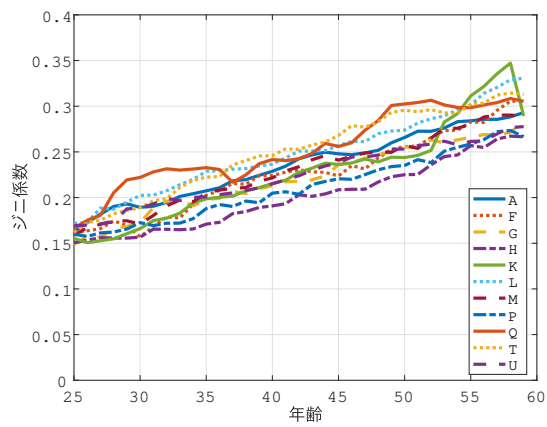
(b) 自治体サイズ：10 万以上 30 万人未満



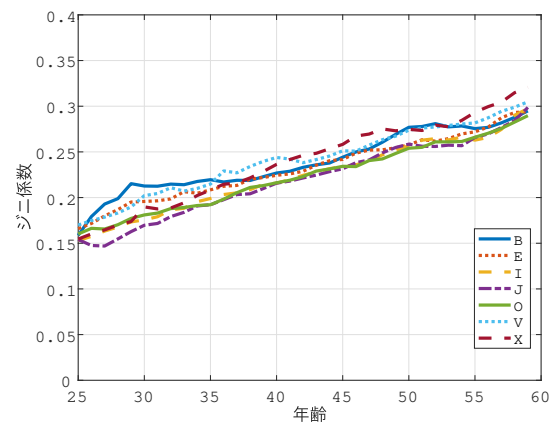
(c) 自治体サイズ：30 万人以上

図 9: ライフサイクル側面からみた自治体ごとの平均給与収入

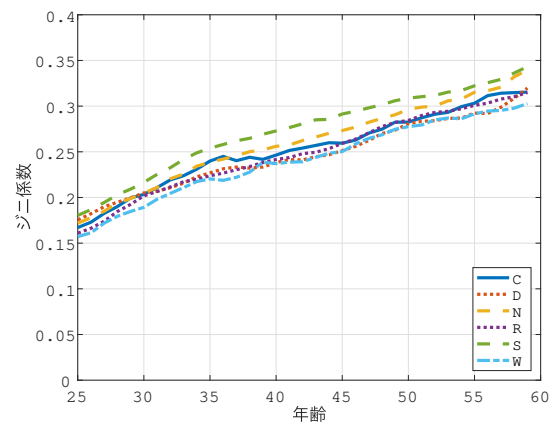
注：それぞれの自治体について、利用可能な全ての年のデータをプールしたサンプル C を使用。



