

事業所・企業系のマイクロデータを用いた データリンケージの可能性

— 賃金構造基本統計調査を例に —

村田磨理子*・伊藤伸介**

要旨

本稿は、企業の業績や財務内容が雇用や賃金に及ぼす影響についての実証分析を目指して、異なる事業所・企業系のデータにおける個票データのリンケージの可能性に関する検証を試みたものである。本稿では、最初に、賃金構造基本統計調査を対象に、複数の調査年をリンケージしたパネルデータを作成するだけでなく、共通の事業所番号に基づいて賃金構造基本統計調査と経済産業省企業活動基本調査に関するリンケージを行った。さらに、賃金構造基本統計調査と法人企業統計調査を対象に、名称・住所等を利用した名寄せによる完全照合と企業規模等の属性を用いた統計的照合の比較・検討を行うことによって、統計的照合の精度に関する検証を行った。本研究では、賃金構造基本統計調査の個票データを中心に、わが国の事業所・企業系の統計調査のデータ特性を踏まえながら、リンケージデータの作成可能性を追究することによって、データリンケージのさらなる展開を図ることが可能になった。

キーワード

賃金構造基本統計調査、マイクロデータ、データリンケージ、経済産業省企業活動基本調査

1. はじめに

マイクロデータのさらなる利用可能性を追求するために、データセットに含まれる属性群に新たな社会変数を追加的に設定することが考えられる。複数のマイクロデータ間の照合 (matching, マッチング) という方法はその1つであって¹⁾、わが国では、1970年代における政府統計のマイクロデータ間の照合に関する松田 (1978) の研究に遡ることができ、照合に

基づく同種の統計調査の個票データにおける縦断的なデータリンケージ (data linkage, データ連結) や、異種の統計調査の個票データにおける横断的なデータリンケージに関する研究が行われてきた²⁾。

ところで、わが国の統計調査は、主に、世帯・人口系の統計調査と事業所・企業系の統計調査に類別されるが、一方では個人や世帯、他方では企業や事業所が調査対象であって、個人・世帯のレコードと事業所・企業のレコードをリンクするための共通の識別子がないことから、世帯・人口系の個票データと事業所・企業系の個票データのデータリンケージ

* (公財)統計情報研究開発センター

e-mail : marikomurata@sinfonica.or.jp

** 中央大学経済学部

e-mail : ssitoh@tamacc.chuo-u.ac.jp

ジは困難である。また、世帯・人口系のデータ同士、ないしは事業所・企業系のデータ同士のリンケージに関しても、リンケージの対象となる統計調査の個票データにおいて共通の直接的な識別子が必ずしも整備されていないことから、データリンケージは容易でないのが現状である。

そうした状況の中で、周防ほか(2009)及び古隅(2015)は、事業所・企業統計調査と法人企業統計調査の調査名簿及び個票データをもとに、住所、企業名や電話番号を用いた名寄せによるデータリンケージの精度を検証し、データリンケージの注意点を指摘している。その一方で、異種の統計調査において識別番号によるリンケージが可能であったとしても、リンケージされたデータの特性を踏まえた上で、実証分析を行うことが必要と思われる。

本稿では、企業の業績や財務内容が雇用や賃金に及ぼす影響についての実証分析を目指して、賃金構造基本調査を例に、異種の統計調査の個票データのリンケージと同種の統計調査におけるパネルデータ化の検証を試みる。

2. 賃金構造基本統計調査におけるパネルデータの作成について

本節では、最初に、賃金構造基本統計調査の個票データの特徴を概括した上で、賃金構造基本統計調査の個票データをもとにしたパネルデータの作成方法を述べる。つぎに、作成された賃金構造基本調査のパネルデータにおけるデータ特性を明らかにする。

賃金構造基本統計調査は、主要産業に雇用される労働者の賃金の実態について、労働者の性別、年齢、学歴等の人口社会的属性や、雇用形態、就業形態、職種、勤続年数、経験年数といった就業属性別に把握することを目的とした基幹統計である³⁾。賃金構造基本統計調査の個票データは、2種類の調査票(事業所票と個人票)に対応した2つのファイルから構成される。事業所票データでは、1事業

所の情報が1レコードに収録され、個人票データにおいては、労働者1個人の情報が1レコードに収録されている。同一の調査年の事業所票と個人票は、都道府県番号と事業所一連番号によって照合できる⁴⁾。本研究の分析に使用した平成20~25年の個票データのレコード数については、事業所票のレコード(事業所)数が約55,000~57,000レコードであるのに対して、個人票のレコード(個人)数は、約120万~130万レコードとなっている。なお、賃金構造基本統計調査は標本調査であることから、すべての事業所をパネル化することはできないことに留意されたい。

ところで、事業所一連番号は、調査年ごとに改めて設定されることから、同じ事業所であっても調査年によって異なる番号が付与される。同様に、労働者の一連番号も調査年ごとにレコードに与えられることから、労働者の一連番号に基づいて同一個人を特定することはできない。一方で、賃金構造基本統計調査の対象事業所は、事業所・企業統計調査または経済センサスをサンプリングフレーム(抽出枠)として抽出されている。抽出枠に含まれる事業所の識別番号(事業所番号)は、賃金構造基本統計調査の個票データにも付与されている。ただし、抽出枠は数年ごとに更新される。

賃金構造基本統計調査の調査対象となる事業所が選ばれる抽出枠が同じ場合には、抽出枠における事業所番号によって、賃金構造基本統計調査と事業所・企業統計調査ないしは経済センサス基礎調査を対象に、個票データ間のリンケージを行うことが可能である。例えば、平成20~平成23年の調査については、平成18年事業所・企業統計調査の事業所番号によってリンケージを行うことができ、平成24と平成25年は平成21年経済センサス基礎調査の事業所番号に基づいて、賃金構造基本統計調査とのリンケージが可能である。他方、異なる抽出枠においては、事業所番号は

必ずしも同一番号では継続されない。しかしながら、事業所・企業統計調査及び経済センサス基礎調査の個票データには、その時点の識別番号だけでなく、当該事業所の前回調査における識別番号も保持しているため、それを經由してリンケージを行うことが可能になる。図1は、抽出枠における事業所番号の収録状況に関するイメージ図を示したものである。

図1で示されるデータの構造を踏まえて、事業所票のパネルデータ化を実際に試みる。本研究では、平成20～25年のそれぞれについて、同一事業所について当年と前年のデータをリンケージした。データリンケージにおいては、抽出枠が同一かあるいは異なるかを考慮した上で、次のような処理が必要になる。

- ① 平成25年と平成24年のリンケージには、平成21年経済センサス基礎調査の事業

所番号を利用する。

- ② 平成24年と平成23年のリンケージに当たっては、平成21年経済センサス基礎調査の個票データを使って、平成21年事業所番号と平成18年事業所番号の対応表を作成する。作成した対応表によって、事業所番号を変換してリンケージする。
- ③ 平成23年調査と平成22年調査、平成22年と平成21年、平成21年と平成20年におけるデータリンケージに関してはいずれも、平成18年事業所・企業統計調査の事業所番号が利用される。

