

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

周防 節雄¹・古隅 弘樹²・宮内 環³

(受付 2009年5月29日；改訂 9月7日；採択 10月1日)

要 旨

統計法に基づいて目的外使用の許諾を得た法人企業統計調査および事業所・企業統計調査それぞれの個票マイクロデータを用いて完全照合によるパネルデータを作成し、企業データベースを構築した。法人企業統計調査のパネルデータの作成では、1985年から2005年に亘る毎年の年報調査および四半期調査それぞれの調査名簿と調査票データについて同一法人企業のレコードを接続した。事業所統計調査、事業所・企業統計調査のパネルデータの作成では、1991年、1994年の事業所統計調査および1996、1999、2001、2004の各年の事業所・企業統計調査について、完全照合の方法により同一事業所のレコードを縦断的に接続した。さらに、2004年時点における法人企業統計調査の調査名簿と事業所・企業統計調査の民営事業所漢字リストテープの企業名や所在地を突合し、相互のパネルデータにおける企業レコード間の接続を行った。本稿では、これらの企業データベースの構築におけるパネル作成および相互リンケージに係る手法について詳解する。

キーワード： パネルデータ、完全照合、文字列照合、統計マイクロデータ。

1. はじめに

我が国の統計調査システムは所轄省庁による分散型であり、カナダのような集中型ではない。また、海外で行われているような税務記録を利用した母集団名簿の整備(Wallgren and Wallgren, 2007)についても未だ実現が難しいのが現状である。このような状況下で、事業所・企業に関する統計調査の母集団名簿を整備するためには、別々に管理された母集団名簿を統合するための完全照合(exact matching)が必要である。なお、先行研究である松田(1991)では、複数の統計調査を接続するための方法論の提示と部分的な実験が行われている。また、Ruggles and Ruggles (1999)では複数の企業マイクロデータの統合と分析が行われている。

著者らが参加する研究プロジェクト(謝辞を参照)では、統計法に基づく目的外使用申請を行い、事業所・企業統計調査と法人企業統計調査それぞれのマイクロ(個票)データが利用可能となった。それらを完全照合手法によってマッチングを行い、相互の接続情報を利用した分析が可能な企業データベースの整備作業を行った。

¹ 兵庫県立大学 神戸学園都市学術情報館：〒651-2197 兵庫県神戸市西区学園西町 8-2-1

² 兵庫県立大学 経済学部：〒651-2197 兵庫県神戸市西区学園西町 8-2-1

³ 慶應義塾大学 経済学部：〒108-8345 東京都港区三田 2-15-45

法人企業統計調査は、平成19年調査までは金融保険業が調査対象から外されてきたが、包括的企業統計としては日本で唯一のもので、年次と四半期で財務情報に関する標本調査が行われている。一方、事業所・企業統計調査の調査対象は事業所概念に基づく悉皆調査であり、ミクロレベルの企業行動分析に適した調査といえる。両者の接続によって企業レベルの情報を事業所レベルで捉えることができ、これまでにできなかった新たな分析集計を行うことが可能になる。

1.1 法人企業統計調査のマイクロデータ

財務省が実査する法人企業統計調査には、営利法人等を調査対象としたその年度における確定決算の計数を調査する「年次別調査」(昭和23年調査開始)と、資本金、出資金又は基金1,000万円以上の営利法人等を調査対象とした四半期ごとに仮決算計数を調査する「四半期別調査」(昭和25年1~3月期調査開始)がある。今回目的外使用申請で使用を許可されたマイクロデータには、1983年~2005年の各年について①年次別と②四半期別の二種類がある。それぞれの種類ごとに「企業名称」や「所在地」などの情報をマッチングキーとしてパネル化を行った。文字列マッチングに関する技術的手法は第2節で、パネル化作業の詳細は第3節で述べる。

1.2 事業所・企業統計調査のマイクロデータ

総務省統計局が実査する事業所・企業統計調査は、民営の事業所及び企業を対象として、事業の種類や従業者数等、事業所及び企業の基本的事項を調査し、行政施策のための基礎資料並びに各種統計調査実施のための事業所及び企業の名簿を得ることを目的としている。当初、「事業所統計調査」の名称で昭和22年に始まり、平成8年調査から企業の実態把握を充実させたことに伴い、「事業所・企業統計調査」に名称を変更し、この中間年の調査は事業所・企業統計調査の「簡易調査」と位置づけられた。平成11年に簡易調査として初めて実施され、平成16年(2004年)の調査は簡易調査としては2回目に当たる。この調査は昭和23年調査から昭和56年調査までは3年ごと、昭和56年以降は5年ごとに実施されている。また、民営事業所を対象として平成元年及び平成6年に事業所名簿整備のための調査も実施された。今回目的外使用申請で使用を許可されたマイクロデータは、1981年~2004年の調査データである。このデータには当該年次のid番号と前回調査時のid番号が付与されている。この情報をマッチングキーとして、2004年調査から1999年調査まで遡って接続したパネルデータを作成できた。さらに、統計局作成の事業所id番号の変遷情報データを使って1991年まで遡って接続したパネルデータを作成できた。なお、2001年データには本研究プロジェクトの別の研究班で行われた作業による「親子企業」と「本所・支所」の名寄せ結果が含まれている。このパネルデータについては第4節で述べる。

1.3 事業所・企業統計調査と法人企業統計調査の完全照合

異なる二つの統計調査の完全照合実験として、2004年の事業所・企業統計調査と法人企業統計調査のマイクロデータを使った。その際のマッチングキーとしては企業名称と所在地の漢字情報を使用した。ユニークなid番号でマッチングする訳ではないので、文字の上では小さな違いがプログラムによる照合結果に大きく影響してしまう。プログラムによる照合の誤判定を軽減するための様々な工夫を施した上で、厳しい制約条件を少しずつ緩めながらプログラムで照合を行い、緩めた照合条件のもとで行った処理結果を、人間の目で見て確認する作業も行った。これについては第5節で述べる。