(a) 自治体サイズ：10 万人未満



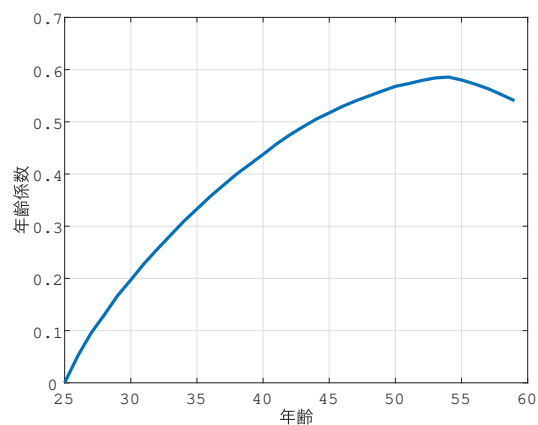
(b) 自治体サイズ：10 万以上 30 万人未満



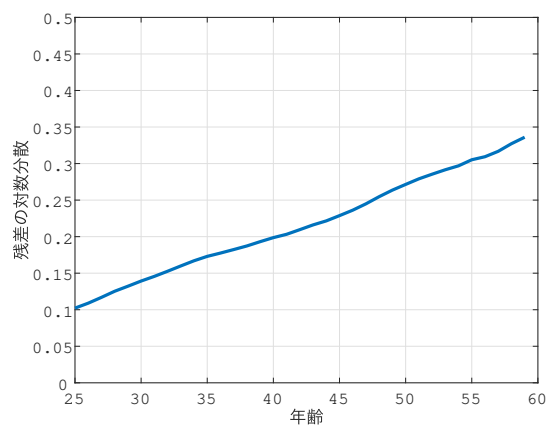
(c) 自治体サイズ：30 万人以上

図 10: ライフサイクル側面からみた自治体ごとの給与収入のジニ係数

注：それぞれの自治体について、利用可能な全ての年のデータをプールしたサンプル C を使用。

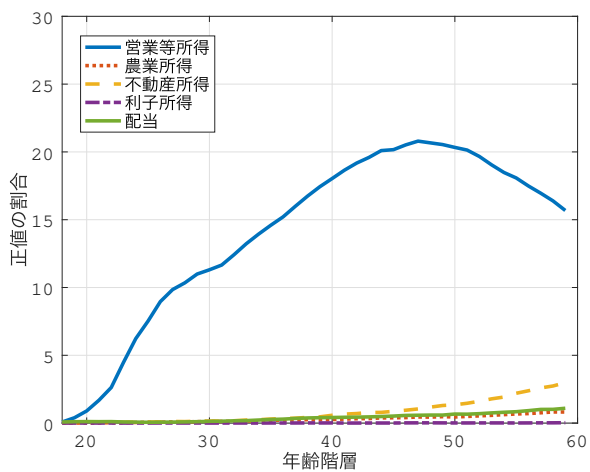


(a) 加齢効果 (β_{age})

