

行政データの学術的利用： 彦根市住民税課税データの 分析事例

井上俊克

Toshikatsu Inoue

滋賀大学 経済学部 / 講師

I はじめに

近年経済学においてデータの多様性が著しく増加している。従来から広く活用されてきた政府統計調査に加え、民間企業が保有する取引データや顧客データ、さらには行政組織が業務を通じて蓄積する各種行政記録データなど、様々なデータが研究に利用されている。このような潮流の中、滋賀大学においても彦根市との連携により、同市が保有する住民税課税データ（以下、彦根市税務データ）の提供を受け、その学術的利用を開始した。本稿では、実際に彦根市税務データを用いた分析例を紹介することで、彦根市税務データの学術研究における活用可能性を広く学内に周知し、今後の利用促進を図る。

まずは、原稿執筆時点における彦根市税務データの概要を紹介する。提供を受けるデータはすべて匿名化処理が施されており、利用者と彦根市との間で締結された秘密保持契約に基づき、厳正な管理体制のもとで利用されている。提供データは、彦根市税務課が住民税徴収業務に用いるデータであり、個人属性（性別・生年など）や収入、所得、税額などの情報が含まれている。個人を識別するIDは匿名化された状態で提供されているが、複数年度にわたる同一個人の追跡が可能であり、パネルデータの構築が可能となっている。また、匿名化された世帯番号も提供されており、世帯の家族構成や配偶者の所得なども分析に利用できる。所得・収入に関するデータは、地方自治体に送付される源泉徴収票および確定申告書に基づいており、年ごとの給与収入・所得、年金所得、事業所得など、収入・所得の発生源ごとに詳細に記録されている。ただし、特定口座における分離課税が適用される金融所得など、一部の所得については本データでは捕捉されていない点に留意が必

要である。税に関しては、所得税や住民税、各種控除額などが記録されているほか、社会保険料控除額から社会保険料の支払額を一定程度把握することも可能となっている¹⁾。

税務データを用いた実証研究は、近年、経済学をはじめとする社会科学分野において注目されている。その理由として次の三点が川口・正木(2022)では挙げられている。第一に、税務データは行政による税の徴収業務の過程で作成されるため、未回答によるバイアスの問題が小さいことが挙げられる。アンケート調査等では回答拒否や無回答の問題が生じやすいのに対し、税務データはほぼ全ての住民に対しデータが収集されている。第二に、所得や収入といった経済変数の測定精度が高いことも大きな利点である。調査票を用いた統計調査においては所得・収入の属する区間の回答を求めることが多いが、税務データでは源泉徴収票や確定申告書に基づく正確な記録が利用可能である。第三に、複数年度にわたるパネルデータの構築ができ、個人の時系列的な変化を追跡でき、統計的因果推論や時系列分析など高度な分析が可能であることも大きな利点である。

税務データの学術利用において、国内で先駆的な取り組みを行っているのが東京大学政策評価研究教育センター(CREPE)である。CREPEでは、およそ20の自治体から税務データの提供を受け、学術研究に活用している。川口・正木(2022)によれば、CREPEが構築したデータベースには、個人住民税の課税情報に加え、所得、年齢、性別、家族構成などの変数が含まれており、複数年度のパネルデータとして整備されている。これらのデータを用いて、既に労働経済学、公共経済学など多岐にわたる分野で実証研究が蓄積されている。彦根市税務データは変数構成の面でCREPEが整備しているデータと部分的に共通しており、すでに

確立された分析枠組みや先行研究の蓄積を参照しながら研究を進めることが可能である。同時に、このことは彦根市税務データが質の高いデータセットではあるものの、データとしての新規性は大きくはないことを意味している。したがって、当然のことではあるが、研究においてはどのような問いを立てるかが重要となる点には注意が必要である。地域特性を活かした分析、他のデータソースとの接続による新たな変数の構築、あるいは独自の政策変化や制度改正に着目した自然実験の設計など、データの特性を理解した上で研究上の付加価値を明確にすることが求められる。

税務データを用いた学術研究のもう一つの重要な取り組みとして、国税庁が実施している確定申告データを用いた共同研究がある。国税庁では共同研究の公募を定期的に行っており、採択された研究者に対して確定申告データを提供している。確定申告データに含まれる情報は、当該自治体の住民のものであれば自治体も把握しており、彦根市税務データにも確定申告を行った個人の情報は含まれている。国税庁が提供する確定申告データの最大の特徴は、全国規模のデータが利用可能である点にある。CREPEや彦根市のような特定の自治体データでは当該地域に居住する住民の情報に限定されるのに対し、国税庁のデータは日本全国の確定申告者を対象としており、サンプルサイズが大きく、全国的な分析や地域間比較が可能となる。他方で、確定申告データには重要な制約も存在する。確定申告データには、年末調整のみで確定申告を行わない給与所得者や、公的年金の源泉徴収のみで納税が完了する年金受給者はデータに含まれない。確定申告を行うのは自営業者、高額所得者、複数の所得源泉を持つ者、医療費控除の適用を受ける者などに限られる点に注意が必要である。一方彦根市税務データは、原則