表1は、リンケージされたデータに含まれる事業所と、その事業所に含まれる個人票に該当するレコード数をそれぞれ示したものである。抽出枠が異なる平成24年と平成23年については、リンケージされた事業所のレ

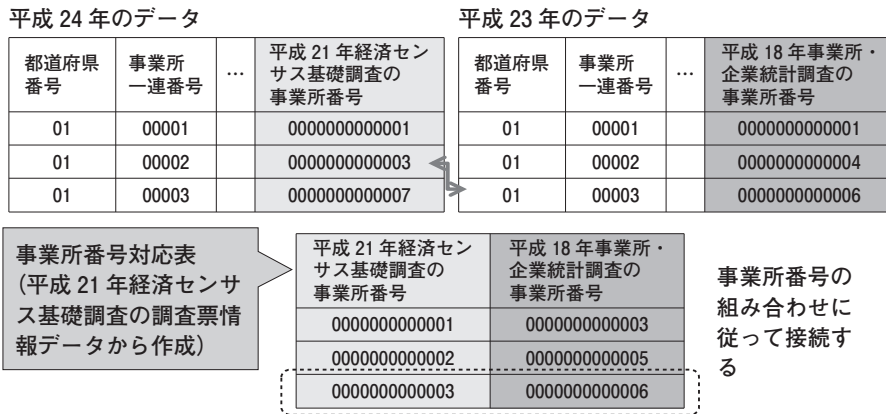


図1 抽出枠が異なる場合のリンケージの例

表1 賃金構造基本統計調査のパネルデータ化における同一事業所のリンケージ状況

リンケージした調査年	リンケージされた事業所のレコード数	同事業所に含まれる個人票のレコード数
平成25, 24年	12,758	358,185
平成24, 23年	8,377	242,817
平成23, 22年	12,899	350,020
平成22, 21年	12,303	345,451
平成21, 20年	10,193	296,684

コード数が他の調査年におけるそれよりも少なくなっている。抽出枠の変更に伴い、調査対象事業所の交代が多くなっていると推測されることから、パネルデータの利用に当たって留意が必要である。

ところで、前述のとおり、賃金構造基本統計調査の調査票情報は、抽出枠である事業所・企業統計調査及び経済センサス基礎調査と事業所番号を介して、複数の調査年にわたってリンケージを行うことが可能である。さらに、事業所・企業統計調査や経済センサ

ス基礎調査の個票データに含まれる調査事項も併せて利用することができる。したがって、賃金構造基本統計調査と経済センサス基礎調査のリンケージを行うことによって、賃金構造基本統計調査の個票データには含まれない調査事項として、例えば、資本金額、本所・支所の別、本所や親会社に関する情報が得られる。

そこで、平成25年調査と平成24年調査をリンケージしたパネルデータに含まれる事業所について分布特性を見たものが、図2と図

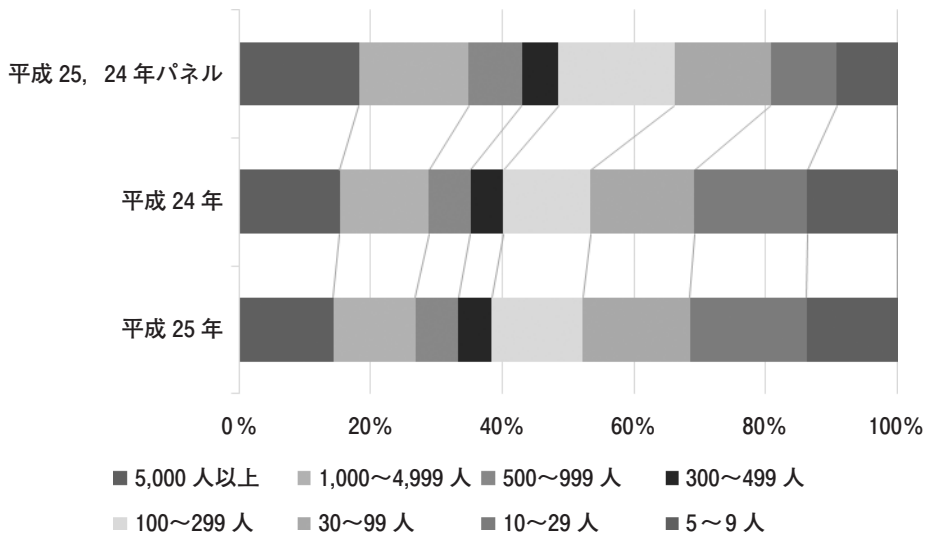


図2 賃金構造基本統計調査のパネルデータにおける企業規模別事業所数の構成割合

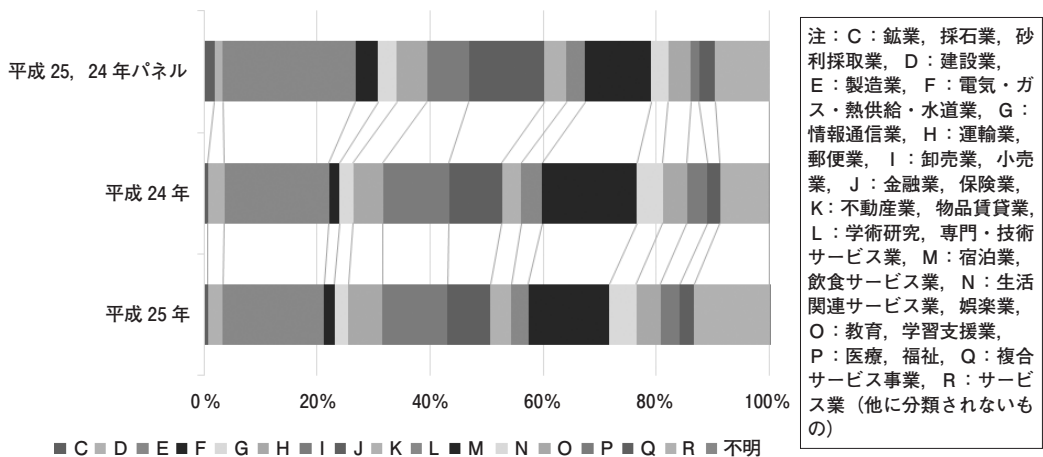


図3 賃金構造基本統計調査のパネルデータにおける産業大分類別事業所数の構成割合

3である。なお、リンケージの対象となる平成25年調査のレコード数、平成24年調査のレコード数、平成25年調査と平成24年調査のパネルデータのレコード数はそれぞれ、57,744レコード、57,426レコード、及び12,758レコードである。図2を見ると、各調査年の対象事業所全体と比較することによって、企業規模の大きい事業所の割合が高く、企業規模の小さい事業所はかなり少ないことがわかる。同様に、産業大分類別に事業所の分布を比較すると、C(鉱業、採石業、砂利採取業)、F(電気・ガス・熱供給・水道業)、J(金融業、保険業)の割合が相対的に大きく、逆に、P(医療、福祉)についてその割合が小さくなっていることが確認できる(図3)。

