



RIETI Policy Discussion Paper Series 12-P-020

生産動態統計調査と工業統計調査の 事業所マッチング法について(2005年-2009年)

小西 葉子
経済産業研究所



Research Institute of Economy, Trade & Industry, IAA

独立行政法人経済産業研究所

<http://www.rieti.go.jp/jp/>

生産動態統計調査と工業統計調査の事業所マッチング法について

(2005年－2009年) *

小西葉子[†] (RIETI)

要旨

本稿では生産動態統計調査と工業統計調査の事業所を名簿ベースで接続する方法を提案している。「生産動態統計－工業統計事業所マッチングコンバータ」は、生産動態統計と工業統計の個票（回答結果）を接続するツールである。事業所マッチング率は約89%で、そのうち両統計の事業所が1:1対応でマッチングされているものは、約97%であった。これにより、生産動態統計の詳細な生産情報に、工業統計調査の豊富な事業所の生産活動情報を補完でき、生産性分析などの経済分析が可能となるであろう。ただし、実証研究する際には、生産動態統計調査が月次・品目ベース、工業統計調査が年次・事業所ベースであるので回答結果の接続には工夫が必要である。

RIETI ポリシー・ディスカッション・シリーズは、RIETI の研究に関連して作成され、政策をめぐる議論にタイムリーに貢献することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、(独) 経済産業研究所としての見解を示すものではありません。

* 本稿は、RIETI プロジェクト『経済変動の需要要因と供給要因への分解：理論と実証分析』の研究成果“Decomposition of Supply and Demand Shock in Production Function Using Current Survey of Production”の遂行のために必要なツールとして考案したものである。その整理、プログラム化の過程でデータ整備プロジェクトのメンバーからアイデア提供や協力を得た。記して感謝したい。また多忙にも関わらず常に丁寧に、細かな疑問にも速やかに対応してくださった、経済産業省大臣官房調査統計グループの大山靖江氏、木下善雄氏、齋藤敬氏、諸岡英行氏に感謝の意を示したい。また新井勝美氏、秦邦雄氏、金子喜則氏には初稿へ有益なコメントを頂いた。なお、本稿の誤謬は全て筆者の責任に帰すものである。

[†] konishi-yoko@rieti.go.jp

1. はじめに

近年、マイクロデータへのアクセス可能性が高まり、計算機やソフトウェアの開発も進み、多くのマイクロデータ解析が行われている。しかし現状でも、大規模なマイクロデータを入手した際には、分析可能な形にクリーニングし、形成するのに多大な時間を要する。さらに、実証研究では複数統計を使用することが多いので、マッチングの作業も必要となってくる。そこで RIETI では、研究者が少しでも効率よく正確な実証分析を行えるよう、経済産業省から提供される統計の個票データのうち、申請頻度が高いものにはデータセット構築のための知識蓄積・共有を行ってきた¹。

生産動態統計調査は、経済産業省がわが国の鉱工業に関する動態を明らかにすることを目的とし、鉱産物及び工業品を生産する事業所に対して毎月調査を行っている。月次調査で速報性があり、1800 にも及ぶ品目が含まれるため、個別産業（品目）の業況判断や需給動向の判断材料に用いられ、産業振興や中小企業対策など多くの政策評価に役立てられている。調査対象となる業種は、生産品目ごとに、①鉄鋼・非鉄金属・金属製品統計、②化学工業統計、③機械統計、④窯業・建材統計、⑤繊維・生活用品統計、⑥紙・印刷・プラスチック・ゴム製品統計、⑦資源エネルギー統計に分類されている。

しかし、RIETI においては、他の経済産業省所管の統計調査と比較して研究者の利用が少ない現状がある。考えられる理由は、第一に調査票が 111 種類存在し、それらの質問項目や単位が調査票ごとに異なり、さらに調査対象となる品目や項目が産業の実態を反映するために頻繁に変更される。そのため特定の調査票に絞る場合を除いて、データハンドリングは研究者にとって非常に難易度が高いことが挙げられる。

第二に、調査の目的上、経済分析で多用される資本（土地、機械、設備）などのデータが含まれていないため、企業や事業所活動に関する分析に、生産動態統計調査単体での応用が難しい点がある

本稿では、この第二のハードルを解決するために、「生産動態統計－工業統計事業所マッチングコンバータ²」を作成する。このコンバータを用いて実証分析を行う際には、いくつかの工夫は必要であるが³、生産動態統計の詳細な生産情報に、工業統計調査の豊富な事業所の生産活動に関する情報を補完することにより、様々な経済分析（特に生産性分析）が可能となるであろう。

¹ RIETI ウェブサイトのマイクロデータ計量分析プロジェクトのページに入ると、各種個票のデータ構築に役立つ情報が掲載してある。具体的には、工業統計調査、企業活動基本調査、海外事業活動基本調査などの事業所や企業を時系列方向に接続してパネルデータを作成する方法や、各調査をマッチングして、事業所と企業が結びついた包括的なデータ構築を行うためのコンバータ作成方法の開発が紹介されている。

² このコンバータは、両統計の対応表（後掲表 6）を作成するためのテキスト形式のプログラムなので、秘匿情報やデータの流出を防ぐこともできる。

³ マッチング後の実証分析の際に注意するべき点は 5 章で述べるが、本稿は事業所マッチングが目的のため詳細は議論しない。詳細は Konishi and Nishiyama(2012)参照。

具体的には、総務省の府省共通事業所コードと両統計の企業名、事業所住所、郵便番号、事業所名による名寄せでマッチングを行った。工業統計では市区町村番号や事業所番号が年により変化するため、阿部他(2012)で作成された「工業統計パネル化のためのコンバータ」をマッチングし、年を通じてユニークな工業統計の事業所 ID を付与している。