本稿の最後でこうした一連のパネル化や異なる統計データでの完全照合の精度について第6節で述べた上で、第7節で問題点や今後の課題について触れる。

2. 文字情報による照合技法

特定の調査客体について、同一の統計調査の異時点間で接続するパネルデータの作成や、異なる統計調査の同一(近接)時点における相互接続を行いたいとき、個々の調査客体を同定して接続するための情報(いわゆるリンクコンスタント)を利用できる場合や、共通の識別番号を使用している場合には問題は無いが、そうでない場合、例えば法人企業統計調査と事業所・企業統計調査における企業データの相互接続用に個々の企業データの接続情報が提供されていない場合には、接続するデータ間で共通する調査項目、例えば名称、所在地、電話番号などの文字情報を用いて独自に照合を行う必要がある。

文字列照合を行う前の事前調査として、比較する項目において使用されている文字の分布を調査するのが望ましい。それによって、想定外の記号や全角・半角の違いなど、当該項目における不適切な入力データや事前処理の不具合が明らかになることが多い。その上で、以下のような事柄について考慮する必要がある。

- 空白文字、区切り記号などの除去の必要性
- 長音記号、句読点などの記号表記
- 文字の半角・全角
- 英字の大文字・小文字
- カタカナ・ひらがな
- 漢字表記における異体字
- 促音および拗音の大書き・小書き

また、今回の作業のように企業名称を照合する際の問題は、「～株式会社」のような企業形態の有無や、「～(株)」のような省略記号の使用による表記揺れが挙げられる。単純にこれら除去してしまうと、株式会社と有限会社で異なるのに同名と判定されたり、これらが名称の前後どちらに付くかで企業が異なるため判別できないといった弊害が出る。そのため、企業名称における企業形態の種類と前後の位置を名称と分けて照合するといった手法が必要となる。

加えて、統計調査によっては住所の情報を正規化(コード化)することで住所の表記揺れによる弊害をなくす工夫がなされているが、このような方策がとられていない場合や、異なる正規化手法が使われている場合には相互の住所に関する文字列情報を照合する必要がある。

所在地の照合における最大の問題は番地情報の表記揺れである。例えば、「1-2-3」という番地表記は「一丁目二番三号」や「1丁目2番地の3」など、種々のパターンで表記されるため、この表記揺れ問題を解決しなければ住所情報の文字列照合は成功しない。つまり、同一の住所正規化手法を用いて共通のコード化を行うか、番地情報の表記を統一するよう変換処理を施すことが必要となる。住所については他に市区町村合併の影響もあり、新旧住所表記の混在に加え、調査名簿の整備時点や調査時期による時間的差異も絡んでくる。

表記を統一する方法は情報の内容を損なわずに照合の成功率を高める方法であるが、すべての表記パターンについて適切な変換処理を行うことは高度なパターンマッチング技術が要求される。そこで、多少の情報の正確さを損なうことになるが、重要な部分をうまく残してマッチングの確率を高めるための照合用キーを作成するアプローチもある。例えば、ひらがな・カタカナはカタカナに統一し、漢数字・算用数字・ローマ数字は算用数字に統一、住所の番地表記も数字だけ(1-2-3は123)にするなどである。

文字情報の照合手順としては、最も厳しい照合条件から徐々に条件を緩和しながら照合し、マッチングできたレコードは以降の照合作業では除外するようにすれば、以降の照合件数を少なくでき、多重接続(1対多, 多対1, 多対多)を抑えることができる。ただし、厳しい条件下で

1対1で接続したレコードは問題ないが、緩い条件下での接続や多重接続の場合については、後で目視チェックをするなど、ケアが必要になる。

3. 法人企業統計調査のパネルデータ作成

財務省が所管する法人企業統計調査について、目的外使用申請によって提供を受けたマイクロデータを用いて法人企業パネルデータを作成した。

法人企業統計調査では、資本金1億円以上の法人企業は管理法人名簿によって整備されている。低階層法人名簿は、資本金1億円未満の法人のうち既定の方法によって抽出された調査対象法人について作成されたものである。上記2つの法人名簿によって標本抽出の方法が異なり、資本金階層別、業種別に分類層化した各セルから標本法人が抽出されている。

標本法人数の資本金階層別の内訳を表1に示す。なお、資本金階層が1億円以上10億円未満の法人企業において、その抽出方法から、実質的に資本金6億円以上については悉皆調査対象である。

3.1 ミクロデータの概要

法人企業統計調査(1983~2005年)における年次別調査および四半期別調査それぞれについて以下のテキストファイル(CSV形式)の提供を受けた。

調査名簿

法人母集団から調査対象となる法人を抽出した名簿データ。

調査票

調査名簿の法人に対して調査を行った結果の回答個票データ。

推計データ

調査結果を基に算出した各集計項目の母集団推計データ。個票データではなく集計量データのため、今回のパネルデータ作成では使用しない。

調査年次別の調査名簿および調査票データにおけるレコード数を表2に示す。なお、調査年次とデータの種類毎にファイルは分割されており、資本金1億円以上の管理法人と1億円未満の低階層法人は分けられておらず、季報データについても同一調査年次の場合は同一ファイルに4回分の四半期調査データが含まれている。

なお、調査項目の大幅な改定が行われた影響で、2002年以前と2003年以降では調査項目が異なっている。また、法人名称や所在地の文字列データが前者では半角のため日本語部分はカ

表1. 法人企業統計調査における資本金階層別標本法人数。

資本金区分	年報	季報
200万円未満	約2,000社	—
200万円以上, 300万円未満	約1,000社	—
300万円以上, 500万円未満	約2,000社	—
500万円以上, 1000万円未満	約2,000社	—
1000万円以上, 2000万円未満	約4,000社	約4,000社
2000万円以上, 5000万円未満	約4,000社	約4,000社
5000万円以上, 1億円未満	約2,000社	約2,000社
1億円以上, 10億円未満	確率比例抽出	確率比例抽出
10億円以上	全数抽出	全数抽出

財務省財務総合政策研究所研修部(2006)より引用

表 2. 法人企業統計調査の調査名簿および調査票データにおける調査年別レコード数.