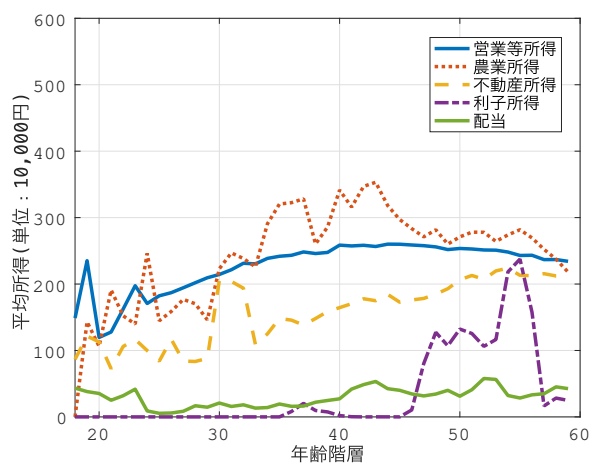


(b) 残差 ϵ_{it} の対数分散

図 11: 年効果とコーホート効果をコントロールした後の給与所得のライフサイクル特性



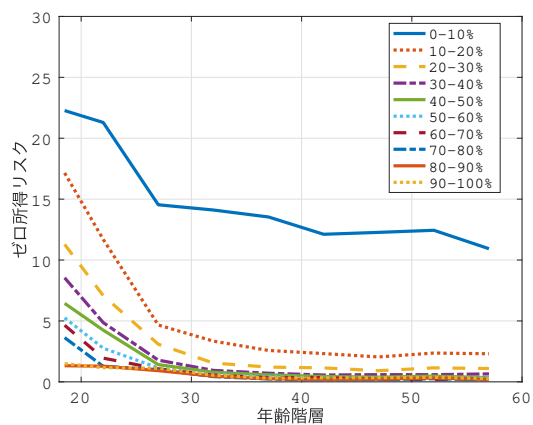
(a) 所得の内訳



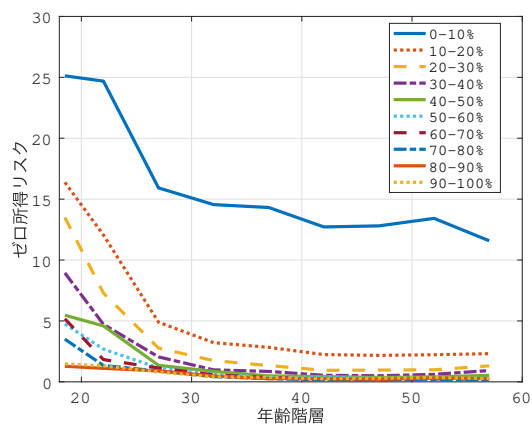
(b) 給与収入以外の所得の平均値

図 12: 給与収入以外の所得の内訳

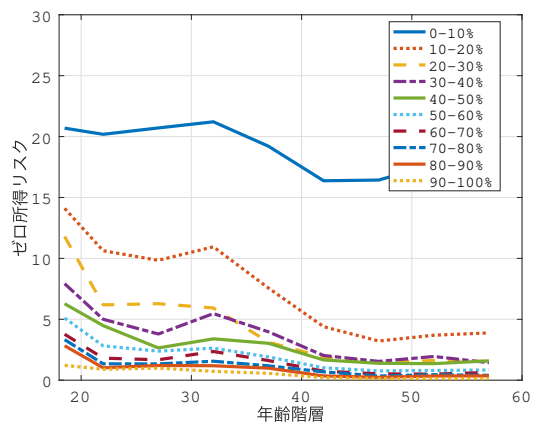
注：18 歳以上 60 歳以下の全サンプルを使用。ただし、給与収入がゼロの個人に限定。



