



ESRI Research Note No.85

生産性分析のための雇用者被用者マッチデータの 構築について

鈴木通雄

June 2024



内閣府経済社会総合研究所
Economic and Social Research Institute
Cabinet Office
Tokyo, Japan

ESRI Research Note は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません（問い合わせ先：<https://form.cao.go.jp/esri/opinion-0002.html>）。

ESRI リサーチ・ノート・シリーズは、内閣府経済社会総合研究所内の議論の一端を公開するために取りまとめられた資料であり、学界、研究機関等の関係する方々から幅広くコメントを頂き、今後の研究に役立てることを意図して発表しております。

資料は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません。

The views expressed in “ESRI Research Note” are those of the authors and not those of the Economic and Social Research Institute, the Cabinet Office, or the Government of Japan.

生産性分析のための雇用者被用者マッチデータの構築について *

鈴木通雄

東北大学、内閣府経済社会総合研究所

1 イントロダクション

雇用者被用者マッチデータは、事業所の売上高、従業者数等の情報と従業員の詳細な属性情報の両方を含むデータであり、賃金や生産性格差の詳細な分析のために各国で研究利用が進んでいる。^{*1} 日本でも様々な公的統計の接合が行われているが、本稿では、製造業の事業所調査である工業統計調査と賃金構造基本統計調査（以下「賃金センサス」と表記する）を接合することによる雇用者被用者マッチデータの構築について検討する。

工業統計調査と賃金センサスの接合データを使用した先行研究としては、1993年から2003年までのパネルデータを用いて賃金と生産性の勤続年数プロファイルを比較したFukao, et al. (2006)、接合データの作成方法の詳細な説明を行った川口、神林 (2010)、2002年と2012年のクロスセクションデータを用いて、賃金の輸出プレミアムを分析した伊藤 (2017) などが挙げられる。本稿では、個別事業所の生産性の決定要因や、集計生産性における労働移動の寄与度についての詳細な分析のための接合データの構築を主な目的として、Fukao, et al. (2006)、川口、神林 (2010) のあとの期間である2006年から2014年までの期間をカバーするマッチデータの構築方法を説明し、その接合結果を検証する。

* 川口大司氏、神林龍氏、深井太洋氏には、調査票情報の接合方法や先行研究についてご教示いただいた。丹後健人氏、家舗弘志氏には、リサーチアシスタントとしてご協力いただいた。内閣府経済社会総合研究所での勉強会参加者の皆さんには貴重なコメントをいただいた。本研究はJSPS 科研費 19K01584、23K01394 の助成を受けたものである。深く謝意を示したい。本稿中の見解等は筆者個人のものであり、また、誤りは全て筆者の責に帰すものである。

^{*1} 雇用者被用者マッチデータを利用した研究については、関連文献のサーベイである Abowd and Kramarz (1999) や Lentz and Mortensen (2010) を参照。より最近の研究例としては、Bonhomme, et al (2019) などがある。

工業統計調査と賃金センサスの事業所個票の接続は、以下の2通りの方法で行う。まず第1に、平成24年経済センサス活動調査が2011年の実績値調査のために、工業統計のかわりに行われるとともに、2014年の賃金センサスの抽出母集団として使用されていることを利用し、平成24年経済センサスを介した工業統計調査と賃金センサスの接合を試みる。第2に、川口、神林(2010)等の既存研究と同様、名簿情報(事業所住所、電話番号など)による接合を行う。工業統計調査、賃金センサスともにパネル化できるため、これらの方法でマッチできるのは当該年だけでなく、その事業所が両調査に参加している全ての期間にわたる。

賃金センサスがパネル化可能である1998年から2014年について、接合作業を行った結果、2014年の接合率は79.1%と高水準となった。これは平成24年経済センサスを介した接合と名簿情報の照合による接合の両方を行ったためである。2006年から2013年までの接合率は58.5%から70.8%で推移したが、2004年と2005年は、賃金センサスの改定やサンプルフレームの変更の影響から接合率が18.5%、29.7%まで低下した。1998年から2003年までは、事業所名簿情報が利用可能ではなかったため、ほとんど接続することができず、接合率は6%から12%となった。この接合結果から、2006年から2014年のデータならば、生産性分析に利用できると考えられる。ただ、この期間の接合率は比較的高いとはいえ、分析サンプルの代表性が損なわれている可能性がある。そこで、全数調査である工業統計調査で利用可能な産出、投入の各変数の平均値を工業統計調査個票データとマッチデータで比較した。その結果、マッチデータの平均を賃金センサスの調査ウェイトで調整することにより、両データ間での各変数の平均の差が1-2%に収まり、さらに設備投資比率の外れ値を除去するとその差がさらに小さくなることが確認できた。

本稿の構成は以下の通りである。第2節では使用するデータの詳細を説明し、第3節ではデータの接合方法を説明する。第4節では接合率などの接合結果を報告し、第5節では今後の課題を述べる。

2 データ

本章では、個票接続を目指す工業統計調査と賃金センサス、そして、接続に必要な名簿情報等を提供する事業所・企業統計調査、経済センサスを解説する。

2.1 工業統計調査

工業統計調査は、製造業の実態把握と行政施策の基礎資料を提供することを目的として、明治42年（1909年）から行われている統計法に基づく基幹統計である。本調査は、日本標準産業分類の「大分類 E-製造業」に属する4人以上の事業所（国に属する事業所を除く）を対象とする全数調査（裾切調査）として、2020年まで行われた。^{*2} 2021年は後述する経済センサス活動調査が行われたため、工業統計調査は実施されず、2022年からは、経済構造実態調査に包摂され、製造業事業所調査として実施されることになった。^{*3}

調査票は従業者数に応じて「甲調査票」（従業者30人以上）と「乙調査票」（従業者4人以上29人以下）の2種類に分かれている。^{*4} 調査項目には、経営組織、従業者数、給与総額、出荷額など両調査票に共通して含まれるものがある一方、有形固定資産や在庫などのように、甲調査票には毎年含まれるものの、乙調査には5年毎しか含まれないものなどもある。調査は経済センサス活動調査実施年をのぞいて毎年6月1日現在で行われ、主に前年1月から12月までの1年間の実績について行われる。2020年工業統計調査（2019年実績）では192,047事業所が対象となり、回収率は95.1%、オンライン提出率は32.5%であった。