調査年	調査名簿					調査年	調査票				
	年報	季報 1	季報 2	季報 3	季報 4		年報	季報 1	季報 2	季報 3	季報 4
1983	28,777	24,083	23,956	23,888	23,828	1983	17,884	13,601	13,480	13,343	13,318
1984	30,859	25,739	25,567	25,508	25,436	1984	19,394	14,419	14,387	14,259	14,202
1985	31,485	26,344	26,199	26,138	26,065	1985	19,427	14,636	14,613	14,417	14,285
1986	33,822	27,772	27,569	27,463	27,393	1986	21,261	15,305	15,414	15,308	15,145
1987	36,657	30,364	30,206	30,114	30,052	1987	21,931	15,997	15,892	15,743	15,521
1988	36,979	30,807	30,608	30,543	30,477	1988	22,184	16,276	16,090	15,976	15,725
1989	40,395	33,770	33,660	33,584	33,517	1989	23,634	17,249	17,129	17,028	16,831
1990	41,708	34,915	34,787	34,687	34,628	1990	24,759	18,143	17,991	17,797	17,681
1991	43,736	37,184	37,024	36,920	36,844	1991	24,994	19,039	18,848	18,611	18,267
1992	45,257	38,723	38,558	38,435	38,128	1992	25,457	19,666	19,398	19,004	18,782
1993	46,344	39,982	39,802	39,639	39,467	1993	26,040	20,101	19,946	19,626	19,332
1994	47,270	40,881	40,700	40,565	40,350	1994	26,217	20,621	20,361	19,796	19,721
1995	47,582	41,306	41,086	40,990	40,810	1995	26,593	20,460	20,580	20,302	19,990
1996	48,272	41,407	41,212	41,119	40,897	1996	25,691	19,296	19,337	19,025	18,684
1997	49,070	42,057	41,844	41,743	41,560	1997	25,394	19,550	19,443	19,007	18,648
1998	50,251	43,286	43,062	42,872	42,746	1998	25,505	19,448	19,045	18,730	18,268
1999	50,328	43,912	43,735	43,602	43,385	1999	24,553	19,669	19,282	18,768	18,332
2000	50,569	44,644	44,359	44,139	43,825	2000	24,337	19,355	19,139	18,519	17,967
2001	49,011	46,430	46,006	45,725	42,735	2001	23,843	19,141	18,974	18,482	17,985
2002	52,392	47,070	46,575	46,370	45,989	2002	23,839	19,241	19,181	18,736	18,119
2003	56,973	49,351	49,352	49,352	49,361	2003	24,898	19,528	19,274	19,104	18,689
2004	59,180	50,570	50,570	50,571	50,571	2004	25,846	20,618	20,327	20,058	19,838
2005	60,084	51,307	51,307	51,307	51,309	2005	25,751	20,223	20,094	19,739	19,264

タカナだったが、後者では全角になり漢字表記が可能になっている。

法人企業のパネルデータを作成するにあたり、資本金1億円以上の管理法人と1億円未満の低階層法人では、法人を識別する法人番号の割り当て方法が調査年次間で異なるため、同じ方法で作業を行うことができない。そのため、これらを別々にパネル化作業を行い、最終的に両者を統合することにした。それぞれのパネル化作業については以下で解説する。

3.2 管理法人部分のパネル化

資本金1億円以上の法人企業、いわゆる管理法人については、40000番未満の法人番号が割り当てられており、年報調査および季報調査の両方で調査対象となっている。また、管理法人については基本的に調査年が変わっても同一法人には同じ番号が割り当てられている。ただし、調査名簿において法人の統廃合や減資によって生じた法人番号の欠番は新規の管理法人に再割当されるため、今回の照合作業が必要になった。

事前準備として照合対象となる法人名と所在地に対して記号処理・表記統一処理を施した後、以下の比較順位で照合して、調査年次間の法人番号による法人企業の同一性、連続性をプログラムで判定した。判定結果は表3に示すように、照合条件と一致度合いに応じた判定コードを付与した。

- (1) 同一法人番号を持つレコードについて、調査名簿データを前後の調査年で比較し、その一致度合を判定コードとして付与する。

比較項目： 法人名称、所在地

- (2) 調査名簿データで判定できない場合、調査票データがあれば、前後の調査年で比較し、その一致度合を判定コードとして付与する。

比較項目： 当初資本金，資本金(期首，期末)

(3) 上記の比較で判定できない場合は新規法人とみなす。

5桁の数値で表記される現行の法人番号では、欠番が再利用されており、1983～2005年の間に調査名簿に掲載された全ての法人企業を一意に識別することができないことから、法人番号の末尾に2桁のサブコードを追加した7桁の法人パネルIDを新たに作成した。なお、法人番号(5桁)の新規法人への重複割り当てがあった場合に、末尾2桁のサブコード(初期値は00)をインクリメントすることで、全調査年次において法人企業を一意に識別できるようにした。

なお、上記の比較手順では以下の規則も適用している。

- 年報名簿と季報名簿では同一調査年において共通の法人番号を持ち、四半期の途中で法人の入れ替えはないものとする。
- 調査名簿データの比較では、前後の調査年において年報と季報(×4)の計5つの項目のいずれかと一致すればよいとする。
- 調査票データの比較では、年報と季報は別々に比較する。
- 調査名簿データにおいて法人番号が途中の調査年で欠番になった場合、その前後は別法人とみなす。

季報調査では、調査対象法人は4～6月期の調査から更新され、1年間継続して調査が行われている。季報調査は四半期毎の調査であることから、調査年次の途中で名称や所在地が変わった場合には、同一調査年次で複数の名称と所在地の情報を名簿情報に保持することになる。この情報は一種の変遷情報であるから、この年次の前後の調査年との照合においては、この変遷情報と比較することで両者の接続が可能となる。また、調査票データの項目については年報と季報では調査項目や調査時期が異なることから比較する際には区別している。調査名簿における法人番号の欠番については、その番号を使用していた法人企業が統廃合や減資によって管理法人名簿から外れたことを意味しており、実際に調査名簿上で法人番号が欠番となった後で同じ法人企業が再び同一番号を割り当てられることは非常に稀なケースであった(なお、プログラム処理結果の目視チェック作業で確認したところ、このようなケースは22,737件中1件だけだった)。