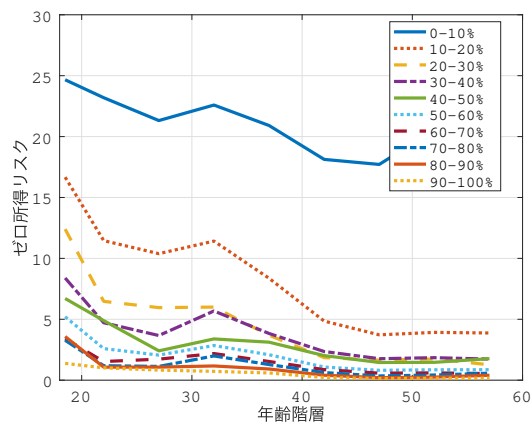
(a) 男性労働者：2018 年-2019 年



(b) 男性労働者：2019 年-2020 年



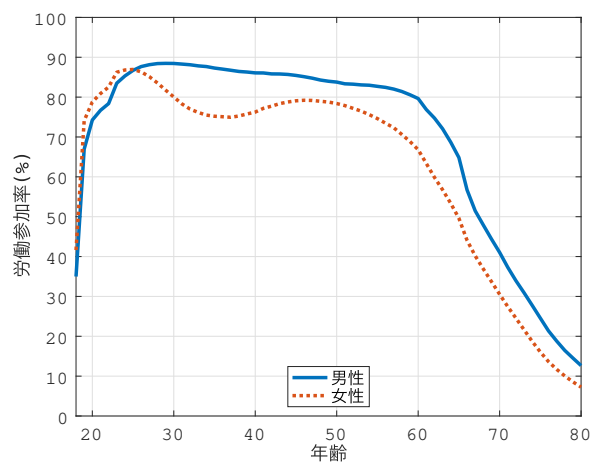
(c) 女性労働者：2018 年-2019 年



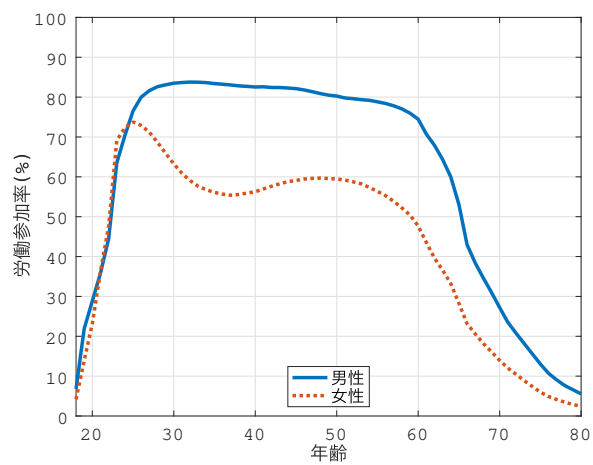
(d) 女性労働者：2019 年-2020 年

図 13: 次年に給与収入がゼロになる確率

注：それぞれの年のすべての自治体のサンプルを使用。



(a) 給与収入が 0 以上



(b) 給与収入が 100 万円以上

図 14: 年齢ごとの労働参加率

注：18 歳以上 80 歳以下の全サンプルを使用。

年	全サンプル	サンプル A	サンプル B	サンプル C	自治体数
2011	64,718	32,281	16,480	13,902	2
2012	103,304	47,791	24,180	19,358	3
2013	104,043	48,012	24,306	19,589	3
2014	162,258	69,856	35,656	28,872	4
2015	1,718,684	732,090	366,899	284,847	9
2016	3,179,636	1,374,008	687,252	540,852	16
2017	5,412,346	2,430,377	1,207,237	952,778	22
2018	5,553,051	2,472,542	1,228,693	972,042	24
2019	5,552,572	2,466,232	1,226,170	970,553	24
2020	5,536,387	2,454,946	1,220,929	962,723	24
2021	5,517,588	2,445,732	1,216,745	960,577	24
2022	612,647	251,020	127,415	104,160	7
2023	351,994	140,991	70,910	56,406	3

表 1: サンプルサイズの推移とサンプルセレクション

自治体コード	A-X	A	B	C	D	E	F	G	H
トップ 1%	1,910	1,200	1,440	2,160	1,990	1,800	1,610	1,190	1,180
トップ 0.1%	4,700	2,480	3,030	6,080	4,960	3,780	4,320	2,880	2,050
トップ 0.01%	6,080	3,180	7,840	6,080	4,960	3,780	4,320	6,740	2,520
自治体コード	I	J	K	L	M	N	O	P	Q
トップ 1%	1,700	1,740	1,710	2,330	1,710	2,150	1,490	1,340	1,670
トップ 0.1%	4,300	4,150	4,370	6,180	3,140	5,500	2,960	2,580	3,270
トップ 0.01%	4,300	10,500	11,900	12,300	5,090	5,500	7,900	6,370	7,720
自治体コード	R	S	T	U	V	W	X		
トップ 1%	1,820	2,080	1,440	1,500	1,440	1,770	1,400		
トップ 0.1%	4,700	4,870	3,280	3,140	3,310	3,340	4,060		
トップ 0.01%	4,700	4,870	7,190	3,140	8,990	6,700	9,790		

表 2: 給与収入の上位 1%、0.1%、0.01% の基準

注：サンプル C を使用。2019 年の値を使用。A-X は 2019 年の全ての自治体をプールして、閾値を計算したもの。

自治体コード	平均 (万円)	ジニ係数	p99/50	p90/p50	p50/p1	p50/p10
A	442.0	0.2601	3.1135	1.9219	2.3615	1.7175
B	481.4	0.2556	3.3898	1.7930	2.5240	1.7023
C	584.3	0.2796	4.2723	1.7320	2.9491	1.8000
D	668.4	0.2847	3.3685	1.8541	3.3767	1.9746
E	648.7	0.2671	3.0689	1.7686	3.3466	1.9569
F	593.2	0.2590	3.0063	1.7582	3.0619	1.8504
G	444.5	0.2377	2.9859	1.7509	2.3666	1.6086
H	478.0	0.2121	2.6511	1.6094	2.5057	1.6126
I	615.4	0.2513	3.0656	1.7088	3.0863	1.7471
J	666.6	0.2516	2.8926	1.7225	3.1916	1.7880
K	674.7	0.2612	2.8047	1.7104	3.2595	1.8553
L	729.3	0.2832	3.6257	1.8064	3.5097	1.8776
M	578.5	0.2666	3.3162	1.7923	3.0363	1.8718
N	646.6	0.2928	3.8458	1.8486	3.2571	1.9298
O	555.9	0.2515	2.9918	1.7529	2.8662	1.7564
P	537.7	0.2366	2.7081	1.6517	2.8532	1.7710
Q	477.9	0.2786	4.0580	1.8617	2.5152	1.8055
R	570.5	0.2717	3.6251	1.7912	2.9448	1.8178
S	600.5	0.3046	4.1067	1.9813	3.0180	1.9392
T	450.4	0.2617	3.6614	1.8333	2.4235	1.6960
U	574.2	0.2524	2.8472	1.7083	3.1239	1.8552
V	478.9	0.2619	3.4128	1.7990	2.5266	1.7581
W	553.3	0.2672	3.6509	1.7761	2.8431	1.7812
X	469.4	0.2605	3.4408	1.7901	2.4468	1.6955
A-X	589.6	0.2852	3.7126	1.8728	2.8888	1.8576