図 1 生動-工業マッチングコンバータ作成フローチャート

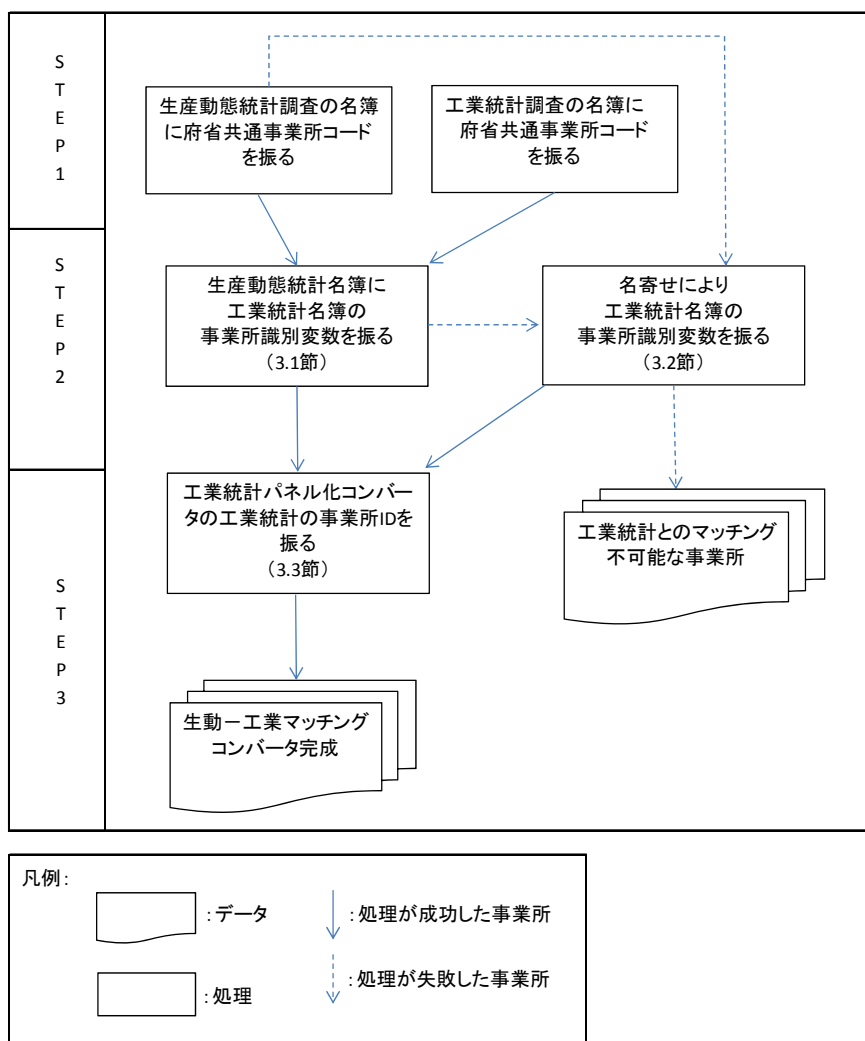


図 1 は作業のフローチャートである。マッチング結果は、マッチング率は 88.71%で、そのうち両統計の事業所が 1:1 対応で接続できているものが約 97%となり、非常に高い率で両統計の事業所を接続することができた。

本稿では、作業の全自動化はできなかったが、高いマッチング率が実現でき、コンバータ作成の方針を立てることができた。この新しいコンバータを共有知識とし、多くの研究者が両統計を用いたより良いマッチング法の開発や実証研究を行うことを期待する。次章では生産動態統計調査と工業統計調査について、マッチングの観点から概観した。

3章では、具体的な作成手順について説明している。4章では、本稿のマッチング作業の評価を行い、5章はまとめと実証分析を行う際の留意点について述べている。

2 生産動態統計調査と工業統計調査についての概観

本章では、生産動態統計調査と工業統計調査について、マッチングのためのコンバータ作成の観点から概観し、各調査の比較を行う。

2. 1 生産動態統計調査について

生産動態統計調査⁴は、経済産業省がわが国の鉱工業⁵に関する動態を明らかにすることを目的とし、鉱産物及び工業品を生産する事業所⁶に対して毎月調査を行っており、第一回目の調査は昭和23年（1948年）に遡る。調査結果は、鉱工業指数、四半期GDP速報、産業連関表作成のための基礎資料として用いられる。調査対象となる業種は、生産品目ごとに、①鉄鋼・非鉄金属・金属製品統計、②化学工業統計、③機械統計、④窯業・建材統計、⑤繊維・生活用品統計、⑥紙・印刷・プラスチック・ゴム製品統計、⑦資源エネルギー統計に分類されている。2012年現在で、これら7つの分類で調査票は計111種類あり⁷、調査品目数は約1800にも及ぶ。

調査対象は、業種ごとに従業員数による裾切りを行っているが、一定規模以上については全調査対象を行っている⁸。調査項目は、生産高、出荷高、在庫高が共通の基本項目⁹となり、原材料、燃料及び電力、労務情報、生産能力・設備に関する項目は品目ごとに必要に応じて調査票に含まれる（表1参照）。個々の事業所には、一意な事業所番号が振られており、それにより事業所を時系列方向につなぐことができる。

以下で、研究の側面からみた3つの特徴について述べる。まず、産動態統計調査は、生産高、出荷高、在庫高が金額ベースのデータのみならず、数量や重量といった質量データにより報告されるのが特徴である。これは、生産性計測で懸念される価格情報が含まれる¹⁰ことによる生産性の上方バイアス推定を部分的に回避することが期待できる。次に、集計結果が翌月末までに公表される速報性を持つ。さらに、非常にユニークな点が、

⁴ 経済産業所ホームページ参照。

<http://www.meti.go.jp/statistics/tyo/seidou/index.html>

⁵ 日本標準産業分類大分類の鉱業と製造業に該当。

⁶ 生産品目の販売管理を行っている事業所、または生産委託を行っている特定事業所も含まれる。

⁷ <http://www.meti.go.jp/statistics/tyo/seidou/gaiyo.html#menu07>で公開されている。

⁸ 対象事業所は約2万事業所で、回収率は約94%である。また複数調査に回答している事業所もある。本稿では、2005年から2010年の事業所を対象とする。

⁹ 一部の調査票では、生産高のみのものである。

¹⁰ 例えば、売上額＝価格×生産量の形になっており、価格を通じて需要要因が含まれることが知られている。

一部の調査票に含まれる「生産能力・設備」という項目である。「生産能力」とは、需要変動に影響されずに標準的な装置と従業者を投入して得られる、各事業所の生産可能量である。また「設備」では、例えば⑤繊維・生活用品の紡績や織物製品で、生産機械の月末保有台数と月間平均実動台数や操業時間といった稼働率に関わる調査している。標準的な生産能力、稼働率、在庫高と売上高や出荷額との乖離を数量的に計測できることは、より正確な需給ギャップを教えてくれるだろう。Konishi and Nishiyama (2012)はこの点に着目し、経済変動における供給要因と需要要因を分解し、従来に計測された生産性から需要要因を取り出すことを目的としている。生産動態統計調査は、速報性があり、1800にも及ぶ品目が含まれるため、個別産業（品目）の業況判断や需給動向の判断材料に用いられ、産業振興や中小企業対策など多くの政策評価に役立てられている。一方、RIETIにおいては、他の経済産業省所管の統計調査と比較して研究者の利用が少ない現状がある。考えられる理由は、第一に、調査票が111種類存在し、それらの質問項目や単位が調査票ごとに異なり、さらに産業の実態を反映するために、調査対象となる品目や項目が頻繁に変更される。そのため特定の調査票に絞る場合を除いて、研究者にとって非常に難易度の高いデータハンドリングとなる。第二に、調査の性質上、経済分析に多用される資本（土地、機械、設備）などのデータが含まれないため、生産動態統計調査単体での、企業や事業所活動についての実証分析への応用が難しい点がある。表1は各調査票の概要である。1.製品の項目のみ全ての品目について調査されており、原材料は品目合計と、各品目をいくつかのグループに集計したもの、月末従事者数は品目合計と事業所全体、生産能力は一部の品目について調査している。全ての項目が各品目に対して調べられているわけではない。1製品と3労務は全ての調査票、4生産能力は6割程度、2原材料はごく少数の調査票で項目として取り上げられている。

表1 生産動態統計調査の調査項目について

項目番号	項目分類	項目名	対象品目	対象調査票
1	製品	生産	全ての品目	全ての調査票
		受け入れ		一部(少数)を除く全ての調査票
		消費		
		出荷		
		月末在庫		
2	原材料	各原材料	品目合計と集計された品目	少数調査票
3	労務	月末従事者数	品目合計と事業所全体	全ての調査票
4	生産能力	生産能力	一部品目	6割程度

2. 2 工業統計調査と阿部他(2012)のパネル化コンバータについて

工業統計調査¹¹は経済産業省がわが国の製造業事業所を対象に、工業の実態を明らかに

¹¹ 経済産業省ホームページ参照

<http://www.meti.go.jp/statistics/tyo/kougyo/index.html>

するために行う年次の統計調査である。調査年の1の位が0, 3, 5, 8の年に全事業所の調査（全数調査）を行い、それ以外の年では、従業員数4人以上の事業所を対象とした調査（裾切り調査）となっている。調査票は従業員数30人以上を対象とする甲票、29人以下を対象の乙票があり、両者の大きな違いは質問項目数にある。付表1に2009年の両調査票の項目がまとめてある。甲票では62、乙票では31の項目がある¹²。