3. 賃金構造基本統計調査と経済産業省企業活動基本調査に関するリンケージデータの作成

賃金構造基本統計調査には、企業の業績や財務内容に関する情報が捕捉されていない。したがって、企業業績や財務内容が雇用や賃金に及ぼす影響についての実証分析を行う場合、企業の収益や財務情報に関する企業属性を得るためには、経済産業省企業活動基本調査(以下「企業活動基本調査」)および財務省所管の法人企業統計調査の個票データを利用してリンケージを行うことが求められる。そこで、本節では、賃金構造基本統計調査と企業活動基本調査⁵⁾におけるリンケージデータの特徴について述べる。

企業活動基本調査における調査対象の企業には、独自の永久企業番号が割り当てられている。さらに、平成23年以降は経済センサス基礎調査の事業所番号が収録されている。したがって、企業活動基本調査の個票データと経済センサスの個票データのリンケージを行うことは容易である。本研究では、平成24年の企業活動基本調査の個票データを用いて、賃金構造基本統計調査とのリンケージの検証

を行っている。

賃金構造基本統計調査と企業活動基本調査のリンケージにおいて留意すべき点としては、この2調査における調査対象の範囲(産業、規模)の違いや、調査の単位(事業所か企業か)の相違が指摘できる。特に、調査の単位の違いについては、例えば、複数事業所からなる企業の場合、1つの事業所(賃金構造基本統計調査)と企業全体(企業活動基本調査)がリンケージされるが、複数事業所企業においては、本所・支所事業所の情報が同時に利用できるか否か等、リンケージされたデータに含まれる情報に留意する必要がある。

リンケージの手順は、以下のとおりである。最初に、平成21年経済センサス基礎調査の事業所番号をリンケージのキー変数として用いることによって、賃金構造基本統計調査における事業所票データに対して、経済センサス基礎調査の個票データ、及び企業活動基本調査の個票データをリンケージする。つぎに、作成された賃金構造基本統計調査のリンケージデータを賃金構造基本統計調査の個人票データにリンケージすることによって、賃金構造基本統計調査における個人に関する属性を追加的に付与する。なお、リンケージにおいては、①個人票データと事業所票データがリンケージされた平成24年賃金構造基本統計調査の個票データに対するデータリンケージ、及び②平成24年と平成25年の賃金構造基本統計調査のパネルデータに対するデータリンケージが行われた。

リンケージデータにおける事業所・企業の分布特性は、表2～表5で示されている。表2を見ると、平成24年賃金構造基本統計調査と平成24年企業活動基本調査においてリンケージができた事業所・企業のレコード数は3,166レコードであって、企業活動基本調査全体の1割程度と少ないことがわかる。さらに、平成24年と平成25年の賃金構造基本統計調査のパネルデータにおいて、平成24年企

表2 賃金構造基本統計調査と企業活動統計調査のデータリンケージの状況

データの種類	リンケージされた事業所のレコード数(企業数)	同事業所に含まれる賃金個人票のレコード数
企活(平成24年)	30,647	-
賃金(平成24年)+企活(平成24年)	3,166	121,366
賃金(平成25, 24年パネル)+企活(平成24年)	1,313	53,626

注:「賃金」は賃金構造基本統計調査,「企活」は企業活動基本調査をそれぞれ表している。以下同様。

表3 賃金構造基本統計調査と企業活動統計調査のリンケージデータの特徴

企業規模別企業数・構成割合

	賃金(平成25年)		賃金(平成24年)+ 企活(平成24年)		賃金(平成25, 24年パネル) + 企活(平成24年)	
	事業所数	割合(%)	企業数	割合(%)	企業数	割合(%)
5,000人以上	8,297	14.4	124	3.9	68	5.2
1,000~4,999人	7,211	12.5	372	11.7	201	15.3
500~999人	3,760	6.5	396	12.5	190	14.5
300~499人	2,881	5.0	315	9.9	125	9.5
100~299人	7,978	13.8	1,359	42.9	565	43.0
30~99人	9,414	16.3	590	18.6	161	12.3
10~29人	10,243	17.7	10	0.3	3	0.2
5~9人	7,960	13.8	0	0.0	0	0.0
計	57,744	100.0	3,166	100.0	1,313	100.0

業活動基本調査とリンケージできた事業所・企業のレコード数は、1,313レコードであることが確認された。また、リンケージデータにおける勤め先企業の規模は、100~299人が最も多くなっている(表3)。企業の産業別に見ると、G情報通信業に該当する企業の割合が相対的に大きくなっている(表4)。また、単一事業所企業が少ないことも確認できる(表5)。

さらに、賃金構造基本統計調査における決まって支給する現金給与額の分布(図4, 表6)をみると、平成24年と平成25年の賃金構造基本統計調査のそれぞれの分布特性や、平成24年と平成25年の2時点間のパネルデータにおける分布において、2つのピークが存

在することが確認できる。それに対して、企業活動基本調査とリンケージした事業所に限定した分布については、下側のピークがないことから、分布特性について違いが見られる。

本研究におけるリンケージの技法では、複数事業所企業の場合、原則として賃金構造基本統計調査における本所事業所のデータのみが、企業活動基本調査における企業全体のデータとリンケージされている。賃金構造基本統計調査で把握される事業所には、本所事業所だけでなく、支所事業所も含まれるため、本所事業所と支所事業所の分布の違いが捉えられていないことが、リンケージデータとの分布の違いの主な理由の1つと推察される(図5)。リンケージの前に本所・支所の統合

表4 賃金構造基本統計調査と企業活動統計調査のリンケージデータの特徴

産業大分類別企業数・構成割合

	賃金（平成25年）		賃金（平成24年）+ 企活（平成24年）		賃金（平成25、24年パネル）+ 企活（平成24年）	
	事業所数	割合（%）	企業数	割合（%）	企業数	割合（%）
C	426	0.7	20	0.6	15	1.1
D	1,496	2.6	47	1.5	17	1.3
E	10,348	17.9	1,677	53.0	672	51.2
F	1,062	1.8	61	1.9	50	3.8
G	1,444	2.5	204	6.4	103	7.8
H	3,439	6.0	13	0.4	9	0.7
I	6,681	11.6	566	17.9	197	15.0
J	4,376	7.6	39	1.2	24	1.8
K	2,139	3.7	55	1.7	26	2.0
L	1,709	3.0	99	3.1	39	3.0
M	8,280	14.3	143	4.5	55	4.2
N	2,765	4.8	106	3.3	43	3.3
O	2,466	4.3	12	0.4	6	0.5
P	1,935	3.4	0	0.0	0	0.0
Q	1,501	2.6	1	0.0	1	0.1
R	7,665	13.3	123	3.9	56	4.3
不明	12	0.0	0	0.0	0	0.0
計	57,744	100.0	3,166	100.0	1,313	100.0

表5 賃金構造基本統計調査と企業活動統計調査のリンケージデータの特徴

単一事業所企業・複数事業所企業別の企業数

	賃金（平成24年）+ 企活（平成24年）	賃金（平成25、24年パネル）+ 企活（平成24年）
単一事業所企業	653	268
複数事業所企業	2,513	1,045
計	3,166	1,313

を行うことなどによって、リンケージできる範囲を拡大する可能性が考えられる。なお、支所を含めたリンケージについては、次節の法人企業統計調査との完全照合を例に論じることとしたい。