表 3: 自治体ごとの給与収入 (2019 年)

注：サンプル C を使用。最終行の A-X は 2019 年の全ての自治体をプールして、平均及び各種の経済格差指標を計算したもの。

年	サンプル A	サンプル A		サンプル C
		男性	女性	
2017	334.8	472.3	199.1	592.7
2018	338.6	474.8	204.2	594.2
2019	341.5	475.6	208.8	594.8
2020	340.3	470.1	211.8	589.6
2021	348.1	477.9	219.7	598.9

表 4: 平均給与収入

注：真ん中の 2 行はサンプル A から男性のみと女性のみでそれぞれ平均給与収入を計算したもの。サンプル A の男性は、定義よりサンプル B に一致する。単位は万円。

年	サンプル A	サンプル A		サンプル C
		男性	女性	
2017	0.5275	0.4309	0.5659	0.2919
2018	0.5211	0.4273	0.5568	0.2893
2019	0.5155	0.4248	0.5487	0.2869
2020	0.5116	0.4246	0.5439	0.2852
2021	0.5076	0.4237	0.5377	0.2843

表 5: 給与収入のジニ係数

注：真ん中の 2 行はサンプル A から男性のみと女性のみでそれぞれ給与収入のジニ係数を計算したもの。サンプル A の男性は、定義よりサンプル B に一致する。

年	サンプル A	サンプル A		サンプル C
		男性	女性	
2017	254.2	376.0	134.0	451.1
2018	258.1	380.1	137.5	454.8
2019	258.9	377.7	141.3	450.9
2020	266.4	382.7	151.3	457.5
2021	274.7	393.0	157.6	467.4

表 6: 平均合計所得金額

注：真ん中の 2 行はサンプル A から男性のみと女性のみでそれぞれ平均合計所得金額を計算したもの。サンプル A の男性は、定義よりサンプル B に一致する。単位は万円。

年	サンプル A	サンプル A		サンプル C
		男性	女性	
2017	0.5765	0.4667	0.6364	0.3590
2018	0.5715	0.4657	0.6262	0.3599
2019	0.5618	0.4567	0.6166	0.3509
2020	0.5480	0.4503	0.5945	0.3434
2021	0.5450	0.4516	0.5870	0.3448

表 7: 合計所得金額のジニ係数

注：真ん中の 2 行はサンプル A から男性のみと女性のみでそれぞれ合計所得金額のジニ係数を計算したもの。サンプル A の男性は、定義よりサンプル B に一致する。