工業統計調査では、各事業所は、都道府県、工業統計調査市区町村、事業所番号の3つの情報により各年の中で識別される。ただし、生産動態統計調査と異なり、事業所番号は5年に一度振りなおされる（最近では1991、1997、2002年に実施されている）。それ以外でも市区町村の統廃合などにより市区町村コードの変更があった場合も事業所番号の変更は行われる。これらの事業所番号の変更情報は「提供用コンバータ」として経済産業省より提供される。

つまり、少なくとも5年に一度は事業所番号が変更されるため、時系列方向に事業所を接続するためには、コード変更の変遷情報を含むコンバータが必要となる。阿部他(2012)¹³は、準備調査名簿の情報を用いて、パネル化のためのコンバータ¹⁴作成法を新たに提案した。それにより、全期間について各事業所に一意なIDが与えられ、研究者がデータを受け取ってすぐにパネル分析ができる環境を提供している。このコンバータでは、期間中1度でも調査対象となった事業所が対象となるため、回答の有無や参入退出行動を全てフラグで記録した大きなデータセットが構築¹⁵できている。RIETIから工業統計調査を目的外申請した場合には、手続きを行えば利用可能となる。

2. 3 生産動態統計と工業統計の比較：マッチングに必要な情報を中心に

本稿では、生産動態統計調査と工業統計調査の事業所をマッチングすることを目的としている。表2にあるように、生産動態統計調査が鉱業と製造業に属する品目を生産している事業所、工業統計調査が主な生産物が製造業に属する事業所を対象にし、工業統計調査が、1の位が0, 3, 5, 8の年に全事業所の調査（全数調査）を行っているので、生産動態統計調査の事業所はほぼ全部工業統計調査にマッチングできると考えられる¹⁶。2.1で述べたように、生産動態統計調査はわが国の鉱工業の産業の実態を明らかにする

¹² 平成22年の調査票は <http://www.meti.go.jp/statistics/tyo/kougyo/gaiyo/chosahyo/h22.html> を参照（HPでは平成12年から22年までの調査票が入手可能）。

¹³ <http://www.rieti.go.jp/jp/publications/summary/12040007.html> でダウンロード可能。

¹⁴ 時系列方向だけでなく、本稿の様に異なる統計を個体認識可能なIDで接続するためのものもコンバータと呼ぶ。

¹⁵ 大きなデータセット構築については、http://www.rieti.go.jp/jp/columns/a01_0348.html を参照。

¹⁶ 鉱業のみ行っている事業所は、工業統計調査に含まれないためマッチングできないが、通常事業所は多種多様な製品を生産しており、鉱業品を加工すればその製品の多くは製造業に属するのでマッチングできないものは少ないであろう。

ことを目的としているため、生産現場や市場動向を反映して、調査票の品目や項目が変更される。例えば、2005年には「機械統計」の民生用電子機器では、テレビの品目はプラズマテレビ、液晶テレビ、その他のカラーテレビであったが、2011年には、テレビというカテゴリ名が薄型テレビに変更され、品目も40型以上、40型未満となっている。一方、調査票自体も新しく加えられたり、名称が変更したり、削除されたりしている。近年では2002年に大幅改訂が行われた。その後2003年から2005年に僅かな改訂が行われ、2006年から2010年までは調査票の構成自体は変更がなかった¹⁷。そのため、本稿では、2005年1月から2010年12月の期間の全事業所を対象としている。工業統計調査は、2005年から2009年を使用する。工業統計調査は、各年で約55万から約61万の事業所が含まれる。

調査項目については、生産動態統計調査は表1にあり、全て調査された場合でも8個程度である。一方年次調査である工業統計調査は付表1にあるように乙票で31個、甲票で62個となっている。これが生産動態統計調査に工業統計調査をマッチングさせる一番大きな動機である。事業所識別は、生産動態統計調査には各事業所に常に一意な事業所番号が振られているので、各年月をつなげてパネルデータにする際には事業所番号をキー変数にすればよい。一方、工業統計調査は都道府県、市区町村、事業所番号の3つをキーにして識別する。しかし2.2で述べたように事業所番号と市区町村番号は変更されるので、パネル化のためのコンバータを利用して、事業所に時間を通じて一意なIDを振る必要がある。

表2 両統計調査の比較

統計名	産業	対象	調査法	個体識別	パネル化
生産動態統計調査	鉱業、製造業	該当品目を生産している事業所	標本調査	事業所番号(常に一意的)	コンバータ必要なし
工業統計調査	製造業	主な生産物が製造業に属する事業所	全数調査あり(1の位が0, 3, 5, 8の年)	都道府県、市区町村、事業所番号(様々な理由で変更される)	コンバータ必要あり RIETIで提供

3. マッチング方法

本稿では、3.1で総務省から付与される府省共通事業所コードでマッチングし、3.2で企業名を軸に事業所住所、郵便番号、事業所名を組合せて名寄せによるマッチングを行う。さらに3.3で阿部他(2012)の「工業統計パネル化のためのコンバータ」と接続し、

¹⁷ 平成9年以降の調査票の変遷が調べられる。調査票が変更されなくても、調査票内の品目などは、毎年適宜変更されている。

http://www.meti.go.jp/statistics/tyo/seidou/chosahinmoku_hensen.html

「生産動態統計調査－工業統計事業所マッチングコンバータ」を作成する。各節の作業の流れを表3で示す。灰色で網掛けしてある部分はエクセルで手作業を行い、網掛けしていない自動化処理は STATA で行っている。

表3 作業の流れ

節番号	手順番号	処理内容	作業/作業doファイル名
3.1	1	ファイル名変更	手作業
	2-3	名簿変数名統一(生動)	SK_meibo_1.do
	4-5	名簿変数名統一(工業)	SK_meibo_2.do
	6-9	府省共通事業所コード接続	SK_meibo_3.do
	10	府省共通事業所コードによるマッチング	SK_matching_1.do
3.2	1	名簿のクリーニング	手作業
	2-7	名寄せによるマッチング	SK_matching_2.do
	8-9	名寄せによるマッチング	SK_matching_3.do
	10	郵便番号修正	手作業
	11	名寄せによるマッチング	SK_matching_4.do
	12-15	名寄せによるマッチング	SK_matching_5.do
	16-19	名寄せによるマッチング	SK_matching_6.do
	19	事業所住所修正	手作業
	20	企業名修正	手作業
	21-23	名寄せによるマッチング	SK_matching_7.do
	24-26	名寄せによるマッチング	SK_matching_8.do
27	目視による名寄せ	手作業	
3.3	1	ファイルの結合と変数名統一	SK_matching_9.do
	2	パネル化コンバータとマッチング	SK_matching_10.do

ここではまずマッチングに必要な生産動態統計調査の生動1：提供用名簿、生動2：事業所対応表、工業1：準備調査名簿、工業2：工業・事業所コード、工業3：統一産業分類付き Wide 形式コンバータの紹介をする。

生動1：提供用名簿

生産動態統計調査では事業所に対してユニークな ID を付与していて、かつ経時的に変化しないため、期間中の事業所をカバーしている名簿が一つ提供される。おそらく企業名、事業所名、所在地は申請年次もしくは調査年次の最新年（本稿の場合は 2010 年）のものであると思われる¹⁸。企業名、企業住所、事業所名、事業所番号、事業所住所などが含まれる。名簿、個票とも変数名が入っていないので研究者が変数名をつける必要がある。

¹⁸ 工業統計調査と同様に、転出や廃業、当該品目を生産しなくなった場合調査対象から削除されることになっているが、それらをどのように名簿で取り扱うのか（一定の年数経過後に削除するのかなど）詳細は本稿では調べていない。