本研究では、1986年から2018年までの工業統計調査個票（2011年、2015年は経済センサス活動調査）を用いる。恒久棚卸法による資本ストックの推計等、一部データの作成では、工業統計調査個票がパネル化可能である1986年からのデータを利用し、賃金構造基本統計調査個票との接合では、1998年から2014年のデータを利用する。また、接合データの生産性分析への応用のため、サンプルを、より詳細なデータが継続して利用可能である30人以上の従業者を持つ事業所に限定する。

2.2 賃金構造基本統計調査（賃金センサス）

賃金センサスは、日本全国の主要産業に雇用される労働者の賃金の実態を、雇用形態、就業形態、職種、性別、年齢、学歴、勤続年数、経験年数等に分けて詳細に明らかにする

^{*2} 調査年の1の位が0、3、5、8の年は従業員1-3人の事業所も対象である。

^{*3} 製造業事業所調査では、製造業における大分類、中分類、小分類又は細分類ごとに売上高を上位から累積し、当該分類に係る売上高総額の9割を達成する範囲に含まれる事業所に調査対象が限定された。また、母集団名簿（サンプルフレーム）も工業調査準備調査名簿から事業所母集団データベースに変更されたため、2022年以前と以降の事業所個票の接続は難しくなる可能性がある。

^{*4} 甲、乙調査票の区分を決定する従業者数は、「この事業所に従事している人の男女計」（工業統計調査調査票）であり、具体的には、個人業主、無給家族従業者、有給役員、常用雇用の合計から、出向・派遣送出者を除き、出向・派遣受入者を足した数である。雇用期間が1か月未満の臨時雇用者は含まれない。

ことを目的として、昭和 23 年から毎年行われている統計法に基づく基幹統計である。本調査は日本標準産業分類に基づく 16 大産業の 5 人以上の常用労働者を雇用する民営事業所と 10 人以上の常用労働者を雇用する公営事業所を対象として、事業所を第一次抽出、労働者を第二次抽出とする層化二段抽出法を用いて行われる標本調査である。調査は、毎年 7 月に 6 月分の賃金等と調査前年 1 年間の賞与、期末手当等特別給与額について行われる。

表 1 は、1989 年（平成元年）から 2019 年（令和元年）までの賃金センサスのサンプルフレームを表す。賃金センサスは、サンプルフレームの事業所 ID（市区町村コード・調査区番号・事業所番号）を収録しているため、サンプルフレームに収録されている事業所名、住所、電話番号などの事業所名簿情報との接続が可能になる。以下で詳述するが、研究利用のための二次利用申請が可能である事業所・企業統計調査と経済センサスはパネル化できるため、それらと接合される賃金センサス事業所個票もパネル化できる。しかし、各事業所の労働者（個人）個票は、たとえ事業所が複数期間にわたって抽出されたとしても、パネル化はできない。

本研究では、厚生労働省から 1989 年から 2019 年までの賃金センサス調査票情報（個人票、事業所票）の提供を受けた。このうち主に用いるのは、パネルデータ化可能な 1998 年から 2014 年のデータである。

2.3 事業所・企業統計調査、経済センサス

事業所・企業統計調査（平成 8 年以前は事業所調査）は、日本国内の事業所および企業の基本的構造を把握し、各種標本調査の母集団情報を収集する目的で実施される統計調査である。調査は、農林漁業の個人経営事業所、家事サービス業、外国公務の事業所を除く全事業所が対象となる。調査項目は、事業所の名称、所在地、業態、従業者数、電話番号などの基本的な情報の他に、平成 8 年以降、企業の国内外活動の多角化や電子商取引の状況など、時代に即した情報が加えられている。

事業所・企業統計調査は昭和 22 年に始まり、昭和 61 年から平成 18 年までは 5 年ごとに実施された。平成 8 年以降は、調査の中間年に簡易調査が実施された。しかし、この期間、各産業の大規模統計調査は産業分野ごとに異なる年次および周期で実施されており、同一時点における包括的な産業構造統計を作成することはできなかった。この問題を解決するために、経済活動を同一時点で網羅的に把握することを目的として、経済センサスが導入された。

まず、事業所・企業統計調査と同様に、事業所および企業の産業、従業者規模などの把

Table1: 賃金構造基本統計調査（賃金センサス）のサンプルフレーム

賃金センサス調査年	サンプルフレーム
1989年（平成元年）	昭和61年事業所・企業統計調査
1990年（平成02年）	昭和61年事業所・企業統計調査
1991年（平成03年）	昭和61年事業所・企業統計調査
1992年（平成04年）	昭和61年事業所・企業統計調査
1993年（平成05年）	平成03年事業所・企業統計調査
1994年（平成06年）	平成03年事業所・企業統計調査
1995年（平成07年）	平成03年事業所・企業統計調査
1996年（平成08年）	平成06年事業所・企業統計調査
1997年（平成09年）	平成06年事業所・企業統計調査
1998年（平成10年）	平成08年事業所・企業統計調査
1999年（平成11年）	平成08年事業所・企業統計調査
2000年（平成12年）	平成08年事業所・企業統計調査
2001年（平成13年）	平成08年事業所・企業統計調査
2002年（平成14年）	平成11年事業所・企業統計調査
2003年（平成15年）	平成11年事業所・企業統計調査
2004年（平成16年）	平成13年事業所・企業統計調査
2005年（平成17年）	平成13年事業所・企業統計調査
2006年（平成18年）	平成16年事業所・企業統計調査
2007年（平成19年）	平成16年事業所・企業統計調査
2008年（平成20年）	平成18年事業所・企業統計調査
2009年（平成21年）	平成18年事業所・企業統計調査
2010年（平成22年）	平成18年事業所・企業統計調査
2011年（平成23年）	平成18年事業所・企業統計調査
2012年（平成24年）	平成21年経済センサス基礎調査
2013年（平成25年）	平成21年経済センサス基礎調査
2014年（平成26年）	平成24年経済センサス活動調査（民営）
2015年（平成27年）	平成25年次フレーム（事業所母集団データベース）
2016年（平成28年）	平成26年次フレーム（事業所母集団データベース）
2017年（平成29年）	平成27年次フレーム（事業所母集団データベース）
2018年（平成30年）	平成28年次フレーム（事業所母集団データベース）
2019年（令和元年）	平成29年次フレーム（事業所母集団データベース）