参考文献

- Aguiar, Mark A. and Erik Hurst (2013) “Deconstructing Lifecycle Expenditure,” *Journal of Political Economy*, Vol. 121, No. 3, pp. 437-492.
- Arata, Yoshiyuki, Daisuke Miyakawa, and Katsuki Mori (2023) “The U-shape Law for High Growth Firms,” *NTC Joint Research Discussion Paper Series, 230200-01HJ*.
- Arellano, Manuel, Stéphane Bonhomme, Micole De Vera, Laura Hospido, and Siqi Wei (2022) “Income Risk Inequality: Evidence from Spanish Administrative Records,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1747-1801.
- Bell, Brian, Nicholas Bloom, and Jack Blundell (2022) “Income Dynamics in the UK and the Impact of the Covid-19 Recession,” *Quantitative Economics*, Vol. 13, pp. 1849–1878.
- Blanco, Andrés, Bernardo Diaz de Astarloa, Andres Drenik, Christian Moser, and Danilo R. Trupkin (2022) “The Evolution of the Earnings Distribution in a Volatile Economy: Evidence from Argentina,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1361-1403.
- Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008) “Consumption Inequality and Partial Insurance,” *American Economic Review*, Vol. 98, No. 5, pp. 1887-1921.
- Bowlus, Audra, Emilien Gouin-Bonenfant, Huju Liu, Lance Lochner, and Youngmin Park (2022) “Four Decades of Canadian Earnings Inequality and Dynamics across Workers and Firms,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1447-1491.
- Busch, Christopher, David Domeij, Fatih Guvenen, and Rocio Madera (2022) “Skewed Idiosyncratic Income Risk over the Business Cycle: Sources and Insurance,” *American Economic Journal: Macroeconomics*, Vol. 14, No. 2, pp. 207–242.
- Drechsel-Grau, Moritz, Andreas Peichl, Kai D. Schmid, Johannes F. Schmieder, Hannes Walz, and Stefanie Wolter (2022) “Inequality and Income Dynamics in Germany,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1593-1635.
- Engbom, Niklas, Gustavo Gonzaga, Christian Moser, and Roberta Olivieri (2022) “Earnings Inequality and Dynamics in the Presence of Informality: The Case of Brazil,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1405-1446.
- Friedrich, Benjamin, Lisa Laun, and Costas Meghir (2022) “Earnings Dynamics of Immigrants and Natives in Sweden 1985–2016,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1803-1847.

- Fukai, Taiyo and Ayako Kondo (2025) “Parental Earnings Trajectories around Childbirth in Japan: Evidence from local tax records,” RIETI Discussion Papers Series, No. 25012.
- Fukuda, Rui (2025) “Administrative Tax Data of Japanese Municipalities: A Review,” Work in progress.
- Gunji, Hiroshi, Kazuki Hiraga, and Kenji Miyazaki (2025) “Analysis of Labor Income and Labor Income Taxes using Final Tax Return Data and Applications of a Heterogeneous-Agent Dynamic General Equilibrium Model,” mimeo.
- Guvenen, Fatih, Serdar Ozkan, and Jae Song (2014a) “How Risky Are Recessions for Top Earners?” *American Economic Review: Papers & Proceedings*, Vol. 104, No. 5, pp. 148-153.
- (2014b) “The Nature of Countercyclical Income Risk,” *Journal of Political Economy*, Vol. 122, No. 3, pp. 621-660.
- Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song (2021) “What Do Data on Millions of U.S. Workers Reveal About Life-Cycle Earnings?” *Econometrica*, Vol. 89, No. 5, pp. 2303-2339.
- Guvenen, Fatih, Luigi Pistaferri, and Giovanni L. Violante (2022a) “Global Trends in Income Inequality and Income Dynamics: New insights from GRID,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1321-1360.
- Guvenen, Fatih, Greg Kaplan, Jae Song, and Justin Weidner (2022b) “Lifetime Incomes in the United States Over Six Decades,” *American Economic Journal. Applied Economics*, Vol. 14, No. 4, pp. 446-479.
- Guvenen, Fatih and Greg Kaplan (2017) “Top Income Inequality in the 21st Century: Some Cautionary Notes,” *Quarterly Review - Federal Reserve Bank of Minneapolis*, Vol. 38, No. 1, pp. 2-15.
- Halvorsen, Elin, Serdar Ozkan, and Sergio Salgado (2022) “Earnings Dynamics and its Inter-generational Transmission: Evidence from Norway,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1707-1746.
- Hoffmann, Eran B., Davide Malacrino, and Luigi Pistaferri (2022) “Earnings Dynamics and Labor Market Reforms: The Italian Case,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1637-1667.