生動 2：事業所対応表

総務省では、同一事業所・企業に対して各省庁の統計調査が過度に集中して実施されることのないよう、母集団となるデータベースを作成し、各事業所・企業に 9 桁の府省共通事業所コードを振っている。この府省共通事業所コードと生産動態統計の事業所番号の対応表である。本稿では、2011 年の名簿に対して付与された対応表を利用している。

工業 1：準備調査名簿

準備調査名簿の電子版利用は、1993 年以降可能である。準備調査名簿には、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にある全ての調査対象事業所が含まれている。事業所を識別する市町村番号と事業所番号が年によって変化するので、名簿は毎年提供される。調査年度、都道府県番号、市区町村番号、産業分類、企業名、事業所名、事業所郵便番号、事業所住所などが含まれる。

工業 2：工業・事業所コード

工業統計調査の事業所に生動 2 と同様の府省共通事業所コードを付与したものの。準備調査名簿と同じく毎年存在する。本稿では、2006 年、2007 年、2008 年分が利用可能である。都道府県番号、市区町村番号、事業所番号、総務省共通事業所コードが含まれる。

工業 3：統一産業分類付き Wide 形式パネル化コンバータ（阿部他、2012 年より作成）
各年について両統計をマッチングした後、工業統計調査の事業所に時系列を通じてユニークな ID（表 4 の項番 1 の id）を振るために使用する。表 4 はパネル化コンバータに含まれる変数一覧である。

表 4 統一産業分類付き Wide 形式パネル化コンバータに含まれる変数
(阿部他、2012、p11 より抜粋)

項番	変数名	内容	備考
1	id	事業所ごとのユニークな番号	キー項目
2	state	都道府県番号	
3	cityYYYY	市区町村番号（YYYY は西暦年）	対象年の数分存在
4	code_estYYYY	事業所番号（YYYY は西暦年）	
5	i_codeYYYY	産業分類（YYYY は西暦年）	
6	kbnYYYY	甲乙区分（YYYY は西暦年）	
7	ans_flgYYYY	回答フラグ（YYYY は西暦年）	
8	i_codeUniYYY	統一産業分類（YYYY は西暦年）	

3. 1 府省共通事業所コードによるマッチング (STEP1)

総務省から提供される府省共通事業所番号を利用して、生産動態統計調査と工業統計調査のマッチングを行う。以下では、図1のフローチャートに従って具体的な作業方法を説明する。府省共通事業所コードとは、国の機関が行う基幹統計調査及び一般統計調査において、同一事業所・企業に対して統計調査が過度に集中しないよう、重複是正のために事業所・企業に割り当てられる9桁のコードである。

1. ファイル名の整理を手作業で行う。

STATAで読み込むため、ファイル名を変更する。生動1：提供用名簿は、「提供用名簿.csv」を“seido_meibo.csv”に変更する。生動2：事業所対応表は「事業所対応.xlsx」を“作成年_seido_somu.csv”（例：2011_seido_somu.csv）とする。工業1：準備調査名簿はファイル名を“調査年_meibo.csv”（例：2005_meibo.csv）に統一する。工業2：工業・事業所コードはファイル名を“調査年_kogyo_somu.csv”（例：2006_kogyo_somu.csv）に統一する。

2. 生動1：提供用名簿の整理には、(SK_meibo_1.do)を実行する。

提供された名簿には変数名が入っていないので、変数名をつける。また、事業所番号、企業名、企業住所、事業所名、事業所住所以外の項目を削除する。手順2-5で変更した変数名はAppendix2に新旧変数一覧がある。

3. 生動2：事業所対応表の整理には、(SK_meibo_1.do)を実行する。変数は生産動態統計調査の事業所番号と総務省の府省共通事業所コードの2変数が入っているので、それぞれ“seido_code”、“somu_code”に変更する。

4. 工業1：準備調査名簿の整理には、(SK_meibo_2.do)を実行する。

年によって変数名が異なる場合があるので、変数名を統一する。また準備調査名簿に都道府県番号、市区町村番号、事業所番号が同じデータが複数存在する場合がある。同じ事業所に対するコンバータが複数件できることを避けるため、重複データを削除する。都道府県番号、市区町村番号、事業所番号、企業名、事業所名、事業所郵便番号地域番号（上3桁）、事業所郵便番号枝番（下4桁）、事業所都道府県名、事業所市区郡名、事業所町村名、事業所番地以外の項目を削除する。

5. 工業2：工業・事業所コードの整理には、(SK_meibo_2.do)を実行する。

各年の都道府県番号、市区町村番号、事業所番号をそれぞれ“state”、“city”、“code_est”に変更する。府省共通事業所コードは、“somu_code”にする。

6. SK_meibo_3.do で、生産動態統計調査の事業所番号をキーに、生動 1：提供用名簿と生動 2：事業所対応表をマージして、提供用名簿に府省共通事業所番号を振る。

7.提供用名簿の重複データの削除 (SK_meibo_3.do) を行う。事業所番号は異なるが、同一の府省共通事業所コードが振られている場合がある。重複している理由も¹⁹、どちらの事業所コードを採用するのも判断すべきではないので、府省共通事業所コードが重複している事業所は全て削除する。削除されたものは、3.2 節の名寄せによるマッチングを行う。