握と各種標本調査の母集団情報の収集を目的とした経済センサス基礎調査が平成 21 年に実施され、それ以降 5 年周期で行われている。さらに、事業所の基本情報に加えて、年間総売上や費用等の経理項目についても調査を行う経済センサス活動調査が平成 24 年に実施された。その後、経済センサス活動調査も平成 28 年および令和 3 年に 5 年周期で実施されている。経済センサスはどちらの調査も当該年の 6 月 1 日に実施され、売上、費用等の経理事項については前年 1 年間の実績値が調査される。

本研究では、総務省から平成 8、11、13、16、18 年の事業所・企業統計調査、平成 21 年経済センサス基礎調査、平成 24 年経済センサス活動調査の調査票情報の提供を受けた。利用が承認された調査事項は、各年の事業所 ID（市区町村コード、調査区番号、事業所番号）、平成 8 年以外の年について、前回調査時の事業所 ID、平成 13 年とそれ以降について、事業所名簿情報（名称、所在地、郵便番号、電話番号）である。

3 雇用者被用者マッチデータの構築

本章は、工業統計調査と賃金センサスの事業所個票の接続方法について解説する。両データの接合は以下のプロセスで進められる：(1) 工業統計調査と賃金センサスのパネル化（3.1 節）；(2) パネル化済み工業統計調査と賃金センサスの接合（3.2 節）。

3.1 各データのパネル化手順

本節では、工業統計調査、事業所・企業統計調査、経済センサス、賃金構造基本統計調査のパネル化手順を解説する。

3.1.1 工業統計調査

各年の事業所は、都道府県、市区町村、事業所番号で一意に識別できるが、事業所番号は原則 5 年ごとに振り直される。（1989、1997、1999、2010 年は事業所番号の振りなおしなし。）また、市区町村の変更があった場合には、該当住所の事業所番号が振り直される。したがって、そのままではパネル化できないが、新旧の事業所 ID（都道府県、市区町村、事業所番号）の対応が個票データとともに提供されるコンバータで得られるため、それを用いてパネル化する。本研究では、阿部、人見、小西、富田、内野（2012）および行本（2015）の方法に従い、準備調査名簿、産業編個票、提供用コンバータを使用してパネル化を行う。

工業統計調査個票のパネルデータ化には、準備調査名簿、産業編個票、提供用コンバー

タの各ファイルを使用する。準備調査名簿は、調査の前に前年の調査結果をもとにして作成される名簿であり、1993年以降、電子ファイルで利用可能で、1年前の調査の結果と当該年の都道府県、市区町村の実査により新たに加えられた地域内の全ての事業所が含まれる。いったん名簿に記載されると一定期間（通常5年）名簿にとどまり、調査対象外（退出、転業など）の事業所も含まれる。これら調査対象外のデータは5年ごとの事業所番号の振り直し時に名簿から削除される。2007年以降は、「アクションコード」（開設、転入、既設、廃業、転業のコード。2008年以降は訂正種別という変数名）が追加され、退出（廃業「5」、転業「6」）を正確に把握できるようになった。産業編個票（甲票（従業者30人以上）、乙票（従業者4-29人））は、調査票の質問項目に対する回答が記載されている。したがって、各事業所の出荷額、従業員数、有形固定資産などの情報はこのファイルに収録されている。事業所ID（都道府県、市区町村、事業所番号）はあるが、企業名、事業所名、住所などの名簿情報は収録されていない。準備調査名簿の作成時点では把握されていなかったが、調査時点で把握された事業所も含まれている（行本（2015））。提供用コンバータは、経済産業省から提供される新旧の事業所ID（都道府県、市区町村、事業所番号）の対応表である。事業所番号の変更が1件もない年については作成されない。5年に1度の事業所番号降り直しでは全ての事業所について作成される。

パネルデータ作成の具体的な手順は行本（2015）に詳しいが、以下で概要を説明する。

1. 準備調査名簿と産業編個票を一对一で接続する。これは、準備調査名簿にのっていないが、産業編個票に載っている事業所が存在するためである。これを「補完済名簿」とよぶ。
2. 補完済名簿と提供用コンバータの接続する。事業所コードの場合は、一对一、自治体番号の場合は、多対一（m:1）で接続する。提供用コンバータに事業所IDの記載がない場合は、同じ事業所IDを転記する。これを「更新済名簿」と呼ぶ。
3. 更新済名簿を接続開始年から、次年度の事業所IDで順次接続する。次年度の事業所IDが欠損の場合、一对一接続ではエラーが発生するので、多対一で接続する。接続後に事業所IDに重複が発生した場合は、郵便番号と電話番号で事業所の照合を行い、正しい接続だけ残す。全ての年で接続作業が終わったら、データをLong形式に変換する。

2011年と2015年の実績データについては、工業統計調査ではなく、経済センサス活動調査（平成24年、28年）が行われたので、これらの個票データについても経済産業省に二次利用申請を行う。工業統計調査個票と経済センサス活動調査個票の間のコンバーターも提供を受けることができるため、2011年と2015年を含むパネルデータの構築が可能

である。ただし、母集団名簿がそれぞれの調査で異なるためか、接続できる事業所数が工業統計調査間での接続と比べて少なくなってしまう。この点は、図 1 で明らかにする。