- Kawaguchi, Daiji and Takahiro Toriyabe (2025) “Nonresponse Bias of Japanese Wage Statistics,” *Journal of Official Statistics*, Vol. 40, No. 1, pp. 57-122.
- Kitao, Sagiri and Tomoaki Yamada (2024a) “Inequality Dynamics in Japan, 1981-2021,” *ESRI Discussion Paper Series*, Vol. No. 392.
- (2024b) “The time trend and life-cycle profiles of consumption,” *Review of Economics of the Household*, <https://doi.org/10.1007/s11150-024-09710-z>.
- (2025) “Earnings, Income, and Wealth Inequality in Japan: A Long-Term Perspective, 1984–2019,” *Japanese Economic Review*, Vol. 75, pp. 231-283.
- Kramarz, Francis, Elio Nimier-David, and Thomas Delemotte (2022) “Inequality and Earnings Dynamics in France: National Policies and Local Consequences,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1527-1591.
- Leth-Petersen, Søren and Johan Sæverud (2022) “Inequality and Dynamics of Earnings and Disposable Income in Denmark 1987–2016,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1493-1526.
- Lise, Jeremy, Nao Sudo, Michio Suzuki, Ken Yamada, and Tomoaki Yamada (2014) “Wage, Income and Consumption Inequality in Japan, 1981-2008: from Boom to Lost Decades,” *Review of Economic Dynamics*, Vol. 17, pp. 582-612.
- McKinney, Kevin L., John M. Abowd, and Hubert P. Janicki (2022) “U.S. Long-term Earnings Outcomes by Sex, Race, Ethnicity, and Place of Birth,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1879-1945.
- Ohtake, Fumio (2005) *Inequality in Japan*, Tokyo: Nihon Keizai Shinbunsha, (in Japanese).
- Ohtake, Fumio and Makoto Saito (1998) “Population Aging and Consumption Inequality in Japan,” *Review of Income and Wealth*, Vol. 44, pp. 361-381.
- Puggioni, Daniela, Mariana Calderón, Alfonso Ceballos Zurita, León Fernández Bujanda, José Antonio Inguanzo González, and David Jaume (2022) “Inequality, Income Dynamics, and Worker Transitions: The Case of Mexico,” *Quantitative Economics*, Vol. 13, No. 4, pp. 1669-1705.
- Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter (2019) “Firming Up Inequality,” *Quarterly Journal of Economics*, Vol. 134, No. 1, pp. 1-50.
- Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron (2004a) “Consumption and Risk Sharing

- Over the Life Cycle,” *Journal of Monetary Economics*, Vol. 51, pp. 609-633.
- (2004b) “Cyclical Dynamics in Idiosyncratic Labor-Market Risk,” *Journal of Political Economy*, Vol. 112, No. 3, pp. 695-717.
- Sudo, Nao, Michio Suzuki, and Tomoaki Yamada (2012) “Inequalities in Japanese Economy during the Lost Decades,” CIRJE Discussion Paper Series, CIRJE-F-856.
- Unayama, Takashi (2023) *Consumption Analysis of Contemporary Japan*: Keio University Press.
- 宇南山卓・佐野晋平・勇上和史・稲葉和洋 (2025) 「所得税データとしての確定申告データ」, 『ファイナンシャル・レビュー』, 第 160 号, 150-181 頁.
- 大野太郎・岡本慎吾・稲葉和洋 (2025) 「所得税における控除の負担軽減効果と再分配効果：税務データを用いた分析」, 『ファイナンシャル・レビュー』, 第 160 号, 124-149 頁.
- 北村行伸 (2025) 「序文：日本における税務データを用いた近年の研究動向と今後の展望」, 『ファイナンシャル・レビュー』, 第 160 号, 92-95 頁.
- 國枝繁樹 (2025) 「我が国における税務データに基づく分析の意義」, *NTC Joint Research Discussion Paper Series, 230100-01ST*.
- 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」, *NTC Joint Research Discussion Paper Series, 230100-01ST*.
- 近藤絢子・深井太洋 (2023) 「市町村税務データを用いた既婚女性の就労調整の分析」, *RIETI Discussion Paper Series 23-J-049*.
- 土居丈朗・別所俊一郎・森克輝 (2025) 「法人税申告書の個票データを用いた欠損法人等に関する実態分析」, 『ファイナンシャル・レビュー』, 第 160 号, 182-258 頁.
- 日景智 (2002) 「所得税と個人住民税との関係について－わが国個人所得課税のメカニズム－」, 『税大論叢』, 第 39 号, 431-504 頁.
- 宮川大介・滝澤美帆・古屋達史 (2025) 「法人税申告データを使ったビジネスダイナミズムの測定」, 『ファイナンシャル・レビュー』, 第 160 号, 259-267 頁.