8. SK_meibo_3.do で、工業統計調査の都道府県番号、市区町村番号、事業所番号をキーに、工業 1：準備調査名簿と工業 2：工業・事業所コードをマージして、準備調査名簿に府省共通事業所コードを振る。

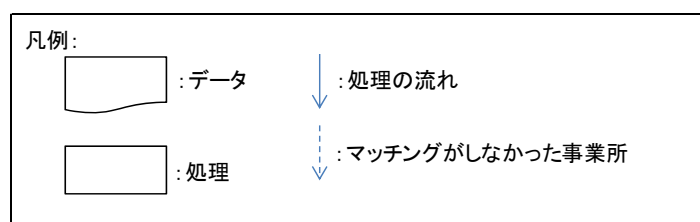
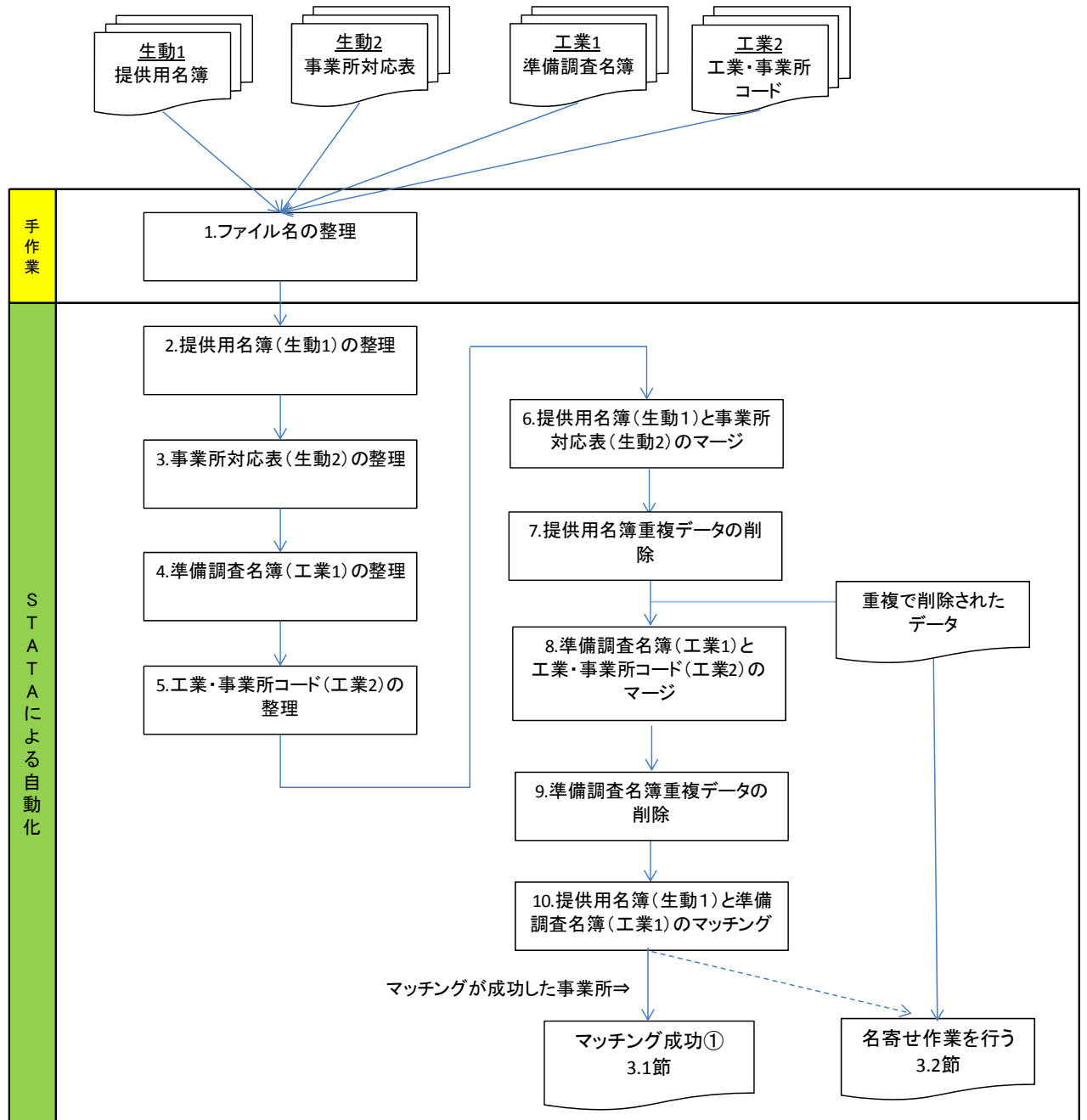
9.準備調査名簿の重複データの削除 (SK_meibo_3.do) を行う。手順 7 と同様に、同一の府省共通事業所コードが異なる都道府県番号、市区町村番号、事業所番号に振られている場合がある。手順 7 と同様の理由で府省共通事業所コードが重複している事業所は全て削除する。

10.生産動態統計調査の提供用名簿と工業統計調査の準備調査名簿を総務省の共通事業所番号をキーにしてマッチングする (SK_matching_1.do)。マッチングできた事業所を保存する。マッチングできなかった事業所は手順 7 のデータと合わせて 3.2 節の名寄せによるマッチングの対象とする。

(マッチング成功①)

¹⁹ 府省共通事業所コードについて問い合わせたところ、正確性が低い可能性があるという回答があった。実際、企業名、住所、事業所名、事業所住所が異なるのに同一の府省共通事業所コードが振られているものがあって、重複があるものは、府省共通事業所コードでのマッチングが不可能と考え、名寄せによるマッチングの対象とした。

図2 府省共通事業所コードによるマッチング法のフローチャート



3. 2 名寄せによるマッチング法

府省共通事業所コードが重複していたために3.1のマッチング候補から削除された事業所と3.1でマッチングできなかった事業所について、名寄せによるマッチングを行う。作業フローチャートは図2で示す。本稿の名寄せでは企業名を軸にして、事業所住所、郵便番号、事業所名を組合せてマッチングする。3.2.1で、名寄せをする際に留意しておくべき問題点などを説明し、3.2.2で手順を説明する。

3. 2. 1 名寄せの際の留意点

前小節で総務省の府省共通事業所コードによるマッチング法の説明を行ったが、その際異なる事業所に同一の府省共通事業所コードが振られている事業所は削除した。この削除された事業所とマッチングすることができなかった事業所を合わせて、名寄せ作業の対象とする。本稿では企業名を軸に、①事業所住所、②事業所郵便番号、③事業所名の3パターンの名寄せを行う。表5は回答事業所や記入者のミスを除いた、それぞれの変数の名寄せに関する起こりうる問題点についてまとめている。

名寄せは完全一致でマッチングするので、僅かな表記の違いも許さない。そのため、表記法の違いについては名簿に対して全角に統一、ひらがなをカタカナに変更、(株)などの略式記号を使用しないなどのクリーニングを行う。

名称変更や住所変更により名寄せできない場合の多くは、2つの統計の調査時期や企業情報の記入時期が異なることによって起こる。例えば事業所の業務も住所も変更がないが、合併や事業拡大によって企業名称のみ変化した場合や、同一市区町村内に引っ越した場合²⁰である。これらの場合、統計調査上は同一事業所とみなされるので、事業所コードの変更はないが、本稿のマッチングではつなげることができない。

しかし、この点については3.3節で、工業統計のパネル化コンバータに接続する際にある程度の解決ができる。実際は期間中存続していても、途中で企業名が変わってしまい名寄せではマッチングできなかったとしても、1年分でも都道府県番号、市区町村番号、事業所番号が接続できていれば、ユニークなidを振ることができるからである。

郵便番号は、工業統計は名簿に記載されているが、本稿では生産動態統計では申請しなかった。そこで事業所住所を基にエクセルのアドインの郵便番号変換ウィザードによって郵便番号を得た。これは最新の情報で構成されているため、それ以前に改訂があった郵便番号とはマッチングできないという問題がある²¹。また、大口事業所は個別郵便番号を有している場合があり、住所から変換した郵便番号とはマッチングできない。これ

²⁰ 都道府県外や同一都道府県内の他の市区町村への引っ越しは、当該事業所が廃業し、新たな事業所が開業したとみなされる。

²¹ 生産動態統計の名簿情報の中にも郵便番号は存在するので、申請することを強く勧める。これによりマッチング結果の向上が多いに期待できる。

らは、目視による修正が必要である。事業所名については、欠損値となっているものが多い。

以上より、名寄せ作業に関わる主な問題は、記入ミス、表記のゆれ、統計調査の記入年のずれが挙げられる。また上記の問題からも、企業名との組み合わせの優先順位を①事業所住所、②事業所郵便番号、③事業所名とする。

表 5 名寄せを困難にする原因（記入ミスや一部記入を除く）

変数名	問題点	備考
企業名	・表記法の違い	(株)、株式会社や ABC をエービーシー、カタカナ、ひらがな表記の違いなど
	・名称変更	合併や事業拡大
事業所住所	・表記法の違い	番地の書き方や旧呼称
	・住所変更	市区町村合併などによる地名の変更 同一市区町村内での引っ越し
事業所郵便番号	・郵便番号変更	市町村合併などによる変更 同一市区町村内での引っ越し
	・個別番号	大口事業所の場合。住所での検索ができない。
事業所名	・表記法の違い	企業名と同じ
	・名称変更	企業名と同じ
	・欠損値	特に工業統計で多い

3. 2. 2 名寄せによるマッチング法の手順（STEP2）

生産動態統計の事業所番号は個体にユニークに振られており、工業統計の都道府県番号、市区町村番号、事業所番号の 3 つの変数の組合せも、各年の中ではユニークである。総務省の府省共通事業所コードはこれらを直接マッチングできるキー変数であるが、全て接続できるわけではないので、企業名を軸にして①事業所住所、②郵便番号、③事業所名で名寄せしていく。

しかし、各統計の事業所識別変数と名寄せに使用する変数が異なるため、名寄せに使用する変数が全て同じでも、異なる事業所の場合もある。また、工業統計で①-③の名寄せの組合せが重複している場合、どの事業所と紐付けるのか判断できないため、予めマッチングの対象としていない。