また、複数の統計調査において重複して捕捉される調査項目（例えば、産業、資本金額、従業員数）において、定義の違い等に起因するずれが生じていることが本研究で確認された。具体的に数字を挙げると、上述①のリン

(単位：百円)

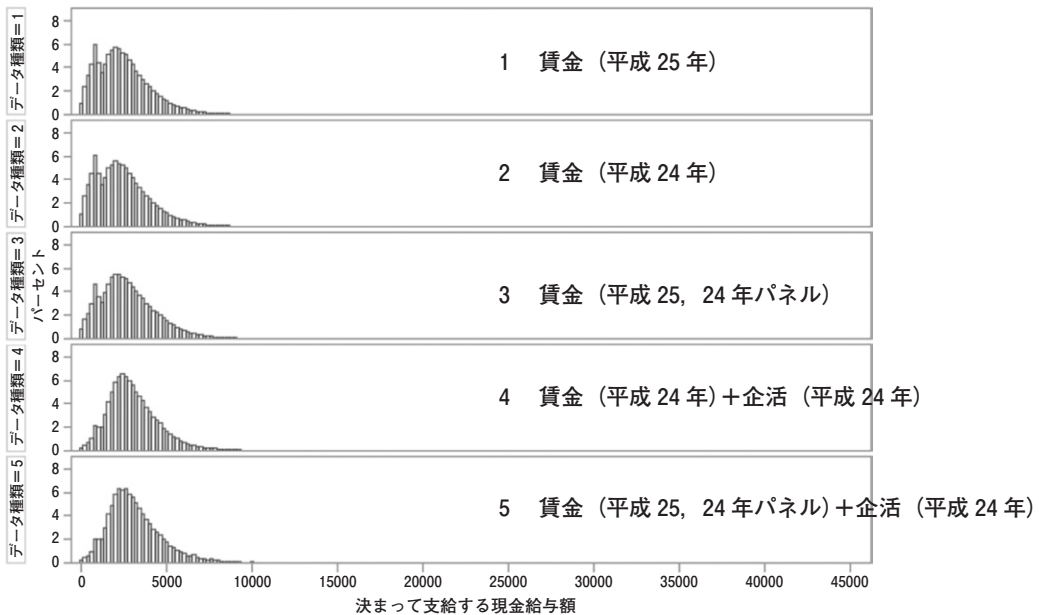


図4 賃金構造基本統計調査と企業活動統計調査のリンケージデータにおける「決まって支給する現金給与額」の分布

表6 賃金構造基本統計調査と企業活動統計調査のリンケージデータにおける「決まって支給する現金給与額」の基本統計量

(単位：百円)

	平均	標準偏差	歪度	尖度	第1四分位数	中央値	第3四分位数
賃金 (平成 25 年)	2,499	1,654	1.513	6.801	1,300	2,259	3,356
賃金 (平成 24 年)	2,494	1,676	1.461	6.044	1,250	2,250	3,365
賃金 (平成 25, 24 年パネル)	2,797	1,732	1.330	5.071	1,592	2,531	3,725
賃金 (平成 24 年) + 企活 (平成 24 年)	3,084	1,596	1.470	5.627	2,038	2,825	3,873
賃金 (平成 25, 24 年パネル) + 企活 (平成 24 年)	3,146	1,625	1.253	3.272	2,058	2,868	3,950

ケージデータにおいて、経済センサス基礎調査と企業活動基本調査の完全照合において企業産業分類が大分類レベルで一致したのは、リンケージされた事業所のうち85%であった。調査時期の違いによる企業・事業所の変化も考えるが、調査ごとの定義や設問形式に起因するずれについては、以下のような点に

留意する必要がある。例えば、企業活動基本調査の調査票では、主要な取扱品名または事業の具体的な名称及び金額が記入され、最も売上高の大きい商品で企業の産業分類が決定されるが、多角化の把握などの分析にはすぐわれないとして、各種商品卸売業、各種商品小売業及び各種物品賃貸業は分類として採用さ

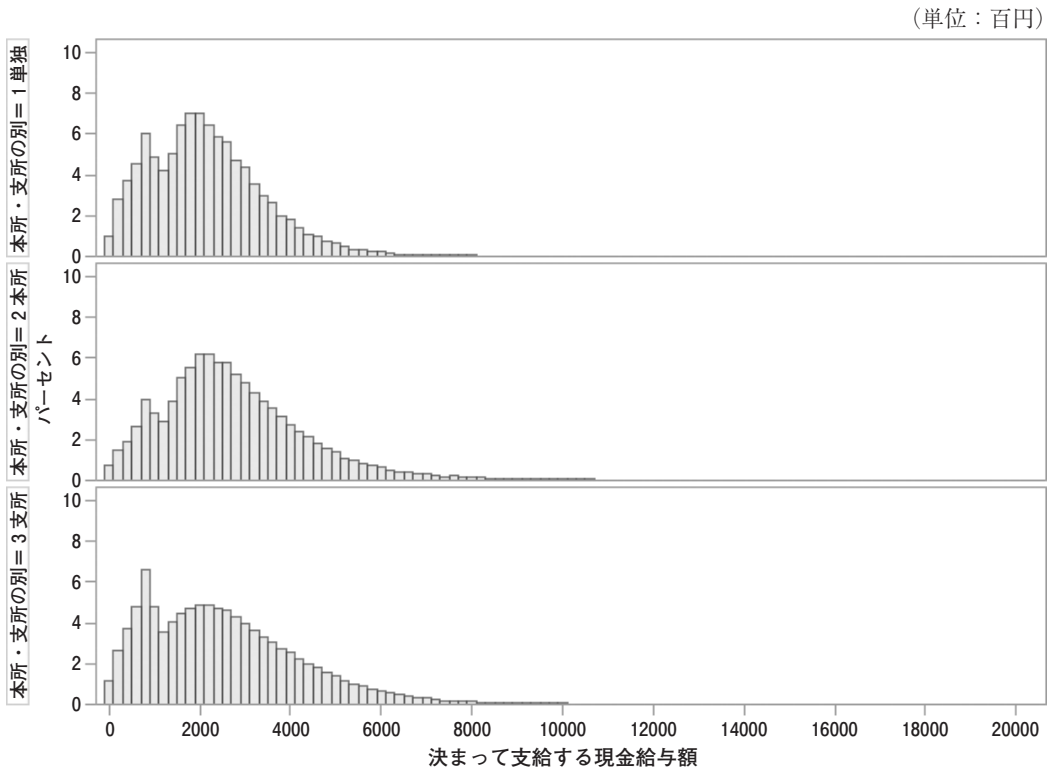


図5 本所・支所別「決まって支給する現金給与額」の分布 平成25年賃金構造基本統計調査、上段：単独事業所、中段：本所事業所、下段：支所事業所

れていない。一方、平成21年経済センサス基礎調査の調査票では、企業の産業分類は、支所を含めた企業全体の主な事業の種類に関する記入内容に基づいて決定される。主な事業の種類は、企業全体の過去1年間の総収入額又は総販売額の最も多いものと定義されているが、金額そのものは記入されないため、実際には、記入者による判断のゆらぎがあると推測される。さらに、経済センサス基礎調査では、事業所の主な事業の種類によって、事業所ごとの産業分類も付与される。したがって、リンケージデータを用いた分析において産業といった調査項目を使用する場合には、分析目的にあった選択が必要である。なお、本研究で集計を行うにあたっては、産業分類に関しては経済センサス基礎調査の企業産業分類を用いている。