プログラムによる判定結果を表4に示す。ただし、2002年以前と2003年以降の法人の比較照合では、文字情報が半角文字(英数・カタカナ)と全角文字(漢字)で異なるため、この間のプログラムによるマッチングの精度には多少の問題があることを留意されたい。この点については、2003～2004年における判定結果のばらつきからも読み取れる。なお、最終的にはプログラ

表3. 管理法人パネルの作成における接続判定条件。

判定コード	2002年以前(半角文字)	2003年(全角文字)	2004年以降(全角文字)
(欠損値)	調査名簿なし	調査名簿なし	調査名簿なし
-1	未判定	未判定	未判定
0	初出	初出	初出
1	法人名が一致	法人名が一致	法人名が一致
2	法人名が部分一致(一方が他方を包含)	法人名が部分一致(一方が他方を包含)	法人名が部分一致(一方が他方を包含)
3	所在地が一致	所在地が一致	所在地が一致
4	所在地が部分一致(一方が他方を包含)	所在地が部分一致(一方が他方を包含)	所在地が部分一致(一方が他方を包含)
5	—	法人名および所在地の半角文字列が一致	法人名および所在地の半角文字列が一致
6	所在地の番地情報が一致	法人名または所在地の半角文字列が一致	法人名または所在地の半角文字列が一致
7	当初資本金が一致(2002年以前のみ)	—	郵便番号が一致(2003年以降のみ)
8	資本金が一致(当該期首と前回期末)	資本金が一致(当該前期と前回期末)	資本金が一致(当該前期と前回当期)
9	新規法人と判定	新規法人と判定	新規法人と判定

表 4. 管理法人パネルの接続結果件数(目視修正前).

調査年	判定コード										合計
	0	1	2	3	4	5	6	7	8	9	
1983	17,059	0	0	0	0	0	0	0	0	0	17,059
1984	1,121	16,526	29	57	0	0	3	0	0	19	17,755
1985	1,198	17,186	28	64	4	0	5	0	0	46	18,531
1986	1,093	17,842	28	81	2	0	4	0	0	66	19,116
1987	2,930	18,504	42	124	0	0	9	0	0	49	21,658
1988	1,808	19,847	77	116	3	0	14	0	0	114	21,979
1989	3,414	20,774	95	113	6	0	19	0	0	56	24,477
1990	2,580	22,722	65	157	3	0	14	0	0	41	25,582
1991	3,088	24,262	78	155	1	0	17	0	0	49	27,650
1992	715	26,234	67	169	2	0	22	0	0	2,167	29,376
1993	15	28,009	50	131	1	0	24	0	0	2,322	30,552
1994	325	29,091	95	139	3	0	33	0	0	1,681	31,367
1995	476	29,832	80	102	2	0	17	0	0	1,317	31,826
1996	441	30,076	56	132	6	0	25	0	0	1,346	32,082
1997	390	30,436	58	112	1	0	24	0	0	1,605	32,626
1998	599	31,198	73	94	6	0	21	0	0	1,969	33,960
1999	10	32,234	55	134	1	0	30	0	0	2,187	34,651
2000	55	32,726	70	147	1	0	34	0	0	2,369	35,402
2001	2,433	33,234	55	160	4	0	38	0	0	1,675	37,599
2002	2	34,386	73	140	3	0	64	0	0	3,612	38,280
2003	1	25,735	1	6,131	421	1,085	1,986	0	267	3,312	38,939
2004	246	24,527	7	957	32	3,281	6,039	0	157	3,865	39,111
2005	0	35,970	10	169	3	40	123	40	5	3,266	39,626
合計	39,999	581,351	1,192	9,584	505	4,406	8,565	40	429	33,133	679,204

注) 判定コード = 0 : 法人番号初出, 判定コード = 1~8 : 継続法人, 判定コード = 9 : 新規法人
 当該年とその前年に調査名簿で同じ法人番号を持つ法人企業同士を比較照合して判定.

ムによる判定結果について目視作業による調整を行ったが、これについては後述する。

表 4 の結果において、1992 年以降で非接続件数が増えているのは商法の改正によって調査対象となる法人企業の増加が影響していると考えられる。なお、この商法改正に伴って、1995 年度の調査から資本金階層区分とその層別標本法人数が変更されている。

3.3 低階層法人部分のパネル化

資本金 1 億円未満の低階層法人については、調査対象法人の負担軽減のために、年次別調査と四半期別調査とで標本法人が重複せず、2 年連続で抽出されないよう配慮されている。また、資本金階層別、業種別に分類層化した各セルにおける最低標本数は、資本金 1 千万円未満の階層(年報のみ)では 30 社、1 千万以上 1 億円未満では 50 社となっている。ただし、前者のセルで 60 社未満、後者のセルで 200 社未満のように母集団の法人数が少ない場合は複数年連続で調査対象になることがある(表 5 参照)。

管理法人部分のパネル化と同様に、1983～2005 年における年報・季報それぞれの調査名簿データを用いて低階層法人パネルデータを作成する。管理法人の場合は年報と季報で共通の法人番号(00001～39999)が割り当てられており、調査年が変わっても法人番号が変わらないことを前提にパネルを作成することができたが、低階層法人の場合は年報と季報で異なる法人番号が割り当てられている(年報：40000～79999, 季報：80000～99999)ことに加え、調査年ご

表5. 低階層法人における複数年連続して調査対象になる(再抽出の)可能性が高い資本金階層と業種.