1) 社会保険料控除額は給与・年金から天引きされた社会保険料と確定申告を行った社会保険料の総額となっているが、納付書などで支払った社会保険料を確定申告しない場合など、必ずしも社会保険料控除額と実際の社会保険料は一致

しない。

として全住民を対象としているため、確定申告をしていない給与所得者や年金所得者、所得のない住民も含めた分析が可能となっている。

彦根市税務データと確定申告データの違いを定量的に明らかにするため、宇南山ほか(2025)が国税庁の確定申告データを用いて示した所得区分の割合を比較する。宇南山ほか(2025)では、所得の収入源をもとに独自の所得タイプ分類を構築しており、同様の手順により彦根市税務データを使い所得タイプを分類した。また、確定申告を行う人で無所得の人は極めて少数であると予想されるが、彦根市税務データには無所得者も多く含まれているため、給与収入やその他の所得が0の場合に分類する無所得者という分類を追加している。

分類した結果、彦根市税務データと国税庁の確定申告データの間には顕著な構成の違いが確認された(図1)。具体的には、彦根市税務データでは給与所得者および年金所得者の割合が大きく増加する一方、事業所得者の割合は低下している。さらに、彦根市税務データには無所得の住民も含まれており、休職や引退した者など、確定申告データでは捕捉されない層も一定数含まれている。この比較により、住民税課税データに基づく彦根市税務データが、確定申告データとは異なるカバレッジを有していることが示されている。

本稿の構成は以下の通りである。第Ⅱ節では、いわゆる「103万円の壁」について、彦根市税務データを用いて定量的に示す。第Ⅲ節では、出産・育児に伴う個人所得の変化を観察する。第Ⅳ節で

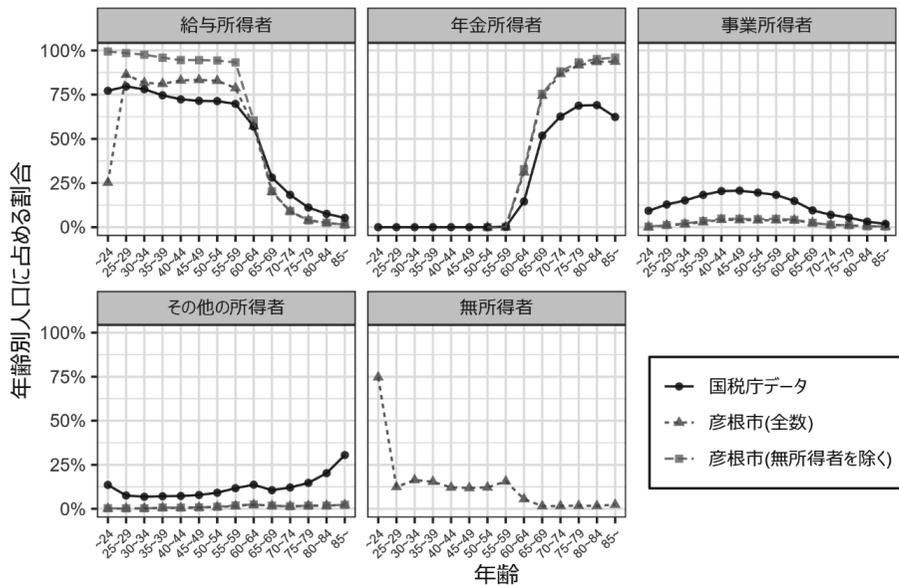


図1:彦根市税務データと国税庁確定申告データにおける所得タイプの割合

図注:国税庁データは宇南山ほか(2025,図1)より。分類は、宇南山ほか(2025)に従い、まず事業所得が全所得の過半数を超えるものを事業所得者と分類し、それ以外の人は給与収入もしくは年金所得などのうち最も高い収入・所得の種類で分類している。

は、機械学習手法を用いた住民税額の予測を行う。第V節では、新型コロナウイルス感染症拡大期における所得変化に着目し、性別による影響の差異を分析する。最後に第VI節で本稿のまとめを行う。

II 配偶者控除の103万円の壁

日本の税制において「103万円の壁」は広く知られた税制の閾値であり、配偶者控除と扶養控除の二つの税制に起因している。配偶者控除は配偶者を持つ納税者に対する控除であり、扶養控除は子どもや親などの扶養親族を持つ納税者に対する控除であり、両制度とも給与収入が103万円を長らく閾値としていた。本研究ではそのうち配偶者控除に関連する壁に焦点を当てる。配偶者控除は、配偶者の年収が一定額以下の場合に課税所得から一定額を控除する制度であり、この控除を受けられるか否かの境界が103万円に設定されていたことから、「103万円の壁」として配偶者の就業調整を誘発する要因として注目されてきた。