図 1 は、今回パネル化できた 1986 年から 2018 年の工業統計調査の事業所（従業者 30 人以上）のサンプルサイズと各事業所の観測期間を示したものである。図 1a は、サンプルサイズの推移を示す。実線は、各年のクロスセクションの事業所数を示し、破線は各年で、当該年と前年の 2 年間どちらも観測される、「一期ラグ観測あり」の事業所数を示す。クロスセクションサンプルサイズに加えて、一期ラグ観測ありのサンプルサイズは、個票データの接続可能性や上記の接続作業に間違いがないかを確認するために有用である。クロスセクションと一期ラグ観測ありのサンプルサイズの差は、2010 年まで安定的に推移しているが、2011 年、2012 年、2015 年、2016 年に一期ラグ観測ありのサンプルサイズが大きく落ち込むことが確認された。2011 年、2015 年は工業統計調査の代わりに、経済センサス活動調査が行われた年であり、この一期ラグ観測ありサンプルサイズの落ち込みは、工業統計調査個票と経済センサス活動調査個票の接合率が工業統計調査間での接合率より低いことを示している。クロスセクションのサンプルサイズは経済センサス活動調査とその前後の工業統計調査ではほぼ同じであるため、名簿情報等を駆使することにより、接合率を改善できる可能性はあり、都道府県名、市区町村名、企業名、従業員数、敷地面積、資本金の照合による追加的な接続を試みたものの、接合率は図 1a の水準でとどまった。クロスセクションのサンプルサイズについては、2000 年まで減少傾向が続いたが、それ以降は 45000 程度で推移している。

図 1b は、事業所の観測期間のヒストグラムを示す。各事業所の観測期間は、従業員数が欠損でない期間として計算した。この図からは、全期間の 33 年間観測されている事業所が一番多いことが確認される。観測期間の平均は 20.8 年間、中央値は 22 年間である。

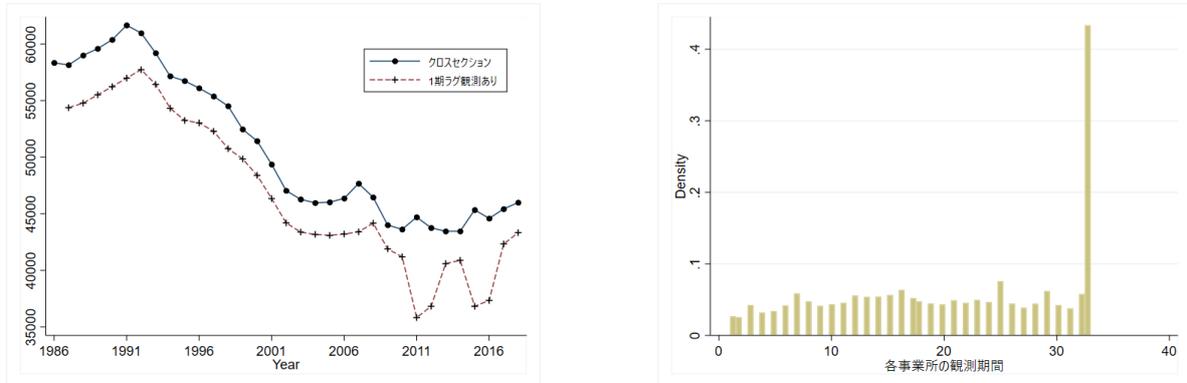
3.1.2 事業所・企業統計調査、経済センサス（基礎調査、活動調査）

事業所・企業統計調査、経済センサスの事業所 ID（市区町村番号、基本調査区番号、事業所番号）も毎回変わる。しかし、総務省への二次利用申請の際、当該年の事業所 ID に加え、前回調査時の事業所 ID 情報を申請することができる。その情報を用いてパネル化する。パネル化した個票データで、各事業所に固有の ID を付与する。

3.1.3 賃金センサス

賃金センサスには、サンプルフレームである事業所・企業統計調査または経済センサスの事業所 ID（市区町村番号、基本調査区番号、事業所番号）が収録されているので、そ

Figure1: 工業統計調査（1986-2018）



(a) サンプルサイズ

(b) 事業所観測期間

注：従業者 30 人以上の事業所のサンプルを使用している。図 1a は、サンプルサイズの推移を示す。実線は、各年の事業所数を表し、破線は各年の事業所の内、当該年とその前年のどちらも観測される事業所の数である。図 1b は、パネル化された 1986 年から 2018 年までの工業統計調査の各事業所の観測期間のヒストグラムである。事業所の観測の有無は、従業員数が欠損でないことでカウントした。

の情報をを用いてサンプルフレームの調査個票と接続可能である。^{*5}前節で説明したように、事業所・企業統計調査、経済センサス（基礎調査、活動調査）はパネル化できるので、接合した賃金センサスの事業所個票もパネル化される。事業所・企業統計調査、経済センサス（基礎調査、活動調査）の中間年の賃金センサスについては、共通のサンプルフレームをもつため、サンプルフレームの事業所 ID でパネル化できる。これは、中間年においては、賃金センサスに収録済みの事業所 ID だけでパネル化できることを意味する。

賃金センサスの調査票情報は個人票と事業所票ファイルで提供されるが、企業名、事業所名、事業所住所や電話番号などの名簿情報は含まれていない。したがって、賃金センサスの事業所名簿情報を取得するためにも、それらの情報が含まれている事業所・企業統計調査、経済センサスと接合することが必要となる。この名簿情報は、以下で示す通り、賃金センサスと工業統計調査に収録される事業所の照合の際に使用される。

図 2 は、今回パネル化できた 1998 年から 2014 年の賃金センサスのサンプルサイズと各事業所の観測期間を示したものである。工業統計調査との比較のため、サンプルは、製造業に分類される常用労働者 30 人以上の事業所を用いる。^{*6} 図 2a は、常用労働者 30 人

^{*5} 2015 年（平成 27 年）以降は賃金センサスのサンプルフレームが事業所母集団データベースに変更され、賃金構造基本調査で提供されるのは市区町村番号と事業所母集団データベースの事業所 ID である共通事業所コードとなった。しかし、事業所母集団データベースは研究目的での二次利用はできない。