資本金	業種
1千万円未満	石油・石炭製品, 電気, ガス・熱供給・水道業等
1千万円以上, 5千万円未満	林業, 石油・石炭製品, 電気, ガス・熱供給・水道業等
5千万円以上, 1億円未満	農業, 林業, 漁業, 鉱業, 木材・木製品, パルプ・紙, 石油・石炭製品, 非鉄金属, 精密機械, 水運, 電気, ガス・熱供給・水道業等

※財務省財務総合政策研究所研修部(2006)から作成

とに法人番号が新規に割り振られており, 管理法人の場合のように法人番号を用いて接続することはできない. さらに連続する調査年次で調査対象になっている法人かどうかを見るためには, 年報と季報の統合パネルを作成する必要がある. 低階層法人パネルの作成は以下の手順で行った.

(1) 事前準備

照合対象となる法人名と所在地に対して記号処理・表記統一処理を施す(管理法人の場合と同様).

(2) 年報パネルの作成

全ての調査年次の組み合わせについて調査名簿間の照合を行い, 法人名と所在地が両方一致することを条件としてリンクテーブルを作成する.

(3) 季報パネルの作成

年報パネルと同様に全ての調査年次の組み合わせについて調査名簿データを照合してリンクテーブルを作成する. ただし, 先行年の第4四半期と後続年の第1四半期で比較する.

(4) 年報と季報のリンケージ

各調査年次の名簿データを用いて法人名および住所の一致を条件としてリンクテーブルを作成する. ただし, 季報データが先行年の場合は第4四半期, 後続年の場合は第1四半期の情報と年報データを比較する. 同一年次の場合は, 第1四半期と第4四半期の両方を比較し, 一方が一致すればよいとする.

(5) 統合パネルの作成

年報と季報のリンクテーブルをもとに, 年報パネルと季報パネルを統合し, 同一法人を識別するための法人パネルID(7桁)を付与する.

管理法人のパネル化では, 法人番号が基本的に変わらないことを前提として, さまざまな条件を設定して判定作業を行ったが, 低階層法人の場合は法人番号が使えないため, 法人名と所在地がともに一致する場合に限り同一法人とみなす最も厳しい条件下で1983~2005年の全年次間の組み合わせで照合を行っている. このため, 法人名や所在地が変わった場合は別法人という判定となり, 目視作業による確認も行っていない.

2002年以前と2003年以降では文字情報が半角文字(英数・カタカナ)と全角文字(漢字)で異なるため, 両者を接続することは技術的に困難であるが, 2003年以降の全角データにおいても, 2002年以前の半角カタカナを全角カタカナに変換しただけのデータも含まれていたため, 可能な限り比較データを半角変換して照合を行った. しかし, 作成したパネルにはこれに起因する大きな断層が2002年以前と2003年以降の間に存在するため, ここでは1983~2002年に

表 6. 低階層法人(年報・季報統合パネル, 1983-2002年)におけるパネル長別法人数.

パネル長 (年数)	パネル数 (法人数)	パネル数 パーセント
1	331,440	85.71
2	39,053	10.10
3	9,016	2.33
4	2,898	0.75
5	1,339	0.35
6	894	0.23
7	547	0.14
8	318	0.08
9	308	0.08
10	177	0.05
11	118	0.03
12	94	0.02
13	95	0.02
14	64	0.02
15	70	0.02
16	64	0.02
17	54	0.01
18	69	0.02
19	52	0.01
20	14	0.00
合計	386,684	100.00

におけるパネル作成作業の結果について述べる。

作成した低階層法人の年報・季報統合パネルについて、パネル長(年数)別の法人企業数の分布を表 6、1983～2002年の間に5回以上抽出された法人(パネル長 ≥ 5)について業種別資本金規模別(セル別)の法人数のグラフを図 1 に示す。(なお、2003年以降も含む1983～2005年の年報・季報統合パネルでは、パネル長=22は7件、パネル長=23は2件であった。)

低階層法人名簿の対象外である資本金1億円以上の法人企業が存在するのは、標本抽出後に回収した調査票データに応じて事後修正したからと考えられる。同一調査年次の年報と季報では標本法人が基本的に重複しないことから、年報・季報統合パネルでは、両方で標本抽出されている1千万以上1億円未満の資本金階層で標本法人数が多くなっている(表 1 参照)。表 5 で示した業種では企業数が少ないため、標本に選ばれる確率がかなり高いことや、放送業(2003年以降は情報通信業)でもすべての資本金階層で再抽出が多いことがわかる。

4. 事業所・企業統計調査のパネルデータ作成

事業所統計調査及び事業所・企業統計調査のマイクロデータを用いたパネルデータの作成について述べる。まずパネルデータ作成に用いた事業所・企業統計調査のファイルについてその概略を述べ、つぎにこれらファイルにおける同一事業所を時系列的に接続したパネルデータの作成について述べることにする。

以下では、記述の簡潔さのために、「事業所統計調査」と「事業所・企業統計調査」を区別する必要のない場合には、これらを「事業所・企業統計調査」で代表させることにする。また、混乱の恐れのない場合には1994年に実施された「事業所名簿整備調査」についても「事業