配偶者控除に関する103万円の壁について、以下では林(2024, 1章)をもとに近年の制度をより詳細に説明する。103万円の壁は、新聞やテレビなどのメディアでも繰り返し取り上げられ、税制の歪みとして一般社会においても広く認知されている。しかし、2017年までの制度においても配偶者の年収が103万円を超えた場合、確かに配偶者控除の対象からは外れるものの一方で配偶者特別控除の対象となるため、夫婦合計の手取り収入は労働時間を増やした際の限界的な増分が減少するとどまり、より働くと手取りが減少するという逆転現象はほぼ発生しない仕組みとなっていた²⁾。さらに、2018年以降は税制改正により、配偶者特別控除の適用範囲・控除額が拡大され、103万円

を超えても手取りの限界的な増分が減少することもなくなった。このように、税制上の103万円を超えることの実質的な不利益は既に大幅に緩和されており、制度上は103万円以下に所得を抑える効果が小さいことが指摘されている。

制度上の不利益が緩和されているにもかかわらず、103万円の壁は依然として社会的に広く認知されており、人々の就労判断に影響を与えている可能性がある。実際に就労調整が生じているかを検証するためには年収分布を詳細に観察する必要がある。その際、単に切りの良い数字である100万円に所得が集中しているのか、それとも税制上の境界として認識されている103万円を意識した就労調整が行われているのかを区別することが重要となる。このような精緻な分析を行うには、個人の所得を少なくとも1万円単位で正確に把握でき、かつアルバイトやパートなどの非正規労働者も含む大規模なデータセットが必要となる。しかし、従来の政府統計や研究者が行う調査では、回答される所得の精度やサンプルサイズの制約から、こうした精緻な分析を行うことは困難であった。

自治体が保有する税務データは、まさにこの条件を満たす理想的なデータである。実際、CREPEの自治体税務データを用いて近藤・深井(2023)は103万円の壁に関する実証分析を行った。彼らの研究では、給与所得の分布において103万円前後に顕著な集中が観察され、就労調整が実際に行われていることが定量的に示された。さらに重要な発見として、2018年の税制改正後も103万円付近での就労調整のパターンに大きな変化は見られず、税制上の実質的な不利益が解消された後も、103万円という数字が依然として就労調整の目安として機能し続けていることが明らかにされた。

2) 配偶者特別控除は配偶者の給与収入が105万円から5万円ごとに段階的に控除額が減っていく。そのため厳密には給与収入の微小な増加によって手取りが減ってしまうケースが発生しうる。

本節では、彦根市税務データを用いて、近藤・深井(2023)の分析のリプリケーションを行う。分析対象は、有配偶者に限定することで、扶養控除による就労調整を行っている者を可能な限り除外し、配偶者控除に起因する就労調整に焦点を絞っている。分析対象者の給与所得を1万円単位で集計し、その分布を観察することで、103万円付近に集中した分布が見られるかを検証する。

図2では彦根市税務データにおける2023年の有配偶者の1万円刻みの給与収入を示しており、年収103万円付近に顕著な所得の集中が確認された。すなわち、CREPEにおける自治体データを用いた近藤・深井(2023)の知見は、彦根市という異なる地域のデータにおいても頑健に再現されることが示されたこととなった。また、近藤・深井(2023)と同様に、60万円・96万円・120万円・180万円、といった12の倍数となる数値においても集中していることが観察される。以上の分析により、

彦根市税務データにおいても「103万円の壁」は先行研究の知見が確認された。

Ⅲ 出産・育児に伴う所得の変化

近年、経済学において男女の賃金格差をもたらす要因を分析する研究が盛んに行われており、その中で出産・育児の影響が重要な要因として注目されている。出産・育児の影響を実証的に明らかにするためには、休職・退職・転職による所得変化を正確に観測できること、出産を経験する家庭を多数捕捉できるサンプルサイズ、同一個人の長期的な所得を追跡できることなど、高水準のデータが求められる。こうした要件を満たすデータとして、行政が保有するデータが注目を集めている。