^{*6} 工業統計調査の個票の甲乙調査票区分は、事業所に従事する臨時雇用以外の人の合計が 30 人以上かどうか

以上のサンプルサイズの推移を示す。賃金センサスは 2005 年に調査票の大幅な改訂が行われ、常用労働者に正社員・正職員かどうかの区分が追加され、臨時労働者についても調査されるようになった。^{*7} 事業所のクロスセクションのサンプルサイズにもその影響がみられ、2005 年以前は 9,000 から 10,000 程度あったものが、2005 年以降は 6,000 程度に減少したことが確認される。

一期ラグ観測ありのサンプルサイズは、サンプル期間を通して減少傾向であるが、特に 2006 年には 118 まで落ち込んだ。表 1 によれば、2006 年は賃金センサスのサンプルフレームが変更された年であり、他の年と比べて抽出事業所が大幅に変更された可能性がある。図 2b は、各事業所の観測期間のヒストグラムを示す。工業統計調査と異なり、1-2 年間しか観測されない事業所が多いことが確認される。事業所の観測期間の平均は 3.8 年間、中央値は 3 年間である。

3.2 工業統計調査と賃金センサスの接合

本節では、パネル化された工業統計調査と賃金センサスの事業所個票の接合手順について解説する。接合は以下の 2 通りの方法で行う：(1) 平成 24 年経済センサス活動調査を介した接続、(2) 事業所名簿情報の照合による接続。表 2 は、それぞれの接合方法で照合に利用する個票情報のタイミングを示す。以下では、各接合方法の詳細を説明する。

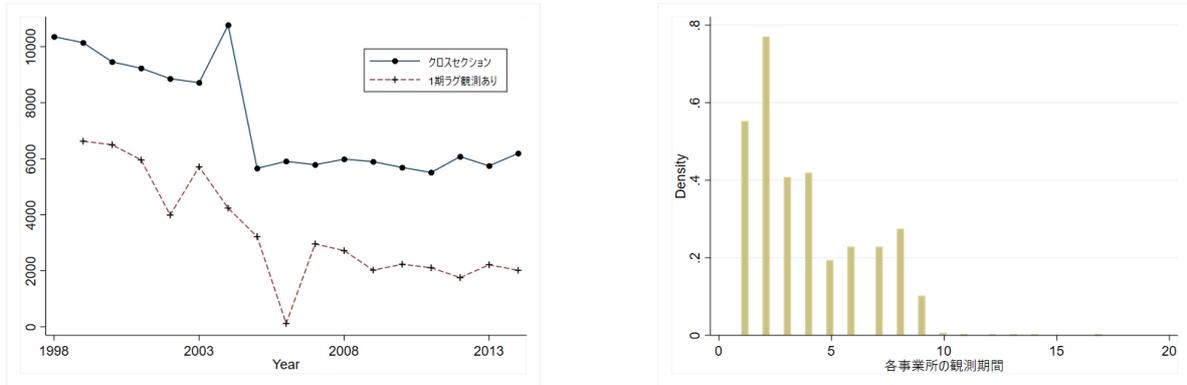
3.2.1 平成 24 年経済センサス活動調査を介した接合

まず、表 1 の通り、2014 年賃金センサスのサンプルフレームは平成 24 年経済センサス活動調査であり、賃金センサス個票に収録される各事業所には平成 24 年経済センサス活動調査の事業所 ID が付与される。さらに、平成 24 年経済センサス活動調査は工業統計調査の代わりに実施され、2011 年の実績データを提供する。前後の年の工業統計調査個票との接続が可能であるため、2014 年データと接続可能な事業所については、平成 24 年経済センサス活動調査の事業所 ID により、賃金センサスと工業統計調査個票の接続が可

かで決定される。これは、出向・派遣の受入や送付がある場合は、常用労働者 30 人以上のサンプルとは一致しない。しかし、工業統計調査においては、その差は平均 0.7% であり、非常に小さいことが確認された。したがって、参照する従業員数の定義の齟齬によるサンプルの違いは限定的であると考えられる。また、常用労働者 30 人以下の事業所も含めた場合でも、図 2a、2b のパターンは変わらない。常用労働者 30 人以下の事業所も含めた場合は、クロスセクションのサンプルサイズが 2005 年以前で 16000-19000 程度、2005 年以降で 10000 程度で推移し、事業所観測期間の平均が 3.1 年、中央値が 2 となる。

^{*7} Lise et al (2010) も賃金センサスの 2005 年の大幅な改定に言及しており、2004 年から 2005 年の間に、女性のパートタイム従業員の数、賃金や労働時間の分散に非連続的な上昇が見られることを報告している。

Figure2: 賃金センサス（1998-2014）



(a) サンプルサイズ

(b) 事業所観測期間

注：工業統計調査との比較のため、製造業に分類される常用労働者数 30 人以上事業所にサンプルを制限している。図 2a は、サンプルサイズの推移を表す。実線は各年の事業所数、破線は各年の事業所のうち、その前年も観測される事業所の数を示す。図 2b は、パネル化された 1998 年から 2014 年までの賃金センサスの各事業所の観測期間のヒストグラムを示す。事業所の観測の有無は、従業員数が欠損でないことでカウントした。

Table2: 工業統計調査と賃金センサスの接合

工業統計調査	賃金センサス		接合 1	接合 2
	調査年	サンプルフレーム		
実績年	調査年	サンプルフレーム	H24 経セ	名簿
2006	2006	平成 16 年事業所・企業統計調査		✓
2007	2007	平成 16 年事業所・企業統計調査		✓
2008	2008	平成 18 年事業所・企業統計調査		✓
2009	2009	平成 18 年事業所・企業統計調査		✓
2010	2010	平成 18 年事業所・企業統計調査		✓
2011 (H24 経セ)	2011	平成 18 年事業所・企業統計調査		✓
2012	2012	平成 21 年経済センサス基礎調査		✓
2013	2013	平成 21 年経済センサス基礎調査		✓
2014	2014	平成 24 年経済センサス活動調査 (民営)	✓	✓

注：「H24 経セ」は平成 24 年経済センサス活動調査を表す。列「接合 1」は平成 24 年経済センサス活動調査を介した接続、「接合 2」は事業所名簿情報による接続であり、✓ は、その年の個票情報を接合に利用したことを示す。工業統計調査と賃金センサスはどちらもパネル化されているので、各年の情報で事業所が接続された場合、それぞれのパネル全体が接続される。