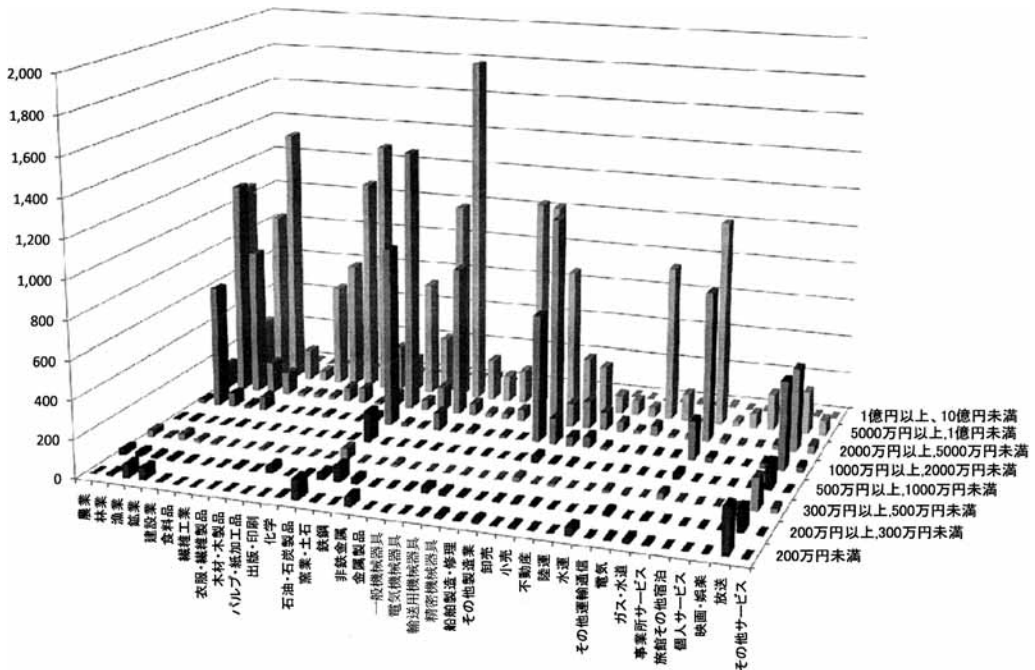


図1. 低階層法人(年報・季報統合パネル, 1983~2002年)における業種別資本金階層別のべ法人人数(パネル長 ≥ 5)。

所統計調査」あるいは「事業所・企業統計調査」と記述する。

4.1 ミクロデータの概略

事業所・企業統計調査のパネルデータの作成には、1991年の事業所統計調査、1994年の事業所名簿整備調査および1996、1999、2001、2004の各年の事業所・企業統計調査のマイクロデータを用いた。

事業所統計調査及び事業所・企業統計について指定統計の目的外使用の許可が得られたのは、1981、1986、1991、1994の各年の事業所統計調査及び1996、1999、2001、2004の各年の事業所・企業統計調査であるが、パネルデータの作成に1981年、1986年の事業所統計調査を含めなかった。というのは、松田(1991)の完全照合の方法でも使われたが、同一事業所を時系列的に接続するための情報が、これまでのところ1991年の事業所統計調査から2004年の事業所・企業統計調査までの範囲に限られていたからである。

事業所・企業統計調査のパネルデータの作成に用いたマイクロデータには大別して2種類のファイルが含まれている。1つは事業所番号などのほかに事業所の名称や所在地などの文字情報を含むファイルで、これを「事業所漢字リストテープ」と呼ぶ。もう1つは事業所の名称や所在地などの文字情報を含まず、事業所番号や従業者数や産業分類などの情報のみを含むファイルで、これを「事業所データテープ」と呼ぶ。

提供されたマイクロデータでは、基本的に民営の法人事業所のみが含まれており、経営組織が個人経営と法人でない団体、および地方公共団体は含まれていない。(ただし、1994年調査の事業所データテープを除く。1994年調査は事業所名簿整備調査として実施され、1994年の事業所データテープには法人以外にも経営組織が個人経営と法人でない団体である民営事業所がす

べて含まれている。他のファイルにこれらの事業所が含まれていないのは統計法に基づく目的外利用申請時の書類の内容によるものである。）

パネルデータ作成に用いたファイルの一覧を表7に示す。以下本稿では記述の簡潔さのために、ファイル番号(#欄)とファイルの略称を用いることにする。参考までに、ファイルに含まれる事業所数(レコード数)と、総務庁統計局(1992, 1995, 1998, 2000)および総務省統計局(2003, 2006)に報告された民営事業所数, うち民営の法人事業所数を併せて記載する。レコード数と公表値に差異があるのは、公表後にもデータ編集が行われていることや、公表値に含まれていない詳細未詳の事業所(未回答, 回答拒否, 移転, 廃業, 等)がデータテープに含まれていることが主たる原因であった。

表8に各年次のデータテープにおける経営組織別の事業所数を示す。ただし1994年の事業所名簿整備調査では事業所の経営組織の別は調査されていないので、1994データテープにおける事業所のレコード数は表8から除いてある。

4.2 パネルデータ作成手順

事業所・企業統計のパネルデータの作成は表7の事業所漢字リストテープと事業所データテープのファイルの双方に含まれる事業所のレコードのうち、各事業所レコードに割り当てら

表7. 事業所・企業統計調査のパネルデータ作成に用いたファイルと公表された事業所数。

#	パネルデータ作成に用いた 事業所統計調査及び事業所・企業統計調査のファイル	本稿で用いる ファイルの略称	ファイル内 レコード数	公表された事業所数	
				民営	うち法人
1	「平成16年事業所・企業統計調査」の事業所漢字リストテープ	2004漢字リストテープ	2,836,728	5,728,492	2,836,717
3	「平成16年事業所・企業統計調査」の事業所データテープ	2004データテープ	4,130,404		
2	「平成13年事業所・企業統計調査」の事業所漢字リストテープ	2001漢字リストテープ	2,971,593	6,138,312	2,971,593
4	「平成13年事業所・企業統計調査」の事業所データテープ	2001データテープ	4,101,788		
5	「平成11年事業所・企業統計調査」の事業所データテープ	1999データテープ	3,985,526	6,203,264	2,924,768
6	「平成8年事業所・企業統計調査」の事業所データテープ	1996データテープ	3,189,284	6,521,837	2,994,096
7	「平成6年事業所名簿整備調査」の事業所データテープ	1994データテープ	7,412,223	6,550,245	...
8	「平成3年事業所統計調査」の事業所データテープ	1991データテープ	2,965,958	6,559,337	2,771,477

注1) 表頭のファイル内レコード数は事業所・企業統計のファイルより筆者集計。

注2) 表頭の公表された事業所数は総務庁統計局(1992, 1995, 1998, 2000)および総務省統計局(2003, 2006)より抜粋。

注3) 1994年事業所名簿整備調査では経営組織について調査を行っていないので、1994データテープに対応する総務庁統計局(1995)には法人の事業所数は報告されていない。