行政保有データを活用した代表的な研究として、Kleven et al. (2019)がある。彼らはデンマークにおける30年超の全国民の家族構成・収入・労働

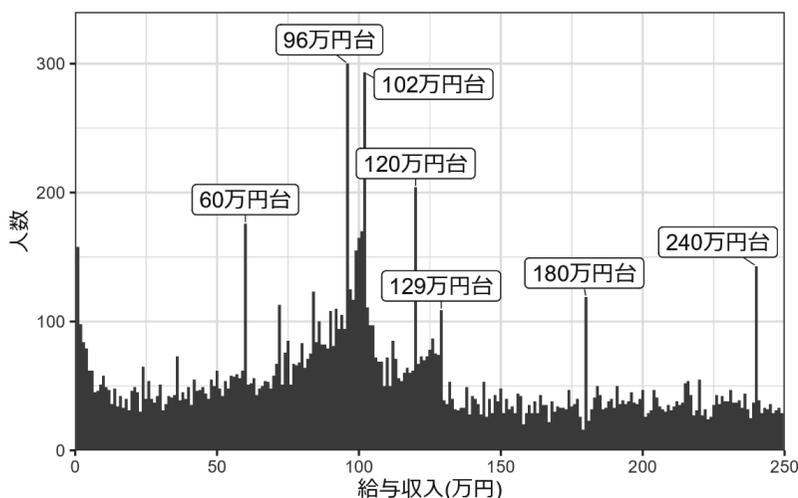


図2:2023年における彦根市の有配偶者の給与収入分布

図注:1~250万円台の人数を1万円ごとにヒストグラムで図示している。給与所得が1万円未満の人は数万人存在し、図に含めた場合レイアウトが大きく変化するため取り除いた。図中のラベルは突出した値のピンを指しており、例えば60万円台ならば600,000円から609,999円を集計するピンを指している。

時間のデータを用いて、第一子出産における収入や労働時間、賃金の変化における性差を分析した。同研究によれば、第一子出産を機に女性は収入・労働時間など大きく低下する一方、男性はほとんど変化が見られず、女性のみが家事を担い仕事を抑えていることが男女格差の大きな要因となっていることが示されている。日本でも、Fukai and Kondo (2025) がCREPEの自治体税務データを用いた分析を行い、第一子出産前後で男性の収入はほとんど変化しない一方、女性の収入は出産直後に60~80%減少することを明らかにしている。本稿では、先行研究でみられる最先端の手法を用いた定量分析は先行研究にまかせることとし、記述統計による彦根市における男女別の給与収入の第一子出産前後の変化の観察を行う。

まず、彦根市税務データにおける年齢・男女別の給与収入をみることで、基本的な収入の年齢による変化を把握し、男女格差の有無を確認する。図3では、25歳から60歳までの男性・女性の給与収入を図示している。男性・女性とも25歳時点での給与収入の平均値や中央値や分布の第一・第三四分位点に大きな違いは見られない。しかしながら、年齢上昇に伴い男性の給与収入の平均値・中央値は年齢とともに上昇している一方、女性の平均値はおおむね横ばいで中央値はやや減少して推移する。その結果、男女間の給与収入の平均値・中央値は年齢とともに拡大する。この傾向はこれまでの研究の知見に基づけば、出産・育児期における女性の休職や退職、時短勤務などの働き方を抑える行動が反映された結果の可能性はある。