能となる。

3.2.2 名簿情報による接合

平成 24 年経済センサス活動調査を介した接続では、2014 年に工業統計調査と賃金センサスの両方で調査された事業所のパネルだけが接続可能である。したがって、2014 年については、高い接合率が期待できるものの、他の年については接続できる事業所が多く残されている可能性がある。そこで、先行研究と同様に、工業統計調査と賃金センサスの事業所名簿情報を照合することによる接続も試みる。この接続においても、もしある年の名簿情報で接続できたら、その事業所のパネル全てを接続する。照合に使用する名簿情報は、都道府県番号、市区町村番号、郵便番号、電話番号である。この情報を用いて、 $m:m$ の接続を行い、もし重複が発生したら、企業名、男女の常用労働者数を照合し、マッチしたものをサンプルに残す。

今回の個票情報の二次利用申請では、1993 年以降の工業統計調査の名簿情報の提供をうけたが、賃金センサスと接合させる事業所・企業統計調査の名簿情報については、2001 年（平成 13 年）以降しか申請できなかった。平成 13 年事業所・企業統計調査は、2004 年賃金センサスのサンプルフレームであるため、2004 年以降の賃金センサスで事業所名簿情報が利用可能である。したがって、名簿情報による事業所個票の接続は、2004 年以降に限られる。

4 接合結果

表 3 は工業統計調査と賃金センサスの両方がパネル化可能である 1998 年から 2014 年までの各年についての両調査個票の接合結果を示す。工業統計調査は全数調査であるのに対して、賃金センサスは標本調査であるため、賃金センサスの方がサンプルサイズが小さい。そのため、表 3 は、賃金センサスとマッチデータのサンプルサイズとそれらの比率である接合率を示す。図 2 で見た通り、賃金センサスのサンプルサイズは、2005 年にそれまでの約 10,000 から約 6,000 に減少している。

接合の状況については、まず、2014 年の接合率が 79.1% と高水準になっていることが確認される。これは、平成 24 年経済センサス活動調査を介した接合と名簿情報の照合による接合の両方を行っているためであると考えられる。しかし、図 1 および 2 でみたように、工業統計調査と経済センサス活動調査個票の接続率が低い点や、賃金センサスにおける各事業所の観測期間が短い点などの問題があり、2014 年以外の年の接合率は低下する。

Table3: 工業統計調査と賃金センサスの接合結果

	1998	1999	2000	2001	2002	2003
賃金センサス	10,356	10,143	9,456	9,231	8,857	8,716
マッチデータ	643	639	631	630	472	1,036
接合率	6.2%	6.3%	6.7%	6.8%	5.3%	11.9%
	2004	2005	2006	2007	2008	2009
賃金センサス	10,771	5,658	5,908	5,786	5,985	5,897
マッチデータ	3,195	1,047	3,454	3,758	4,167	4,047
接合率	29.7%	18.5%	58.5%	64.9%	69.6%	68.6%
	2010	2011	2012	2013	2014	
賃金センサス	5,687	5,509	6,076	5,745	6,193	
マッチデータ	3,953	3,569	4,263	4,069	4,896	
接合率	69.5%	64.8%	70.2%	70.8%	79.1%	

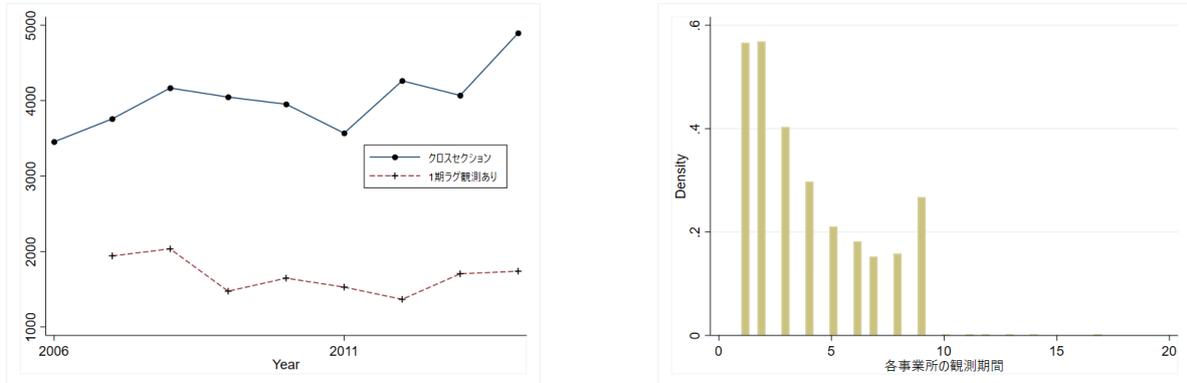
注:「賃金センサス」の行は、製造業に属し、常用労働者 30 人以上の事業所数を示し、「マッチデータ」の行は、工業統計調査と賃金センサスを接合したデータにおける各年のサンプルサイズを示す。「接合率」は、賃金センサスとマッチデータのサンプルサイズの比率である。マッチデータは、工業統計調査の甲調査票区分（事業所に従事する臨時雇用以外の人の数が 30 人以上）に属する事業所にサンプルを限定している。事業所に従事する臨時雇用以外の人の数は、常用労働者数とは厳密には等しくないが、その差は小さいため、接合率への影響は限定的であると考えられる。

それでも、名簿情報での接合が可能な 2004 年から 2014 年については、賃金センサスの改定やサンプルフレームの変更の影響を受けたと考えられる 2004 年、2005 年以外の年では、60-70% の接合率が達成された。^{*8} 1998 年から 2003 年については、名簿情報の照合ができないため、接合率はさらに下がり、6% から 12% の水準で推移した。