4. 賃金構造基本統計調査と法人企業統計調査の名寄せによる完全照合の結果と統計的照合の評価

企業の収益や財務内容等の企業属性に関しては、法人企業統計調査を利用して分析を行うことも考えられる。そこで、本節では、賃金構造基本統計調査と法人企業統計調査の名寄せによる完全照合の状況と、統計的照合に関する評価結果について述べる⁶⁾。賃金構造基本統計調査とのリンケージにおいては、経済センサスの事業所番号のような共通の識別番号を持たないことから、本研究では、最初に、名称、所在地の文字情報や本所・支所の情報を用いて、個票データの完全照合を行った。つぎに、2つの統計調査に共通するキー項目に基づいて統計的照合を試み、統計的照合の精度評価を行った。なお、本研究では、平

成23年の賃金構造基本統計調査と法人企業統計調査の個票データを使用する。

前述のとおり、2つの調査に共通のIDは存在しない。また、賃金構造基本統計調査の個票データには事業所の名称、住所がないことから、抽出枠である平成18年事業所・企業統計調査における企業（ないしは事業所）の名称と住所を用いる。前節の企業活動基本調査とのリンケージでは、原則として本所事業所のデータのみが対象となったため、リンケージできた事業所・企業の数が少なかった。そこで、平成18年事業所・企業統計調査の個票データにある本所の名称・住所を用いることによって、支所事業所のデータについてもリンケージの対象とした。また、名寄せによる完全照合を行う上で、名称は「株」・「株式会社」を除去し、さらに余分な空白の除去やカナ・英数字を全角に変換する等の処理（正規化）が適用され、住所は市区町村レベルまでの一致で判定を行うために用いられた。なお、周防ほか（2009）では、事業所・企業統計調査と法人企業統計調査の完全照合において、法人企業統計調査の調査名簿を使用し、文字情

報による照合技法について、種々の条件による機械的処理と目視チェックを組み合わせる評価を行っている。本研究では、法人企業統計調査の個票データを使用したため、照合対象となる法人数は比較的少なく、機械的処理では住所は市区町村までの一致という緩い条件としたが、全体の目視チェックにより正確さを補っている。

本研究で作成されたリンケージデータについては、同じ企業の本所事業所と支所事業所に関する属性情報が、別々のレコードに含まれている点に留意する必要がある。ただし、事業所・企業統計調査の項目を使うことによって、本所・支所を統合することや、事業所の情報と企業の情報の両方を分析に利用することができる。

表7と表8は、名寄せによる賃金構造基本統計調査と法人企業統計調査の個票データの完全照合の結果を示したものである。検証結果から、賃金構造基本統計調査の9,337事業所が法人企業統計調査の3,191企業とリンケージすることが明らかになった。本分析によれば、法人企業統計調査の1つの企業に対

表7 賃金構造基本統計調査と法人企業統計調査のリンケージデータの特徴
企業規模別企業数

	賃金（平成23年）+ 法企（平成23年）
5,000人以上	229
1,000～4,999人	746
500～999人	467
300～499人	349
100～299人	691
30～99人	382
10～29人	202
5～9人	94
4人以下，不詳	31
計	3,191

注：「法企」は法人企業統計調査を表している。以下同様。

表8 賃金構造基本統計調査と法人企業統計調査のリンケージデータの特徴
産業大分類別企業数

	賃金(平成23年)+法企(平成23年)
A, B 農林水産業	3
C 鉱業, 採石業, 砂利採取業	23
D 建設業	154
E 製造業	1,098
F 電気・ガス・熱供給・水道業	69
G 情報通信業	128
H 運輸業, 郵便業	203
I 卸売業, 小売業	457
J 金融業, 保険業	448
K 不動産業, 物品賃貸業	179
L 学術研究, 専門・技術サービス業	53
M 宿泊業, 飲食サービス業	149
N 生活関連サービス業, 娯楽業	95
O 教育, 学習支援業	18
P 医療, 福祉	5
R サービス業(他に分類されないもの)	109
合計	3,191

して、賃金構造基本統計調査の複数の事業所がリンケージする場合が含まれることがわかった。法人企業統計調査における企業規模の分布特性をみると、約8割は企業規模100人以上の企業であり(表7)、企業の産業大分類別で見た場合には、「E 製造業」、「I 卸売業, 小売業」、「J 金融業, 保険業」に該当する企業が多いことが明らかになった(表8)。

つぎに、本研究では、賃金構造基本統計調査と法人企業統計調査を対象に、完全照合に先立ち、統計的照合を行っている。その統計的照合の精度検証について、併せて説明する。統計的照合については、2つの調査に共通して利用することが可能な、本社所在地、業種と企業規模を用いて行った。本社所在地については市区町村に関する地域区分を使ってお

り、業種に関しては原則として法人企業統計調査の業種の分類を用いた。さらに、企業規模については原則として賃金構造基本統計調査の企業常用労働者数の階級区分を使用し、法人企業統計調査は期中平均従業員数を階級化している。

ここで、住所は町名、番地等の文字情報を利用しないことに加えて、賃金構造基本統計調査と法人企業統計調査に共通の項目(業種)に関しても、分類区分に関する相違があることに留意する必要がある。そのため、業種分類は同じ範囲になるよう上位レベルに統合するなどの処理が求められる。また、企業規模に関しても、賃金構造基本統計調査における「企業常用労働者数」と法人企業統計調査における「期中平均従業員数」は、定義が異

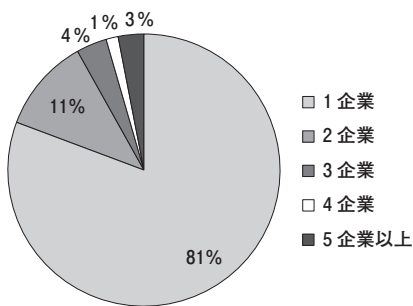
なり、正確に一致しないが、これらの項目については企業規模を表す項目として近似しているものと仮定して使用している。なお、本研究における検証対象としては、賃金構造基本統計調査の調査対象となった民営事業所の中の単独事業所・本社事業所に限定した。

図6は、本社所在地、業種、企業規模の組み合わせが同一のグループを構成する企業数別の割合を示したものである。本社所在地、業種と企業規模別にグループ分けを行うと、法人企業統計調査の対象となるレコードは18,772グループ、賃金構造基本統計調査の対象となるレコードは21,374グループにそれぞれ分けることができる。一方、1グループ

当たりの企業数の分布を確認すると、それぞれの調査においてグループの中の8割は1企業のみで構成されていることがわかる。2企業で構成されるグループが1割程度あるものの、3企業、4企業、5企業以上で構成されるグループは、それぞれ5%未満と少ない。

賃金構造基本統計調査と法人企業統計調査において、同一属性のグループを構成する企業数の観点で対応を見たのが、表9である。表9では、同じ属性のグループが照合されている。表9を見ると、一方しか該当企業がないグループが多く、「1企業」対「1企業」で照合したグループに含まれる企業数は、約2,000と少ない。