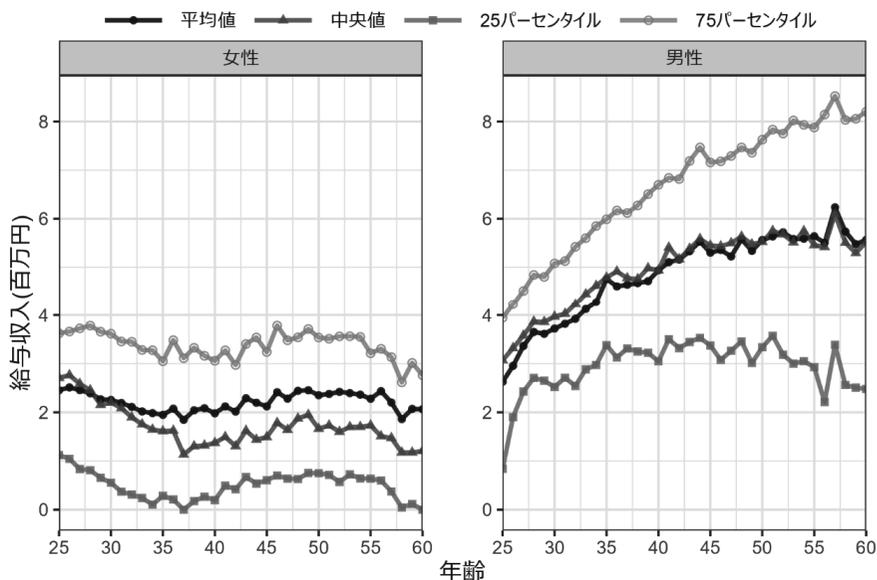


図3: 男女別の年齢と給与収入の関係
 図注: データは2023年の給与収入を用いている。

IV | 税収予測

次に、第一子の年齢別の給与収入を観察する。第一子は同一世帯内にいる最も年齢の高い子の生年とし、その生年から第一子の誕生前後の給与収入の変化を分析する。図4から男女の給与収入の平均値と中央値を比較すると、出産・育児に伴う女性にのみ発生する給与の変化が一層明確となる。男性は第一子の誕生前後に目立った変化はなく年齢に伴う給与の増加が反映されたであろう右肩上がりの直線となっている。一方、女性では第一子出産後に大きく低下する。第一子の年齢が1歳となると女性給与収入の中央値は0円となり、多くの女性が退職や休職したことが示唆される。以上の結果から第一子の出産前後において女性のみが収入が大きく減少するという先行研究と同様の結果が確認された。

自治体にとって予算編成段階における税収の正確な予測は、財政運営の安定化につながる重要な課題である。本稿で利用している彦根市税務データは市町村住民税（以下、住民税）の徴収業務に用いられるデータであり、この住民税収入は令和5年度決算において彦根市の税収全体の22%を占める。このデータを統計学的手法により将来の住民税収の予測に活用することができれば、自治体の税収予測の精度向上を通じて、より安定した財政運営に貢献することが可能となる。そこで本節では、彦根市税務データを用いて自治体の住民税収の予測精度の向上が可能であるかを検証する。

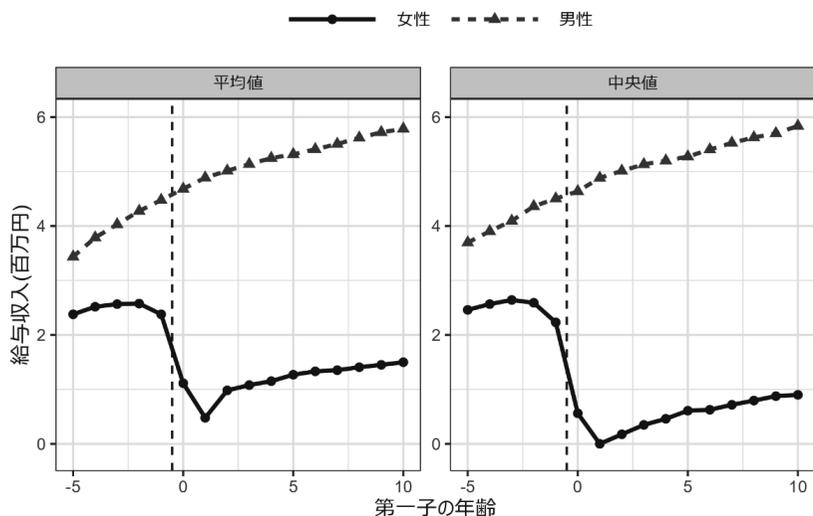


図4: 第一子の年齢別の男女の給与収入

図注: データは2014年から2023年の給与収入を用いている。横軸の負の値は第一子の生まれる(-X軸の値) 年前の収入を表している。縦の破線は0と-1の間にあり、厳密ではないが第一子の誕生時期を表す。

本節は個人レベルデータを用いた住民税予測の先行研究である深井(2022)の予測フレームワークを参考にしている。深井(2022)は、CREPEの自治体税務データから個人レベルの納税額を予測しその合計額と、他の予測と組み合わせることで自治体の住民税収の予測を行っている。本稿では、この研究の枠組みを参考に、彦根市税務データを用いて個人レベルの税収予測モデルを構築し、そのモデルが自治体の税収予測の精度向上に貢献するかを検討する。

本研究の特徴は、個人レベルのデータを活用した予測アプローチを採用した点にある。従来、マクロ経済変数などを用いて税収を予測するアプローチは数多く存在するが、このアプローチにおいては公表されているデータを利用することで十分であり、個人レベルのデータを利用する必要性は低い。そのため本稿の分析では、マクロ経済変数や年度自体の情報は用いずに個人レベルの情報のみにより住民税額の予測を行うことによって、個人レベルのデータが予測精度に貢献するかどうかを検証する。

データの構築においては、転入者の税収予測が課題となった。ある年に彦根市に転入した場合は転入翌年から住民税を納税することとなるが、最初の住民税額の予測に対してはインプットとなるデータが存在せず困難である。このような転入者による納税額は全体の4%程度であったため、