本稿では、生産動態統計に工業統計の各年の個体識別 3 変数を振る。よってマッチングが成功している結果は、名寄せの組合せ条件の下で生産動態統計と工業統計が 1:1 対応、

m:1(m>1)対応の場合のみである²²。作業の途中で、複数対複数(以下 STATA のマージコマンドにならって m:m と表記)のマッチングをすることもあるが、最終的には目視で 1:1 か m:1 にマッチングしている。以下では、図 3 の作業フローチャートに沿って手順の説明をしていく。

1. まず準備として、府省共通事業所コードでのマッチングができなかった生産動態統計の事業所名簿について、事業所住所からエクセルの郵便番号変換ウィザードを使用して郵便番号を振る。

工業統計調査の準備名簿は全事業所を対象とし、工業統計調査の郵便番号を"000-0000"の形式に変換し、4つに分かれている住所を連結する。

次に表 5 にある表記のゆれを修正するためにクリーニングする。両名簿に対して、企業名、事業所名、事業所住所について、①半角文字を全角文字に変換、②スペースの削除、③中点の削除、④ひらがなをカタカナに変換する。さらに企業名については「有)、(有)」を有限会社に、「(株)、(株)」を株式会社に、「(資)、(合資)」を合資会社に、(協)を協同組合に、(企)を企業組合に、(名)を合名会社に、(業)を協業組合にそれぞれ変換する。これらの作業はエクセルで行った。

2. 両統計の名簿で企業名と事業所住所が重複する事業所を全て削除する。削除した事業所は手順 3 でマッチングしなかった事業所と合わせて手順 4 の処理の対象となる。手順 2-7 は、STATA で自動化処理され、名寄せ作業 A とする(SK_matching_2.do)。

3. 企業名と事業所住所によるマッチングを行う。マッチングしなかった事業所は手順 4 の処理の対象となる。

(マッチング成功②)

4. 両統計の名簿で企業名と郵便番号が重複する事業所と郵便番号が欠損値の事業所を全て削除する。削除した事業所は手順 5 でマッチングしなかった事業所と合わせて手順 6 の処理の対象となる。

5. 企業名と事業所住所によるマッチングを行う。マッチングしなかった事業所は手順 7 の処理の対象となる。

(マッチング成功③)

6. 両統計の名簿で企業名と事業所名が重複する事業所と事業所名が欠損値の事業所を全て削除する。削除した事業所は手順 7 でマッチングしなかった事業所と合わせて手順

²² マッチング成功結果の約 97%が 1:1 で対応している。

8 の処理の対象となる。

7. 企業名と事業所名によるマッチングを行う。マッチングしなかった事業所は手順 8 の処理の対象となる。なお、マッチング結果は両統計名簿の事業所住所で確認することを薦める。

(マッチング成功④)

8. 生産動態統計の名簿で企業名、事業所住所、事業所名の 3 つが重複する事業所は、名寄せパターン A において常に削除される。これらの事業所を(SK_matching_3.do)で抜き出す。非対象の事業所は手順 10 の対象となる。手順 8、9 は(SK_matching_3.do)で処理される。

9. 重複している事業所について「`duplicates drop f_name (企業名) address_e (事業所住所) e_name (事業所名), force`」で重複事業所のうち 1 事業所だけ残して削除する。削除されなかった事業所に対して、名寄せ作業 A を実行する。マッチングしなかった事業所は手順 10 の対象となる。マッチングした事業所に振られた都道府県番号、市区町村番号、事業所番号を企業名、同じ企業名、事業所住所、事業所名を持つ重複事業所に振る。この作業では異なる生産動態統計の事業所に同一の工業統計の都道府県番号、市区町村番号、事業所番号が接続される、m:1 マッチングである。

(マッチング成功⑤)

10. 生産動態統計の郵便番号はエクセルのウィザードで得たが、郵便番号が振られず欠損値になっているものがある。それらに対して手作業で郵便番号入力を行う。

11. 郵便番号の欠損値が修正された事業所に対し、名寄せ作業 A を(SK_matching_4.do)で実行する。マッチングできなかった事業所は手順 12 の対象となる。

(マッチング成功⑥)

12. 工業統計の企業名を「`duplicates drop f_name (企業名), force`」で企業名が重複している事業所のうち 1 事業所のみ残して削除し、企業名でマッチングをする。工業統計に同一の企業名が無く、マッチングしなかった事業所は手順 20 の対象となる。手順 12-15 は(SK_matching_5.do)で処理する。

13. 異なる事業所で同一の企業名、事業所住所を持つ事業所に対して、企業名と事業所住所の重複を全て削除した工業統計名簿をマッチングする(m:1 のマッチング)。マッチングしなかった事業所は手順 14 の対象となる。

(マッチング成功⑦)

14. 異なる事業所で同一の企業名、郵便番号を持つ事業所に対して、企業名と郵便番号の重複を全て削除した工業統計名簿をマッチングする(m:1のマッチング)。マッチングしなかった事業所は手順15の対象となる。

(マッチング成功⑧)

15. 異なる事業所で同一の企業名、事業所名を持つ事業所に対して、企業名と事業所名の重複を全て削除した工業統計名簿をマッチングする(m:1のマッチング)。マッチングしなかった事業所は手順16の対象となる。

(マッチング成功⑨)

16. 事業所住所の入力ミスや表記違いを修正するために、事業所の所在都道府県の入力ミスは無いものと仮定して、企業名と事業所の都道府県名でマッチングする。ここでマッチングしなかった事業所については、本稿のアプローチではこれ以上のマッチングは不可能と判断する。手順16-19は(SK_matching_6.do)で実行し、名寄せ作業Bと呼ぶ。

17. 工業統計の企業名と事業所住所の重複を全て削除して、企業名と事業所都道府県名でマッチングする。マッチングできなかった事業所は、工業統計の名簿で企業名と事業所住所に重複がある事業所でこれ以上のマッチングは不可能である。

18. マッチング結果で、生産動態統計の事業所番号が複数存在している事業所については、一つの生産動態統計の事業所に対して複数の工業統計の事業所が対応している。どの事業所とマッチングすればよいかの判断はできないのでマッチング不可能とする。

19. 手動で事業所住所のチェックを行いマッチングする。生産動態と工業統計の企業名と事業所の対応として、1:1、m:1、m:mのものが含まれている。特にm:m対応になっているものを目視でマッチングするときには企業情報を企業HPや資料で調べ慎重に行うこと。マッチングできなかったものはマッチング不可能とする。

(マッチング成功⑩)

20. 手順12で工業統計の企業名と生産動態名簿の企業名をマッチングしてマッチングしなかった事業所を対象に手動で企業名の修正を行う。

21. 修正した結果に対して企業名でマッチングを行う。マッチングしなかった事業所は

手順 24 の処理の対象となる。手順 21-23 は(SK_matching_7.do)で実行する。

22. 名寄せ作業 A を行う。マッチングしなかった事業所は手順 23 の処理の対象となる。
(マッチング成功⑩)

23. 名寄せ作業 B を行う。マッチングしなかった事業所は手順 17-19 と同様の理由で
マッチング不可能とする。
(マッチング成功⑪)