表8. ファイル内の事業所レコードのうち本所・支所の別、経営組織の別が報告されている事業所レコード数。

経営組織	データテープ				
	2004	2001	1999	1996	1991
民営事業所合計	2,836,717	2,971,593	2,924,817	2,994,096	2,771,477
個人経営	0	0	0	0	0
株式会社	1,507,245	1,594,743	1,581,835	1,643,017	1,563,501
有限会社	986,194	1,030,138	1,001,678	1,009,626	883,848
合名・合資会社	24,669	27,593	29,532	32,490	50,779
相互会社	7,831	10,466	11,662	14,090	
外国の会社	1,943	2,410	1,900	2,294	2,127
会社以外の法人	3,08,835	306,243	298,210	292,579	271,222
法人でない団体	0	0	0	0	0

注) ファイル内の事業所レコード数は事業所・企業統計のファイルより筆者集計。

れた識別コードに相当する地域番号や事業所番号を手がかりに完全照合の方法によって同一の事業所のレコードを接続した。以下にこの方法について述べる。

表7の事業所漢字リストテープと事業所データテープに含まれる事業所のレコードにはすべて市区町村番号、調査区番号、事業所番号が付与されており、これらを合わせると個々の事業所の識別番号として利用できる(なお、極めて稀なケースとして重複番号が存在したが、これらはデータ編成上のミスとして処理した)。これを以下では便宜上「現行事業所ID」と呼ぶことにする。

ただしこの現行事業所IDは、菅・宮川(2008)に示されるビジネスレジスターなどの例に見られる特定の事業所に付与される永久事業所番号とは異なり、事業所・企業統計調査の実施年ごとに事業所番号が付け直されるなどして変更されているため、異なる調査年の事業所・企業統計調査の事業所漢字リストテープや事業所データテープに同一の事業所のレコードが含まれているとしても、その同一の事業所のレコードを現行事業所IDによって同定することはできない。

しかし幸いなことに、事業所・企業統計調査のマイクロデータでは、当該実施年における事業所が、前回実施年においても調査対象であり回答があった場合には、前回実施の調査において付与された事業所ID(これを以後は「回顧事業所ID」と呼ぶ)が当該実施年における事業所レコードに付与されている。なお、表7の事業所漢字リストテープと事業所データテープのうち、2004、2001、1999のファイルには、現行事業所IDと回顧事業所IDとの対応の情報が各事業所のレコードに含まれている。この対応を用いて2004の事業所漢字リストテープあるいは事業所データテープから、2001、1999そして1996の事業所データテープにまで遡ってこれらのファイルにまたがって存在する同一の事業所のレコードを接続することでパネルデータの作成が可能である。さらに1996年から1994年、さらに1991年については、別途提供されたこれらの調査年における事業所IDの対応表を用いて同一事業所のレコードの接続を行った。(この対応表を以下では1996年から遡る事業所IDという意味で「1996コンスタント」と呼ぶことにする。なお「コンスタント」とは事業所IDが事業所に決め付けで付与される定数であることからこのように呼ばれる。)

具体的には表9に示された要領でレコードの接続作業を行った。「接合に用いた事業所ID」欄はこの作業で完全照合に用いた2つの事業所IDを左右に対応させて示している。これら事業所IDのうち、左側が表頭に「接合したファイル」とある欄のファイルに含まれる事業所ID、右側が一つ前の接合作業で得られたファイルに含まれる事業所IDで、表の同一行に記された左右の2つを完全照合に用いた。なお、「接合に用いた事業所ID」欄にある括弧内の番号は、接合の試みの順序を表し、(1)の試みで接合ができなかった場合に限り(2)を試し、(2)で接合ができなかった場合に限り(3)を試した。

4.3 作成したパネルデータについて

事業所漢字リストテープや事業所データテープに含まれる現行事業所IDや回顧事業所IDを用いた2004年から1996年までの事業所・企業統計調査パネルデータの作成においては、事業所のレコードについての複数対応のケース、つまり、当該調査時の複数の事業所が前回調査時の単一事業所へ接続するような回顧事業所IDが重複するケースについては、例外的ともいえるほどその件数は少なかったが、1996コンスタントを用いた1994年と1991年においては非常に多くの複数対応のケースが見られた点は注意を要する。

事業所・企業統計調査のマイクロデータのファイルの接合作業の概要を表10に、データテープおよび漢字リストテープの接合状況を図2に示す。また、本稿で作成した事業所・企業統計パネルデータにおいて、2004年調査時点で当該パネルデータに含まれる事業所のレコードを起

表 9. 事業所・企業統計調査のパネルデータ作成のためのファイルの接合手順.

#	接合したファイル	接合に用いた事業所 ID	
1	2004 漢字リスト		
2	2001 漢字リスト	(1) 2001 漢字リストの 2001 現 ID	(1) 2004 漢字リストの 2001 回顧 ID
3	2004 データ	(1) 2004 データの 2004 現 ID	(1) 2004 漢字リストの 2004 現 ID
		(2) 2004 データの 2001 回顧 ID	(2) 2001 漢字リストの 2001 現 ID
4	2001 データ	(1) 2001 データの 2001 現 ID	(1) 2001 漢字リストの 2001 現 ID
		(2) 2001 データの 2001 現 ID	(2) 2004 漢字リストの 2001 回顧 ID
		(3) 2001 データの 2001 現 ID	(3) 2004 データの 2001 回顧 ID
5	1999 データ	(1) 1999 データの 1999 現 ID	(1) 2001 漢字リストの 1999 回顧 ID
		(2) 1999 データの 1999 現 ID	(2) 2001 データの 1999 回顧 ID
6	1996 データ	(1) 1996 データの 1996 現 ID	(1) 2001 漢字リストの 1996 回顧 ID
		(2) 1996 データの 1996 現 ID	(2) 1999 データの 1996 回顧 ID
7	1994 データ	(1) 1994 データに付与済みの 1996 コンスタント中の 1994-ID	(1) 1996 データに付与済みの 1996 コンスタント中の 1996-ID
8	1991 データ	(1) 1991 データに付与済みの 1996 コンスタント中の 1991-ID	(1) 1994 データに付与済みの 1996 コンスタント中の 1994-ID

注) 表中では本文中の言葉を次のように略記した.