法人企業統計（全 18,772 グループ）



賃金構造基本統計（全 21,374 グループ）

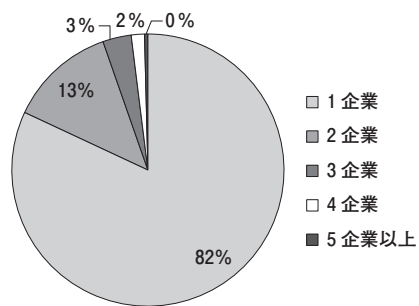


図6 本社所在地、業種、企業規模の組み合わせが同一のグループを構成する企業数別グループ数の割合

表9 3項目の組み合わせが同一のグループについて、構成する企業数別の対応関係

		賃金構造基本統計調査 1グループ当たり企業数						合計
		なし	1企業	2企業	3企業	4企業	5企業以上	
法人企業統計調査 1グループ当たり 企業数	なし	-	15,020	2,110	502	208	183	18,023
	1企業	12,482	1,981	385	136	70	96	15,150
	2企業	1,566	336	113	26	26	33	2,100
	3企業	461	120	43	21	14	21	680
	4企業	177	47	17	16	6	7	270
	5企業以上	375	89	58	25	13	12	572
	合計	15,061	17,593	2,726	726	337	352	36,795

注：3項目は、本社所在地、業種、企業規模

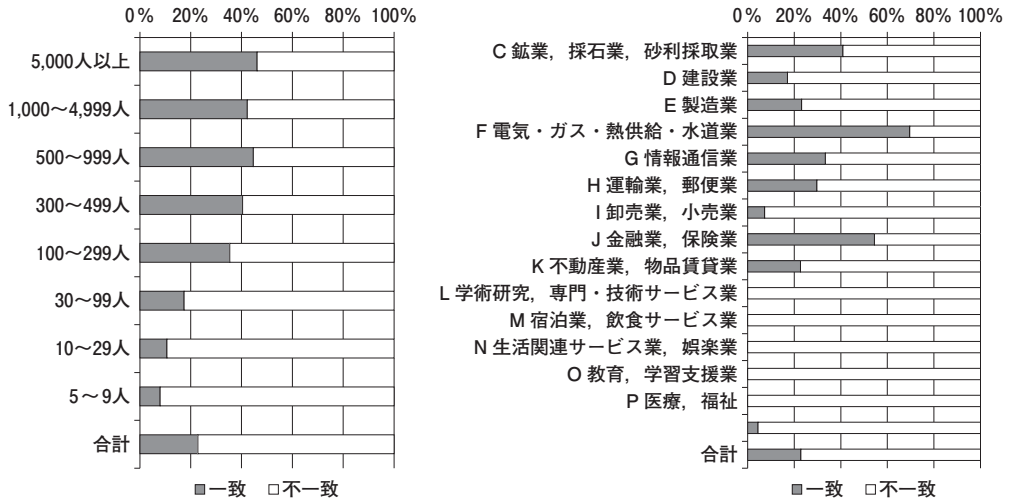


図7 統計的マッチングにおける一致率 左：企業規模別，右：産業大分類

そこで、「1企業」対「1企業」で照合した約2,000組について、完全照合の結果と比較し、統計的照合の一致率を確認したのが、図7である。図7からは、企業規模が大きいほど一致率が高くなるのがわかる。また、産業大分類別にみると、「電気・ガス・熱供給・水道業」、「金融業、保険業」においては、照合された組のほぼ半数が一致しているが、「宿泊業、飲食サービス業」など5業種では、それらのすべてが一致していないことが確認できた。本結果によれば、両方の調査が標本調査であることから、企業規模や業種分類で類別された母集団となる法人企業群において企業数が少ない場合、該当する企業は両方の調査のいずれにも捕捉され、統計的照合がなされたレコードにおいて正しく照合されている可能性が高くなるのがわかった。一方、そうでない場合には、一方の調査のみにおいて捕捉される対象となっていることから、正確に照合されていないと考えられる。

5. おわりに

本稿では、賃金構造基本統計調査の個票データにおけるデータリンケージの可能性を追究した。本研究の独自性として、企業の業

績や財務内容が雇用や賃金に及ぼす影響についての実証分析を指向する形で、賃金構造基本統計調査における縦断的なデータリンケージ、さらには賃金構造基本統計調査の経済産業省企業活動基本調査あるいは法人企業統計調査に対する横断的なデータリンケージを行い、リンケージデータのデータ特性について比較・検証を行った点が指摘される。事業所・企業統計調査や経済センサス基礎調査の事業所番号を介することによって、賃金構造基本統計調査と企業活動基本調査のリンケージが可能であることはわかったが、その一方で、事業所と企業のマイクロデータ間の照合を行うことから、それぞれの統計調査における調査項目の定義や分類区分の相違、さらには単一事業所と複数事業所におけるデータ特性の相違が、リンケージの精度に影響を及ぼすことが、本研究から明らかになった。事業所・企業系のマイクロデータのリンケージにおいては、リンケージの前に本所と支所に該当するレコードの統合を行ったり、統計調査における調査項目の定義を調整したりすることによって、リンケージできる範囲の拡大、およびリンケージデータの精度を高めることが求められよう。他方、賃金構造基本統計調査

と経済産業省企業活動基本調査のリンケージデータについては、企業規模の大きな企業が、リンケージされやすい傾向になるため、リンケージデータによる実証分析⁷⁾を行う場合には、セレクションバイアスの可能性について留意する必要があると思われる。

他方、本研究では、賃金構造基本統計調査と法人企業統計調査を対象に、名称・住所・電話番号等による名寄せによる完全照合を行うだけでなく、本社所在地、業種と企業規模に基づく統計的照合を行い、完全照合と比較することによって、統計的照合の一致率を検証した。その結果、法人企業統計調査において標本として選ばれやすい従業者規模の大きな企業については、統計的照合における一致率が高くなっていること、産業大分類別にみると、産業によって一致率が大きく異なることが実証的にも確認された。

近年、わが国においては、政府統計に関するデータアーカイブが検討されてきた。第Ⅱ期「公的統計の整備に関する基本的な計画」(平成26年3月25日閣議決定)においても、「統計データ・アーカイブ(仮称)」については、調査票情報等の提供及び活用の促進の基礎として、引き続き具体化に向けた検討を行うことの必要性が指摘されている。こうした点を踏まえると、本研究において、異種の

統計調査、とりわけ事業所・企業系の統計調査の個票データにおけるデータリンケージの可能性を追究したことは、わが国において政府統計のデータアーカイブを展開する上でも意義があると思われる。

その一方で、将来的には、政府統計間のマイクロデータだけでなく、政府統計マイクロデータと行政記録データのリンケージや、行政記録データ同士のリンケージの可能性も考えられる。その意味では、本研究において、リンケージを行うために使用した識別子(名前、住所、個体識別番号等)をどういった形で保管するかは、データアーカイブの将来的な課題であるだけでなく(森(2008)、伊藤(2016))、政府統計の二次的利用のさらなる展開を図る上で重要かと考える。他方、リンケージデータを分析する観点からは、共通の母集団データベースにおいて、企業と所属する各事業所との対応関係、各種の標本調査における調査対象の該当・非該当が把握できれば、リンケージを容易に行うことができる。また、調査項目の定義についても網羅的に確認できることが望ましい。このようなデータリンケージに必要な属性情報に関するメタデータの整備についても、データアーカイブにおける今後の課題と言えよう。