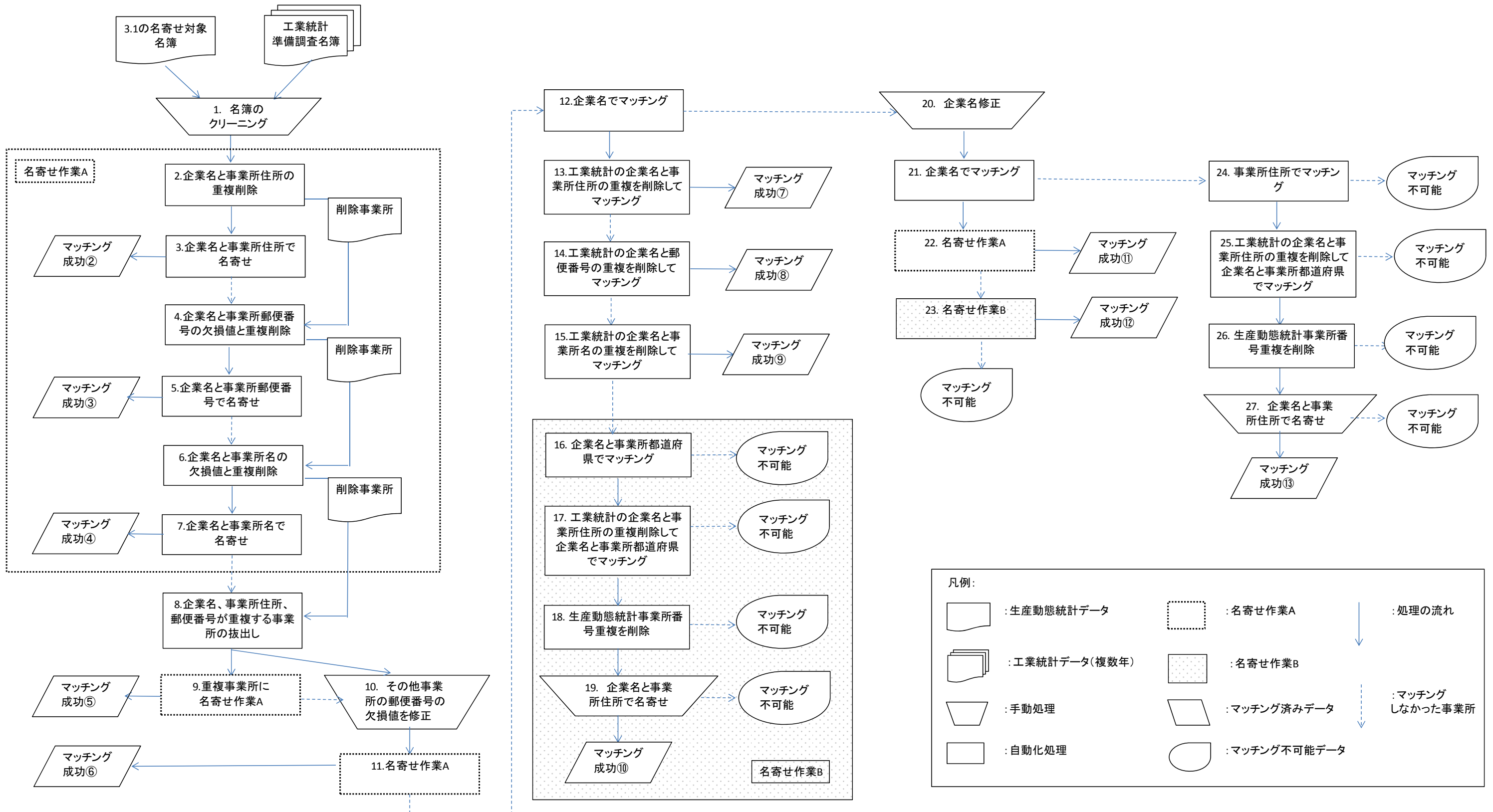
24. 修正後に企業名でマッチングできなかった事業所に対して、事業所住所でマッ
チングする。マッチングできなかった事業所はマッチング不可能とする。手順 24-26 は
(SK_matching_8.do)で実行する。

25. 工業統計の企業名と事業所住所の重複を全て削除して、企業名と事業所住所でマッ
チングする。マッチングできなかった事業所は、工業統計の名簿で企業名と事業所住所
に重複がある事業所でこれ以上のマッチングは不可能である。

26. マッチング結果で生産動態統計の事業所が複数存在している事業所については、一
つの生産動態統計の事業所に対して複数の工業統計の事業所が対応しておりどの事業
所とマッチングすればよいかの判断はできないためマッチング不可能とする。

27. 手動で企業名と事業所住所のチェックを行いマッチングする。生産動態と工業統計
の企業名と事業所の対応として、1:1、m:1、m:m のものが含まれている。特に m:m 対
応になっているものを目視でマッチングするときには企業情報を調べ慎重に行うこと。
マッチングできなかったものはマッチング不可能とする。
(マッチング成功⑬)

図3 名寄せによるマッチング法のフローチャート



3. 3 工業統計パネル化コンバータとのマッチング (STEP3)

前節までで生産動態統計の事業所に、工業統計の各年の事業所識別変数(都道府県番号、市区町村番号、事業所番号)が紐付けられた。これらを基に、阿部他(2012)で提案されたパネル化コンバータで表4の項番1のidを振る。これにより、生産動態統計調査に含まれない日本標準産業分類コードや工業統計調査の回答結果を接続することができる。表6は完成した「生産動態統計－工業統計事業所マッチングコンバータ」に含まれる変数一覧である。項番8－13のように各年を列でつないだ形式のパネルデータをSTATAではWide形式と呼ぶ²³。作業手順を以下で示す。

表6 生動－工業マッチングコンバータに含まれる変数 (Wide形式)

項番	変数名	内容	備考
1	seido_code	事業所ごとのユニークな番号	生産動態統計
2	id	事業所ごとのユニークな番号	工業統計
3	f_name	企業名	両統計より
4	address_f	企業住所	生産動態統計
5	e_name	事業所名	両統計より
6	address_e	事業所住所	両統計より
7	state	都道府県番号	工業統計
8	cityYYYY	市区町村番号 (YYYYは西暦年)	工業統計 対象年の数分存在
9	code_estYYYY	事業所番号 (YYYYは西暦年)	
10	i_codeYYYY	産業分類 (YYYYは西暦年)	
11	kbnYYYY	甲乙区分 (YYYYは西暦年)	
12	ans_flgYYYY	回答フラグ (YYYYは西暦年)	
13	i_codeUniYYY	統一産業分類 (YYYYは西暦年)	

1. 3.1節と3.2節でマッチングに成功した13個のファイルをSK_matching_9.doで結合する。その際変数名の整理を行う。

2. SK_matching_10.doでマッチングに成功した事業所と工業統計のパネル化コンバータをマッチングする。まず、2005年の都道府県番号、市区町村番号、事業所番号でidを振り、マッチングできなかった事業所を2006年の都道府県番号、市区町村番号、事業所番号でマッチングする。これを2009年まで繰り返す。パネル化コンバータを利用することにより、1年でも都道府県番号、市区町村番号、事業所番号が振られていれば

²³ 対して、例えばcityという変数名を作って、各年を縦につないだパネルデータをLong形式という。データの形や分析目的によってどちらかを選ぶ。

一意な id を得ることができるというメリットがある。

4. マッチング結果とその評価

前章までで、「生産動態統計－工業統計事業所マッチングコンバータ」の作成方法を提案した。これは、生産動態統計調査と工業統計調査の個票（回答結果）とをつなぐコンバータとなる。

本稿では、生産動態統計の事業所のうち、2005 年から 2010 年の全事業所を対象としている。表 7 はマッチング結果である。対象事業所中、工業統計の事業所 id を 88.71% の事業所に接続することができた。本稿ではなるべく多くの事業所をマッチングするために、手作業や目視を含む多くの工程を要したが、総務省から提供される府省共通事業所コードによるマッチングで 57.63%（マッチング成功①）、その後の名寄せ自動化作業のみで接続できたものが 83.55%（マッチング成功①－④）もあった。残りの作業で接続したのは 5%以下であった²⁴。