図 3 は、図 1 および 2 と同様の方法で、2006 年から 2014 年までのマッチデータのサンプルサイズの推移と事業所の観測期間分布を示したものである。対象期間での接合率が 60% から 80% であるので、図 2a と比較すると、クロスセクション、一期ラグ観測ありのサンプルサイズは小さくなっている。経年の変化のパターンについては、接合率が 2014 年に向けて増加することを反映し、特にクロスセクションのサンプルサイズが増加する。図 3b も図 2b と同様に 1-2 年間の観測期間が一番多くなっているが、3 年以上の観測期

^{*8} 接合の状況をより詳しく見るために、名簿情報を使わず、平成 24 年経済センサス活動調査を介した接合のみでの接合率も計算したところ、2014 年は予想通り 76.4% と高い値であったが、2013 年には 29.8% と大きく減少し、そこから徐々に減少しながら 2006 年に 23.2% となった。2005-2006 年は、賃金センサスパネルの接続率が大幅に減少するため、工業統計調査との接合率も 2005 年には 0.12% まで低下する。したがって、2014 年以外の年では、名簿情報による接合が重要であることがわかった。

Figure3: マッチデータ (2006-2014)



(a) サンプルサイズ

(b) 事業所観測期間

注：工業統計調査と賃金センサスを接合したマッチデータのサンプルを使用。図 3a は、サンプルサイズの推移を表す。実線は各年の事業所数、破線は各年の事業所のうち、その前年も観測される事業所の数を示す。図 3b は、2006 年から 2014 年までのマッチデータ各事業所の観測期間のヒストグラムを示す。事業所の観測の有無は、従業員数が欠損でないことでカウントした。

間を持つ事業所の割合が大きくなっている。観測期間の平均は 3.9 年、中央値は 3 年である。

賃金センサスは、事業所とその労働者をそれぞれ抽出する層化二段抽出法を用いて行われる標本調査である。したがって、変数の集計をする際には、調査（サーベイ）ウェイトを用いて行うことが必要となる。賃金センサスでも事業所と常用労働者の抽出率が調査ウェイトとして提供されている。表 3 で見たように、2006 年から 2014 年の期間で、マッチデータのサンプルサイズは、賃金センサスの 6-8 割程度であるため、セレクションバイアスの懸念が生じる。そこで、ここでは簡易的な検証として、工業統計調査に収録されている出荷額等の変数の平均を工業統計調査とマッチデータのサンプルで比較する。

表 4 は、全数調査である工業統計調査とマッチデータにおける各変数の平均値を比較する。変数のリストは以下である： y_{it} 産出（出荷額、加工賃収入額、修理料収入額の和）の実質対数値、 m_{it} 中間投入（原材料、燃料、電力使用額、委託生産費の和）の実質対数値、 l_{it} 従業者数の対数値、 k_{it} 恒久棚卸法で計算した有形固定資産の実質対数値、 I_{it}/K_{it-1} 設備投資比率、輸出比率は、製造品出荷額等に占める直接輸出額の割合。列「工業統計」、「マッチデータ（ウェイトなし）」、「マッチデータ（ウェイトあり）」はそれぞれ、全数調査である工業統計調査における各変数の平均、マッチデータにおける各変数の平均、マッチデータにおける調査ウェイトを用いた各変数の加重平均を示す。まず、「工業統計」と

Table4: 工業統計調査変数の平均

	(1) 工業統計	(2) マッチデータ (ウェイトなし)	(3) (2)-(1)	(4) マッチデータ (ウェイトあり)	(5) (4)-(1)
y_{it}	11.88	12.67	0.79	11.90	0.02
m_{it}	10.98	11.80	0.81	11.01	0.02
l_{it}	4.36	4.95	0.59	4.37	0.01
k_{it}	10.48	11.42	0.93	10.51	0.02
I_{it}/K_{it-1}	0.93	0.33	-0.60	0.46	-0.47
輸出比率	0.024	0.047	0.023	0.023	-0.001

注：2006-2014年のデータを使用。 y_{it} 、 m_{it} 、 l_{it} 、 k_{it} 、 I_{it}/K_{it-1} はそれぞれ、産出（出荷額、加工賃収入額、修理料収入額の和）の実質対数値、中間投入（原材料、燃料、電力使用額、委託生産費の和）の実質対数値、従業者数の対数値、恒久棚卸法で計算した資本ストックの実質対数値、そして、設備投資比率である。輸出比率は、製造品出荷額等に占める直接輸出額の割合である。列「工業統計」は工業統計調査における各変数の平均、「マッチデータ（ウェイトなし）」はマッチデータにおける各変数の平均、「マッチデータ（ウェイトあり）」はマッチデータにおける調査ウェイトを用いた各変数の加重平均を報告する。

「マッチデータ（ウェイトなし）」の比較では、各変数の平均値が大きく異なることがわかる。生産の産出と投入の各変数（ y_{it} 、 m_{it} 、 l_{it} 、 k_{it} ）についてはレベルで60%から80%ものちがいがあり、設備投資比率、輸出シェアについても大きな差がある。この平均の違いは、調査ウェイトを用いると大幅に減少することが、「マッチデータ（ウェイトあり）」との比較で明らかになった。産出、投入の各変数については1-2%の差におさまり、輸出比率の差も0.001まで小さくなることが確認された。^{*9} 一方、設備投資比率の差は以前として大きなままである。そこで、設備投資比率の外れ値を除去したサンプルで、この平均の差をさらに検証する。

表5は、設備投資比率 I_{it}/K_{it} の値が2を超えた場合にその観測値を除去するという処理をした時の各変数の平均値の比較である。第(3)列から明らかなように、設備投資比率の外れ値を除去すると、工業統計調査とマッチデータサンプルにおける各変数の平均値の違いはほとんどなくなる。したがって、マッチデータでは賃金センサスからさらにサンプルが落ちることによるバイアスが懸念されるが、少なくとも、ここで確認した平均の計算においては、依然として調査ウェイトは有効であり、母集団平均に近づけられることが示唆された。

^{*9} 同様に、常用労働者数の平均を賃金センサスとマッチデータでも比較した。対数をとった常用労働者数の調査ウェイトによる加重平均は、賃金センサスで4.31、マッチデータで4.27、その差は0.04となった。前述の通り、出向・派遣の受け入れ、送付がある場合は、賃金センサスの常勤労働者数は工業統計調査の従業者数($\exp(l_{it})$)と異なりうる。