謝辞

本研究は、平成26年度一橋大学経済研究所共同利用共同研究拠点事業プロジェクト研究「企業の業績と雇用政策が家計の就業・資産選択に及ぼす影響に関する計量分析」(研究代表者：北九州市立大学 林田実)及び平成27年度日本学術振興会科学研究費補助金「家計の資産選択や労働供給に関する政策評価のためのマイクロシミュレーション分析」[基盤研究(C)(課題番号 15K03399)](研究代表者：中央大学 伊藤伸介)における研究成果の一部を発表するものである。

本研究において使用した「賃金構造基本統計調査」、「事業所・企業統計調査」、「経済センサス基礎調査」、「企業活動基本調査」及び「法人企業統計調査」の調査票情報は、統計法第33条に基づき提供を受けたものであり、本稿で作成した集計表等は提供を受けた調査票情報を独自集計したものである。記して関係各位に御礼申し上げます。

注

- 1) ミクロデータ間の照合(マッチング)は、完全照合(exact matching)と統計的照合(statistical matching)に大別される(美添・荒木(1999, 2000))。アメリカでは、1950年代半ばから1960年代にかけて、製造業センサスデータの完全照合による事業所縦断データ(Longitudinal Establishment Data)の編成が、ラグルス夫妻(R. Ruggles and N. Ruggles)等によって試みられた(Ruggles and Ruggles(1999)p.xvi)。また、1970年代に入ると、個人の所得申告ファイル(tax return)、アメリカセンサス局の経常人口調査(Current Population Survey)、及びアメリカ連邦準備制度理事会(Federal Reserve Board)の消費者家計状況調査(Survey of Financial Characteristics of Consumers)に関する統計的照合が、連邦商務省の経済分析局(Bureau of Economic Analysis)によって実施された(Budd(1971))。さらに、Okner(1972)、Ruggles and Ruggles(1974)等によって、異種のマイクロデータ間における横断的なデータリンケージについての研究が展開されてきた。
- 2) わが国では、総務省統計局が実施する家計調査と貯蓄動向調査におけるリンケージに関する研究が行われただけでなく(美添・荒木(1999))、経済産業省企業活動基本調査(以下「企業活動基本調査」)と総務省の科学技術研究調査のリンケージ等、異種のマイクロデータ間のデータリンケージが行われてきた(松田(1999, 200頁))。さらに、2009年における統計法の全面施行後、様々な統計調査間のデータリンケージに基づく実証分析が行われてきた。例えば、栗原・坂田(2015)や坂田(2015)は、「法人企業統計調査」と「法人企業景気予測調査」の個票データにもとに作成したリンケージデータを用いて、法人企業を対象にした判断項目と企業実績との関連性に関する計量分析を行っている。なお、マイクロデータを用いた統計的照合の精度検証に関する実証研究については、例えば栗原(2015)等を参照されたい。
- 3) 調査の対象は、全国の5人以上の常用労働者を雇用する民営事業所(5~9人の事業所については、企業規模が5~9人の事業所に限定)、及び10人以上の常用労働者を雇用する公営事業所で日本標準産業分類(平成19年11月改定)に基づく16大産業[C: 鉱業, 採石業, 砂利採取業, D: 建設業, E: 製造業, F: 電気・ガス・熱供給・水道業, G: 情報通信業, H: 運輸業, 郵便業, I: 卸売業, 小売業, J: 金融業, 保険業, K: 不動産業, 物品賃貸業, L: 学術研究, 専門・技術サービス業, M: 宿泊業, 飲食サービス業, N: 生活関連サービス業, 娯楽業, O: 教育, 学習支援業, P: 医療, 福祉, Q: 複合サービス事業, R: サービス業(他に分類されないもの)]である。
- 4) 利用申請者に対して提供された賃金構造基本統計調査の個票データにおいて、事業所の名称と所在地、及び労働者の氏名は、収録されていない。一般に、統計法33条に基づく調査票情報の利用申出において、個体の名称や所在地といった直接的な識別子は提供されない。
- 5) 企業活動基本調査は、日本標準産業分類に掲げる大分類C-鉱業, 採石業, 砂利採取業, 大分類E-製造業, 大分類F-電気・ガス・熱供給・水道業(中分類35-熱供給業及び中分類36-水道業を除く。), 大分類G-情報通信業の一部, 大分類I-卸売業, 小売業, 大分類J-金融業, 保険業の一部, 大分類K-不動産業, 物品賃貸業の一部, 大分類L-学術研究, 専門・技術サービス業の一部, 大分類M-宿泊業, 飲食サービス業の一部, 大分類N-生活関連サービス業, 娯楽業の一部, 大分類O-教育, 学習支援業の一部及び大分類R-サービス業(他に分類されないもの)の一部に属する事業所を有する企業のうち、従業者50人以上かつ資本金額又は出資金額3,000万円以上の会社を対象としている。
- 6) 法人企業統計調査は、法人の企業活動の実態を明らかにし、併せて法人を対象とする各種統計調査のための基礎となる法人名簿を整備することを目的に、財務省が統計法に基づく基幹統計調査として実施している調査である。営利法人等を対象とする標本調査(年次別調査と四半期別調査)であって、平成20年度調査からは、「金融業, 保険業」が調査対象に含まれている。なお、調査事項としては、資産・負債・純資産の状況や損益等について確定した決算の計数が捕捉される。
資本金1億円以上の管理法人は、基本的に、調査年が変わっても同一法人には同じ法人番号が割り当てられており、パネルデータ化が比較的容易である。しかしながら、他の統計調査とのリンケージについては、経済センサス等の共通の識別番号を持たないため、名称、所在地等による事業所ないしは企業との名寄せが必要となる。
- 7) 伊藤・出島(2015)は、本稿において議論された賃金構造基本統計調査と企業活動基本調査のリン