本稿では、1 つでも多くの事業所に工業統計の情報を結びつけることを目的とし、事業所対応は 1:1 に加え m:1 も許している。つまり異なる生産動態統計の事業所に同一の工業統計の事業所が対応しているものもある。これは同住所、同郵便番号、同事業所名であるが、異なる生産動態統計の事業所番号が振られているものがあることを意味する。この様な事業所をどのように分析に使うかは、研究者の目的、期間、対象業種でも異なってくるだろうが、1:1 対応が扱い易いのは明らかである。

本稿の結果では、マッチングされた事業所のうち 97%以上が 1:1 でマッチングされていた。これは、本稿で採用した方法が上手く機能している証拠と成り得る。

最後に、本稿の方法でどのような事業所がマッチング不可能かを述べる²⁵。まず、①両統計の年次が厳密には等しくならないことからそのずれている期間に新規名称、新規調査対象、開業した事業所はマッチングできない。次に、②鉱業やエネルギー生産（例えば電力会社）のみを行っている事業所も製造業の対象とならないのでマッチングできない。また、③工業統計において企業名と事業所住所、企業名と郵便番号、企業名と事業所住所のいずれかに重複がある場合は対象とならない。

この種のマッチングは、記入ミスの修正や期間調整が必要となり、全て自動化することが困難である。そのためその手作業部分が研究者により方法に違いが出てくる。どこまで完璧にするか、時間をかけるかは個々人の考え方に依るが、かけるコストとマッチング率上昇にはトレードオフの関係がある。

²⁴ 筆者はこの 5%以下の部分の作業を通じて、多くのマッチング時の問題点やデータ知識の増大、またプログラムの効率化が実現できたことを記しておく。

²⁵ 記入ミスは除く。

表7 マッチング結果

	マッチング 成功番号	マッチング率	累積比率
自動化	①	57.63%	
	②	18.91%	76.54%
	③	5.32%	81.86%
	④	1.46%	83.32%
	⑤	0.23%	83.55%
手動＋ 自動化	⑥	2.13%	85.69%
	⑦	0.47%	86.16%
	⑧	0.05%	86.20%
	⑨	0.02%	86.23%
	⑩	0.45%	86.68%
	⑪	0.54%	87.22%
	⑫	0.02%	87.24%
	⑬	1.47%	88.71%
合計		88.71%	---

5. まとめと分析に対するいくつかの留意点

本稿では、「生産動態統計－工業統計事業所マッチングコンバータ」を作成した。その際、総務省から提供される府省共通事業所コードと両統計の企業名、事業所住所、郵便番号、事業所による名寄せでマッチングを行った。工業統計では事業所番号が年により変化するため、阿部他(2012)で作成された「工業統計パネル化のためのコンバータ」をマッチングし、年を通じてユニークな工業統計の事業所 id を付与している。マッチング率は 88.71%で、そのうち両統計の事業所が 1:1 対応で接続しているものは、約 97%であった。この新たなコンバータにより、生産動態統計調査と工業統計調査の個票（回答）データを接続することが可能となる。

本稿では両統計の「事業所」を名簿ベースで接続することを目的としているが、両統計の個票（回答）を接続して実証分析を行う際の留意点を記しておく。まず、①生産動態統計調査は月次調査で、工業統計調査は年次調査（12 月末日に調査）である。②生産動態統計調査は品目ベースであるので、調査項目は品目ごとにある。対して工業統計調査は事業所ベースの調査であり、事業所全体の生産活動に関する情報が含まれている。今回は、全自動化はできなかったものの、高いマッチング率が実現でき、コンバータ作成の方針を立てることができた。この方法は一例であり、表 3 にあるように 6 か所に手作業が入る。この部分は個々の研究者によりマッチング率を上昇させる余地がある部分である。

今回のマッチングで、特に、府省共通事業所コードのマッチングの効率化への寄与は非

常に大きかった。今後、このような事業所、企業を包括する共通コードが整備されていけば、マッチング作業の自動化が進み、精度の高いマイクロデータセット構築が実現するであろう。このような包括的なマイクロデータにより、エビデンスに基づく経済政策の立案が活発となることを期待する。

参考文献

- [1] 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(1993-2009年)」、RIETI-PDP、12-P-007.
- [2] Y. Konishi and Y. Nishiyama (2012) "Decomposition of Supply and Demand shock in Production Function using Current Survey of Production", RIETI-DP, *forthcoming*.

Appendix 1 工業統計調査の調査項目

甲票		乙票	
2009		2009	
No.	変数名	No.	変数名
1	調査年度	1	調査年度
2	甲乙区分	2	甲乙区分
3	都道府県番号	3	都道府県番号
4	市町村番号	4	市町村番号
5	調査区	5	調査区
6	事業所番号	6	事業所番号
7	産業コード	7	産業コード
8	他事有無	8	他事有無
9	経営組織	9	経営組織
10	資本金額	10	資本金額
11	正社員男	11	正社員男
12	パートアルバイト男	12	パートアルバイト男
13	出向・派遣男	13	出向・派遣男
14	従業者合計男	14	従業者合計男
15	臨時雇用者男	15	臨時雇用者男
16	正社員女	16	正社員女
17	パートアルバイト女	17	パートアルバイト女
18	出向・派遣女	18	出向・派遣女
19	従業者合計女	19	従業者合計女
20	臨時雇用者女	20	臨時雇用者女
21	個人男	21	個人男
22	個人女	22	個人女
23	従業者合計	23	従業者合計
24	延べ常用労働者		
25	常用雇用者給与		
26	その他の給与		
27	現金給与合計	24	現金給与合計
28	原材料使用額		
29	燃料使用額		
30	電力使用額		
31	委託生産費		
32	原材料使用額等	25	原材料使用額等
33	年初土地		
34	年初有形計		
35	取得中古土地		
36	取得新規有形計		
37	取得新規建物		
38	取得新規機械		
39	取得新規その他		
40	除却土地		
41	除却有形計		
42	減価償却額		
43	建仮の増		
44	建仮の減		
45	年初在庫製造品		
46	年初在庫半製品		
47	年初在庫原材料		
48	年初在庫合計		
49	年末在庫製造品		
50	年末在庫半製品		
51	年末在庫原材料		
52	年末在庫合計		
53	製造品出荷額計	26	製造品出荷額計
54	在庫金額積上計		
55	加工賃収入額計	27	加工賃収入額計
56	修理料収入額	28	修理料収入額
57	出荷額等合計	30	出荷額等合計
58	内国消費税額	31	内国消費税額
59	敷地面積		
60	建物面積		
61	原料用		
62	淡水用途別合計		

Appendix 2 変数名一覧 (3.1 節)

統計名	作業ファイル	旧変数名	新変数名	作業手順
生産動態統計	提供用名簿	事業所番号	seido_code	2
		企業名	f_name	
		企業住所	address_f	
		事業所名	e_name	
		事業所住所	address_e	
		事業所番号	seido_code	
	事業所対応表	総務省コード	somu_code	3
		事業所番号	seido_code	
工業統計	準備調査名簿	総務省コード	somu_code	4
		都道府県番号	state	
		市区町村番号	city	
		事業所番号	code_est	
		企業名	f_name	
		事業所名	e_name	
		事業所郵便番号3桁	zip_1	
		事業所郵便番号4桁	zip_2	
		事業所都道府県名	add_1	
		事業所市区郡名	add_2	
		事業所町村名	add_3	
		事業所番地	add_4	
	工業・事業所コード	都道府県番号	state	5
		市区町村番号	city	
		事業所番号	code_est	
		総務省コード	somu_code	