Table5: 工業統計調査変数の平均（外れ値除去）

	(1) 工業統計	(2) マッチデータ (ウェイトあり)	(3) (2)-(1)
y_{it}	11.930	11.934	0.004
m_{it}	11.048	11.051	0.003
l_{it}	4.375	4.378	0.004
k_{it}	10.507	10.526	0.020
I_{it}/K_{it}	0.108	0.104	-0.004
輸出比率	0.025	0.024	-0.001

注：2006-2014年のデータを使用。さらに、 I_{it}/K_{it} の値が2を超えた場合、外れ値として除去。変数の定義は、表4の脚注を参照。

5 今後の課題

雇用者被用者マッチデータは、その豊富な従業員の属性情報により、事業所の生産性成長率に対する労働投入の「質」の貢献度を詳細に分析することができる。さらに、集計生産性成長率は、労働や資本投入の配分の効率性にも影響するが、雇用者被用者マッチデータでは、年齢、性別、教育水準など、詳細な属性グループごとに労働者の移動パターンの効率性を分析できる可能性がある。少子高齢化が進み、生産年齢人口が減少する日本では、特に、ミクロ、マクロレベルでの生産性改善が経済成長のために非常に重要であり、適切な政策設計のためにも雇用者被用者マッチデータを用いた分析の必要性が今後も高まることが予想される。本節では、公的統計を用いた雇用者被用者マッチデータの作成のための今後の課題について述べる。

まず、経済センサス活動調査個票と工業統計調査個票の間の接合率が低いため、サンプルサイズが減少し、バイアスの懸念が生じる。両調査が異なる母集団名簿を用いていることが接合率の低下の原因だと考えられるが、コンバーターの見直しや名簿情報の照合により、接合率の改善をはかることは、研究上非常に重要であると思われる。

次に、今回の個票データの二次利用申請では、平成13年事業所・企業統計調査以降でしか名簿情報を申請できなかった。平成13年事業所・企業統計調査は、2004年賃金センサスのサンプルフレームであるため、名簿情報による工業統計調査との接合は、2004年以降に限られる。対して、Fukao, et al. (2006)、川口、神林 (2010) は、1993年-2003年のデータを名簿情報の照合で接合しているため、以前はこの期間に対応する事業所・企業統計調査についても名簿情報が提供されていたということである。賃金センサスパネルに

2005年–2006年の間で断絶が起きているので、1993年から2014年の全期間をカバーするパネルデータの構築は難しいが、それでも、1993年から2003年の雇用者被用者マッチデータを追加的に使用できることのメリットは大きいと考える。また、今後に向けて、収集された調査票情報が、統計作成や研究のために適切に保持されることが重要であると考える。

本研究で接合データの構築が2014年までであったのは、2015年以降の賃金センサスのサンプルフレームである事業所母集団データベースの統計的研究目的の二次利用が認められていないためである。もし事業所母集団データベースの研究利用が認められ、賃金センサス個票と事業所名簿情報を接続することができれば、他の年と同様に工業統計調査と接合できる。さらに、近年、賃金センサスだけでなく、経済センサス活動調査、経済構造実態調査、そして企業活動基本統計調査も事業所母集団データベースを母集団名簿とするように整備されてきている。もし事業所母集団データベースの事業所ID（共通事業所コード）が各調査の個票情報に追加されれば、名簿情報の照合によらずに各調査の接合が可能になる。賃金センサスの各事業所の観測期間が短い問題は残るものの、共通事業所コードによる接合では高い接合率が期待できるため、研究利用が認められることの価値は非常に大きいと考える。

最後に、2024年度に、事業所・企業統計調査と経済センサス調査票情報の二次利用申請が認められなかった。総務省「調査票情報の提供に関するガイドライン」の改正により、統計的研究でのパネルデータ作成と他の調査との接合目的での調査票情報の提供ができなくなったということである。本研究でのデータ作成はあくまで中間生成物であり、個別事業所に着目するような分析結果は公表しないため、ガイドラインに反しないという解釈も可能なのではないかとも思うが、いずれにせよ、雇用者被用者マッチデータの政策分析における重要性も勘案し、調査票情報のオンサイト利用も含めて、二次利用ができる環境が整備されることを願いたい。

参考文献

- 阿部武司, 人見和也, 小西葉子, 富田秀昭, 内野泰助 (2012)、「工業統計調査のパネル化のためのコンバータ (1993年-2009年)」RIETI Discussion Paper Series 12-P-007, Research Institute of Economy, Trade and Industry
- 伊藤公二 (2017)、「日本における賃金の輸出プレミア : employer-employee data を利用した分析」RIETI Discussion Paper Series 17-J-050, Research Institute of Economy,

Trade and Industry

- 川口大司, 神林龍 (2010)、「政府統計の接合データの作成と利用—工業統計調査と賃金構造基本統計調査の例」北村行伸編著『応用ミクロ計量経済学』、pp. 131–160、日本評論社
- 行本雅 (2015)、「工業統計調査のパネル・データ整備の現状について」 KIER Discussion Paper Series 1506, Kyoto University
- Abowd, J. M., and F. Kramarz (1999). “The Analysis of Labor Markets Using Matched Employer-Employee Data,” *Handbook of Labor Economics*, vol. 3, pp. 2629–2710.
- Bonhomme, S., Lamadon, T., and Manresa, E. (2019). “A Distributional Framework for Matched Employer Employee Data,” *Econometrica*, vol. 87, no. 3, pp. 699–739.
- Fukao, K., R. Kambayashi, D. Kawaguchi, H. U. Kwon, Y. G. Kim, and I. Yokoyama (2006). “Deferred Compensation: Evidence from Employer-Employee Matched Data from Japan,” *Hi-Stat Discussion Paper Series*, 187
- Lentz, R. and D. T. Mortensen (2010). “Labor market models of worker and firm heterogeneity,” *Annual Review of Economics*, Vol. 2, No. Volume 2, 2010, pp.577–602.