テージデータをもとに、企業業績や財務情報が企業の個別労働者における賃金率に及ぼす影響を実証的に明らかにしている。

参考文献

- [1] 伊藤伸介・出島敬久(2015)「企業業績が個別労働者の賃金に与える効果に関するマイクロデータ分析—企活と賃金センサスのデータリンクテージをもとにして—」,『経済学論纂』(中央大学) 56巻, 1・2合併号, 13~37頁.
- [2] 伊藤伸介(2016)「わが国における政府統計のデータシェアリングの現状と課題」『情報管理』, Vol. 58, No. 11, 836~843頁.
- [3] 栗原由紀子(2015)「統計的マッチングにおける推定精度とキー変数選択の効果—法人企業統計調査マイクロデータを対象として—」,『統計学』108号, 1~15頁.
- [4] 栗原由紀子・坂田幸繁(2015)「企業判断の情報特性と期待形成モデルの比較—『法人企業景気予測調査』および『法人企業統計調査』のリンクテージデータから—」,『統計研究参考資料』No. 116, 法政大学日本統計研究所.
- [5] 坂田幸繁(2015)「景気予測調査における判断カテゴリー選択のミクロ的数量特性」『立教経済学研究』第69巻第2号, 55~72頁.
- [6] 周防節雄・古隅弘樹・宮内環(2009)「法人企業統計調査と事業所・企業統計調査の統合データによる企業データベース:1983~2005年」,『統計数理』, Vol. 57, No. 2, 277~303頁.
- [7] 古隅弘樹(2015)「公的産業統計を組み合わせた遡及的データベースの構築」,『産業構造の寛容と公的統計の利用(中間報告)』, 統計数理研究所共同研究リポート337, 5~23頁.
- [8] 松田芳郎(1978)『データの理論』岩波書店.
- [9] 松田芳郎(1999)『マイクロ統計データの描く社会経済像』日本評論社.
- [10] 森博美(2008)「情報資産としての統計と政府統計データアーカイブ」『統計学』第94号, 15~25頁.
- [11] 美添泰人・荒木万寿夫(1999)「家計調査と貯蓄動向調査の統計的マッチング」(財統計研究会『平成10年度 統計的マッチングにより発生する誤差の要因等の検証に関する調査研究報告書』, 9~97頁.
- [12] 美添泰人・荒木万寿夫(2000)「完全照合」松田芳郎・伴金美・美添泰人編『講座マイクロ統計分析②:マイクロ統計の集計解析と技法』日本評論社, 239~250頁.
- [13] Budd, E.C. (1971) "The Creation of a Microdata File for Estimating the Size Distribution of Income", *Review of Income and Wealth*, Series 17, No. 4, pp.317-334.
- [14] Okner, B.A. (1972) "Constructing a New Data Base From Existing Microdata Sets. The 1966 MERGE File", *Annals of Economic and Social Measurement*, Vol. 1, No. 3, pp.325-342.
- [15] Ruggles, N.D. and Ruggles, R. (1974) "A Strategy for Merging and Matching Microdata Sets", *Annals of Economic and Social Measurement*, Vol. 3, No. 2, pp.353-371.
- [16] Ruggles, N.D. and Ruggles, R. (1999) *Macro- and Microdata Analyses and Their Integration*, Cheltenham; Northampton, Mass.

Potential of Data Linkage using Business Microdata based on Basic Survey on Wage Structure

Mariko MURATA^{*}, Shinsuke ITO^{**}

Summary

This paper explores the potential of data linkage for conducting detailed empirical analysis into the influence of corporate financial conditions on employment and wage structure using business microdata. In order to create data for this analysis, exact matching was used to generate longitudinally linked data from the Japanese Basic Survey on Wage Structure. In addition, original data from the Basic Survey on Wage Structure and the Basic Survey of Japanese Business Structure and Activities were linked via cross-sectional data linkage based on the business establishment ID number.

In a separate step, exact matching (based on name, address etc.) and statistical matching were conducted using data from the Basic Survey on Wage Structure and the Financial Statements Statistics of Corporations by Industry, and the results were compared in order to determine the accuracy of data linkage.

This paper aims to advance the usage of data linkage as an analytical tool for corporate attributes contained in business microdata.

Key Words

Basic Survey on Wage Structure, Microdata, Data Linkage, Basic Survey of Japanese Business Structure and Activities

^{*} Statistical Information Institute for Consulting and Analysis

^{**} Faculty of Economics, Chuo University

編集委員会からのお知らせ
機関誌『統計学』の編集・発行について

編集委員会

1. 常時、投稿を受け付けます。
2. 次号以降の発行予定日は、
第111号：2016年9月30日、第112号：2016年3月31日です。
3. 投稿に際しては、「投稿規程」、「執筆要綱」、「査読要領」などをご熟読願います。
4. 原稿は編集委員長（下記メールアドレス）宛にお送り願います。
5. 原稿はPDF形式のファイルとして提出して下さい。また、紙媒体での提出も旧規程に準拠して受け付けます。紙媒体の送付先は編集委員長宛（住所は会員名簿をご参照下さい）をお願いいたします。
6. 原則としてすべての投稿原稿が、編集委員会またはレフェリーによる査読の対象となります。
7. 通常、査読から発刊までに要する期間は、査読が順調に進んだ場合でも2ヶ月間程を要します。投稿にあたっては十分に留意して下さい。

編集委員会や投稿応募についての問い合わせは、
下記メールアドレス宛に連絡下さい。
また、編集委員長へのメールアドレスも下記になります。

editorial@jsest.jp

編集委員長 朝倉啓一郎（流通経済大学）
副委員長 藤井輝明（大阪市立大学）
編集委員 前田修也（東北学院大学）
山田 満（関東支部所属）
橋本貴彦（立命館大学）

[注記] 『統計学』の定期刊行を守るために、できるかぎり早期のご投稿をお願いいたします。111号（2016年9月30日発行予定）への掲載を想定すると、A：「論文」「研究ノート」の場合、2016年7月初旬、B：その他の原稿の場合、2016年7月末、をメ切の目途にご投稿いただく必要があります。

以上

編集後記

研究成果をご投稿下さいました執筆者の皆様、査読に関わって下さいました皆様に心より御礼申し上げます。次号より朝倉啓一郎新編集委員長にバトンタッチをします。次回大会の会員総会に向けて、機関誌活性化にむけた課題を議論し始めております。着手できることから始めて参りますので、会員の皆様方からの積極的な投稿を期待いたします。今後ともよろしくお願いいたします。

（長澤克重 記）

執筆者紹介 (掲載順)

村田磨理子 ((公財) 統計情報研究開発センター)
伊藤伸介 (中央大学経済学部)

支部名

事務局

北海道	062-8605	札幌市豊平区旭町 4-1-40 北海学園大学経済学部 (011-841-1161)	水野谷武志
東北	986-8580	石巻市南境新水戸 1 石巻専修大学経営学部 (0225-22-7711)	深川通寛
関東	192-0393	八王子市東中野 742-1 中央大学経済学部 (042-674-3424)	芳賀寛
関西	525-8577	草津市野路東 1-1-1 立命館大学経営学部 (077-561-4631)	田中力
九州	870-1192	大分市大字旦野原 700 大分大学経済学部 (097-554-7706)	西村善博

編集委員

長澤克重 (関西) [長] 朝倉啓一郎 (関東) [副]
前田修也 (東北) 橋本貴彦 (関西)
山田満 (関東)

統計学 No.110

2016年3月31日 発行	発行所	経済統計学会 〒194-0298 東京都町田市相原町4342 法政大学日本統計研究所内 TEL 042(783)2325 FAX 042(783)2332 http://www.jsest.jp/
	発行人	代表者 菊地進
	発売所	音羽リスマチック株式会社 〒112-0013 東京都文京区音羽1-6-9 TEL/FAX 03(3945)3227 E-mail: otorisu@jupiter.ocn.ne.jp 代表者 遠藤誠

STATISTICS

No. 110

2016 March

Articles

Potential of Data Linkage using Business Microdata based on Basic Survey on Wage Structure
..... Mariko MURATA, Shinsuke ITO (1)

Activities of the Society

Activities in the Branches of the *Society* (19)

JAPAN SOCIETY OF ECONOMIC STATISTICS