本研究では予測することは避け、簡易的な対処ではあるが、総額を過去実績値の平均値から算出し最終的な予測値に加算することとした。

個人レベルの税収を予測するモデルは機械学習モデルを用い、学習には機械学習で一般的に用いられる枠組みである、データをTrain-Validation-Testに分割するアプローチを採用した(表1)。モデルの予測対象は個人レベルの住民税であり、モデルのインプットは前年までの住民税並びに住民税の徴収に用いられた予測対象の2年前までの収入・所得データとなる。訓練用のデータを2016年から2022年の予測対象とそれに対応するインプットとし、2023年の予測対象とそれに対応するインプットをモデルの選択やハイパーパラメータのチューニングのためのValidationデータ、2024年のアウトプットとそのインプットをモデルのTestデータとする。モデルは決定木ベースの予測モデルであるLightGBMとLassoを用いる。評価指標はL1ノルムを用い、Validationデータでのサンプル外予測精度をもとに特徴量の選択、ハイパーパラメータの選択、モデルアンサンブルのウェイトを決定する。

Validationデータを用いたモデル選択は試行回数が多く全過程を記述すると膨大な量となるため主要な結果を2点のみ説明する。第一に、特徴量エンジニアリングを行うとLightGBMにおいては予測精度が大きく上昇した。とりわけ、過去の

表1:税収予測におけるデータの分割

	Train				Validation	Test
予測に用いるデータの期間	2015年まで	2016年まで	...	2021年まで	2022年まで	2023年まで
予測対象の期間(住民税額)	2016年	2017年	...	2022年	2023年	2024年

納税額や所得・収入を1年前だけではなく複数年に遡り、過去の値やそれらの変化額をインプットデータに加えることは予測精度を大きく向上させた。また、予測対象を住民税額そのものではなくその昨年度からの変化額にすることもLightGBMにおいては精度向上に寄与した。第二に、LightGBMとLassoではLightGBMの方が精度が高く、さらにLassoのアンサンブルのウェイトを0にした場合が予測精度が最も高かったため、最終的にLightGBMのみの予測モデルを採用した。

最終的な予測結果は、個人レベルでは予測精度が向上したものの、集計レベルでは単純に予測した場合と大きな差はなかった(表2)。機械学習モデルの対抗馬として「前年の住民税額マイナス1600円」という簡易的な予測との比較を行った。なお、この1600円というのは2023年までの個人レベルにおける住民税の減少額の平均値である。誤差の絶対値で計測された個人レベルの予測誤差は、機械学習モデルを使った場合には大きく減少している。しかしながら、個人の予測誤差を平均した誤差においては機械学習モデルが単純な予測の3%程度精度に予測しているが、これは予測精度の向上はわずかであったと言わざるを得ない。なぜなら、この個人レベルの予測値を合計し

それに転入者の納税額の予測値を足し合わせた市全体の税収の予測値において700万円程度の差にしかならないためである。さらに、この自然体の税収予測値は実際の税収と大きな乖離が生まれている³⁾。この実際の値との乖離と比べても、集計レベルの予測においては機械学習を使ったモデルは単純な予測と比較し予測精度の向上にはそれほど寄与しないという結果となった。

V | コロナ禍の所得の減少

経済学では経済理論や経済モデルの妥当性を検証するためにデータを用いた実証研究が広く行われている。彦根市税務データは所得や家族構成に関する詳細かつ包括的な情報を含んでおり、こうした実証研究においても有用である。本節では、こうした実証研究の一例として、新型コロナウイルス感染症が日本の労働市場にもたらした影響を分析したKikuchi, Kitao and Mikoshiba (2021) (以後、KKM) の事後検証を彦根市税務データを用いて行う。

KKMは、新型コロナウイルス感染症の初期段階で観測されたデータをもとに、マクロ経済モデルを用いてその後の雇用や所得への影響を予測

表2: 税収予測の結果

	個人レベルの予測		集計レベルの予測
	誤差の絶対値の平均(円)	誤差の平均(円)	税収の予測値(億円)
機械学習モデル	17,090	2,264	61.05
差分による予測	18,290	2,325	61.12
実績値			60.61

3) 乖離の要因としては2024年に行われた定額減税が考えられる。彦根市においては市町村住民税の所得割から平均して約4000円減額されている。

し、厚生評価を行っている。彼らの研究では、労働者の属性による影響の異質性に注目しており、正規雇用か否か、リモートワークが可能な職種か、感染リスクが高い産業か、といった要因によって経済的なショックがどのように異なるかを分析しその後の影響を予測している。

KKMの分析では様々な労働者属性が考慮されているが、そのすべてを検証することはデータの制約から困難である。そこで本稿では、彼らの主要な分析結果の一つである労働収入に対する影響の男女間の差異(KKM Fig. 12. a)に焦点を絞って検証を行う。KKMによれば、女性の方が男性に比べより非正規雇用率が高くリモートワークを行うことが難しい職種である割合が高いため、新型コロナウイルス感染症の影響をより大きく受けることが予測された。特に、男性に比べて女性の方がより大きく労働収入が減少し、その影響が長期化することが予測されている。

以上のように予測されたコロナ禍における労働収入の変化を検証するのに彦根市税務データは適したデータである。このデータは休職や失業による収入減少を捕捉でき、かつ地域の全人口をカバーしているという利点があるためである。本節の分析では、2019年に100万円以上の給与収入があった25歳から55歳までの男女を対象とし、その後の給与収入の変化を追跡する。分析対象を設定する際、以下の2点に留意した。第一に、給与収入が100万円以上という基準を設けた理由は、労働者を識別するためである。KKMの研究は労働収入に焦点を当てているが、本稿のデータでは雇用者であるか否かを厳密に判別できない。そこで、一定以上の給与収入がある25歳以上の者を労働者とみなす簡便な方法を採用した。第二に、年齢を55歳以下に限定した理由は、定年退職の影響を排除するためである。分析期間は最長で4年間

であるため、期間終了時点でも60歳未満となるよう対象を設定した。

まず、図5において2019年からの給与収入の変化率を単純に男女で比較すると、男性の方が女性に比べより所得の変化率が高いことが示されている。ただし、これは給与の年齢に対する増加率の男女間の違いを反映している可能性があり、コロナ禍の影響とは無関係である可能性も否定できない。

そこで、2014年からの変化と2019年からの変化を比較することで、コロナ禍の影響をより明確に捉えることを試みる。この比較を行った図6からは、男性は2015年時点と比べて給与所得の伸びが鈍化している一方、女性においては平均値はほぼ変わらず、中央値においては2019年以後ではより給与が増加していることが観察される。したがって男女の比較をすれば、コロナ禍以後の所得の変化は男性が女性に比べ給与が伸び悩んだことが示唆される。しかしながらこの結果も、女性の就業率の増加の影響や、女性がパート・アルバイトの割合が高いことから最低賃金上昇の影響をより大きく受けた可能性があり、コロナ禍の影響を適切に抽出できているかは議論の余地がある。

以上の簡単な分析からKKMの予測が実際の経済と整合的であったかは十分には判明せず、単純な男女比較ではKKMと整合的な結果となり、2014年との比較ではKKMの予想に反する結果が示されたこととなった。

VI | おわりに

本稿では、彦根市税務データを用いて様々な角度から分析を行った。国税庁の確定申告データとの比較では、給与や年金が主な収入源となる人の割合が大きく異なることが示された。また、経済学

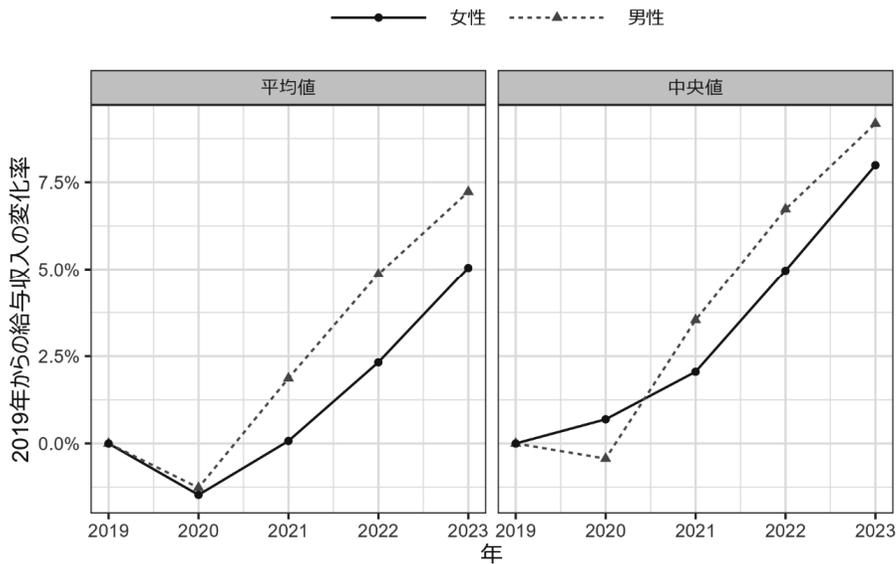


図5: 給与収入の男女別変化率

図注: 2019年に100万円以上の給与所得があった25歳から55歳までの男女を対象としている。各年で平均値・中央値を計算したのち、それらの2019年からの変化率を計算している。

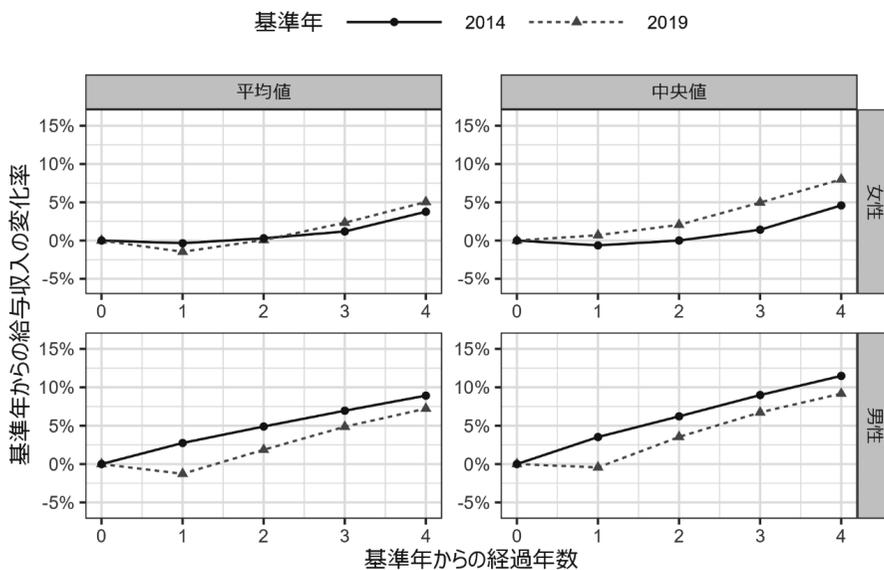


図6: 給与収入の2014年と2019年からの変化率の比較

図注: それぞれの基準年において給与所得が100万円以上あった25歳から55歳までの男女を対象にその年以後の給与所得のデータから計算している。変化率は給与所得の平均値や中央値を計算したのちに変化率を計算している。

において注目されている103万円の壁や、第一子の出産前後の給与変化の性差を明瞭に観察できることを示した。さらに、機械学習を用いた税収予測や経済理論の検証へ活用可能であることを示した。本稿で用いたデータは、本学データサイエンス学部の佐藤正昭教授を中心に管理されており、本学教員、大学院生、学部生は一定の手続きを経て学術利用が可能である。興味のある方は積極的にコンタクトを取っていただき、今後このデータを活用した様々な研究成果が生まれることを期待したい。

【付記】

本研究は、彦根市住民税課税データを学術利用可能にした彦根市役所の西田朋人氏ならびに協力していただいた職員各位、本学データサイエンス学部の佐藤正昭教授、本学データサイエンス・AIイノベーション研究推進センターの中川哲也氏のご尽力なしには実施できなかったものであり、深く感謝申し上げます。また、本学データサイエンスセミナーにおいて報告の機会をいただき、参加者の皆様から貴重なコメントを賜ったことに心より感謝する。なお、本稿で述べられている見解は著者個人のものであり、所属機関の公式見解を示すものではない。

参考文献

- ◎ Fukai, T., & Kondo, A. (2025). Parental earnings trajectories around childbirth in Japan: Evidence from local tax records. RIETI Discussion Paper Series No. 25-E-012. Research Institute of Economy, Trade and Industry. Available at: <https://www.rieti.go.jp/en/publications/summary/25020003.html> (Accessed: 2025-11-17).
- ◎ Kikuchi, S., Kitao, S., & Mikoshiba, M. (2021). Who suffers from the COVID-19 shocks? Labor market heterogeneity and welfare consequences in Japan. *Journal of the Japanese and International Economies*, 59, Article 101117. DOI: <https://doi.org/10.1016/j.jjie.2020.101117>
- ◎ Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11 (4), 181–209. DOI: <https://doi.org/10.1257/app.20180010>
- ◎ 宇南山卓・佐野晋平・勇上和史・稲葉和洋 (2025). 所得データとしての確定申告データ.『ファイナンス・レビュー』令和7年第2号(通巻第160号), 150–181.
- ◎ 川口大司・正木祐輔(2022). CREPEによるプロジェクト設立の背景とねらい：行政データと実証経済学①. 経済セミナー編集部note. Available at: <https://note.com/keisemi/n/n58b05049ba73> (Accessed: 2025-11-17).
- ◎ 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 No. 23-J-049. 独立行政法人経済産業研究所. Available at: <https://www.rieti.go.jp/jp/publications/nts/23j049.html> (Accessed: 2025-11-17).
- ◎ 林正義 (2024). 税制と経済学: その言説に根拠はあるのか. 中央経済社
- ◎ 深井太洋 (2022). 自治体税務データを用いた住民税予測：行政データと実証経済学③. 経済セミナー編集部note. Available at: <https://note.com/keisemi/n/nd01483e505de> (Accessed: 2025-11-17).

Academic Use of Administrative Data: A Case Study of Hikone City Resident Tax Data Analysis

Toshikatsu Inoue

Academic research using administrative data has expanded significantly in recent years. In line with this trend, Hikone City has granted researchers at Shiga University access to income and tax data from its resident tax records since 2025. This paper introduces the dataset and presents analytical case studies that demonstrate its research potential. We hope this will encourage both further utilization of the dataset and participation in this research project by other researchers.