「漢字リストテープ」を「漢字リスト」, 「データテープ」を「データ」, 「現行事業所 ID」を「現 ID」, 「回顧事業所 ID」を「回顧 ID」, 「1994 事業所 ID」を「1994-ID」, 「1991 事業所 ID」を「1991-ID」.

点として過去に遡ることができる事業所のレコードを表 11 に示す.

5. 法人企業統計調査と事業所・企業統計調査の完全照合

ここまで法人企業統計調査と事業所・企業統計調査それぞれのパネルデータ作成について解説してきたが, それぞれのパネルデータを接続することができれば, 法人企業の財務情報と, その傘下の事業所の詳細な情報を接続することができる. しかし, 相互の統計調査の個票データを接合する情報, いわゆるリンクコンスタントが無いため, 独自に接続を試みることにした.

名称や所在地を漢字データで照合可能でかつ同一調査年が存在するのは 2004 年であることから, 今回の作業で使用したデータおよび照合対象レコード数は, それぞれ, パネル ID 付与済み法人企業統計調査 2004 年度調査名簿データ (70,639 件), パネル ID 付与済み事業所・企業統計調査 2004 年漢字リストテープ (1,513,098 件) である.

5.1 準備作業

照合作業に先立ち, 相互の比較項目に対する共通化作業や, 文字列情報の調整作業を行った. 詳細を以下に示す.

1) 照合項目における表記・コードの統一

名称

法人名簿の法人名称は基本的に会社組織名(株式会社, 有限会社, 等)を含まないが, 事業所企業の正式名称ではこれらを含むため, 事業所名称から分離した. また, 本社事業所名称に特有の語句(本社, 本部, 本店, 等)も併せて除去した.

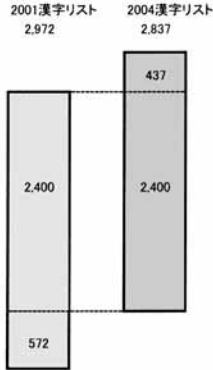
所在地

番地表記の統一. 丁目, 番地, 号はハイフン区切りに.

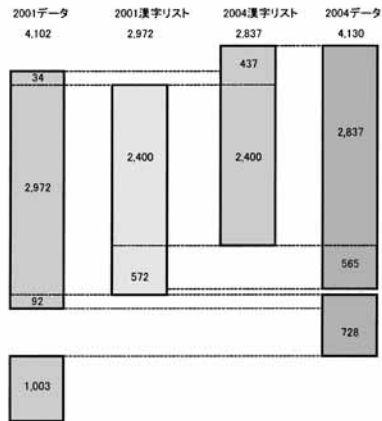
文字列共通

拗音・促音は大書き, 長音記号の統一, 空白や記号(ピリオド, 中黒, 等)の除去.

2004年、2001年事業所漢字リスト・リンク件数(千件)



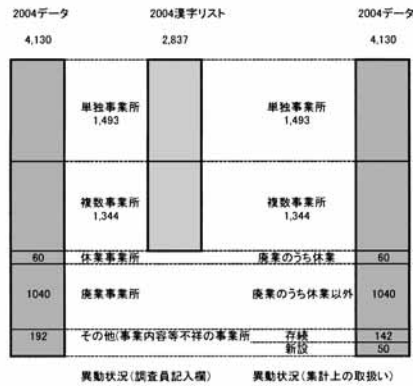
2004年、2006年事業所データ・リンク件数(千件)



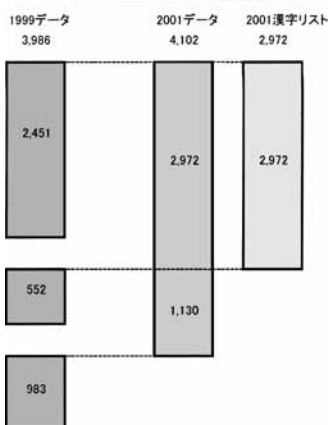
2001年事業所データ・リンク件数(千件)



2004年事業所データ・リンク件数(千件)



1999年事業所データ・リンク件数(千件)



1996事業所データ・リンク件数(千件)

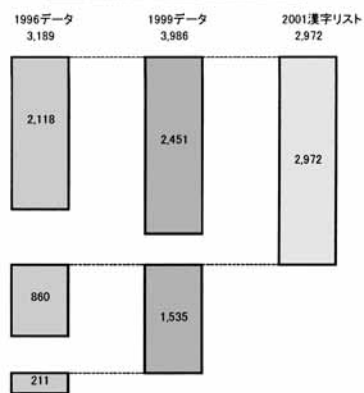


図 2. 事業所データテーブルの接合状況.

表 10. 事業所・企業統計調査のパネルデータ作成のためのファイル接合状況概要.

#	接合したファイル	レコード	接合に用いた事業所 ID	件数
1	2004 漢字リスト	2,836,728		
2	2001 漢字リスト	2,971,593	(1) 2004 漢字リストの 2001 回顧 ID (接続できなかった件数) (複数対応のケース件数)	2,399,581 572,012 0
3	2004 データ	4,130,404	(1) 2004 漢字リストの 2004 現 ID (2) 2001 漢字リストの 2001 現 ID (接続できなかった件数) (複数対応のケース件数)	2,836,728 565,188 728,497 9
4	2001 データ	4,101,788	(1) 2001 漢字リストの 2001 現 ID (2) 2004 漢字リストの 2001 回顧 ID (3) 2004 データの 2001 回顧 ID (接続できなかった件数) (複数対応のケース件数)	2,971,593 34,493 92,313 1,003,427 38
5	1999 データ	3,985,526	(1) 2001 漢字リストの 1999 回顧 ID (2) 2001 データの 1999 回顧 ID (接続できなかった件数) (複数対応のケース件数)	2,450,995 551,550 983,009 28
6	1996 データ	3,189,284	(1) 2001 漢字リストの 1996 回顧 ID (2) 1999 データの 1996 回顧 ID (接続できなかった件数) (複数対応のケース件数)	2,117,642 860,141 211,519 18
7	1994 データ	7,412,223	(1) 1996 データに付与済みの 1996 コンスタント中の 1996-ID (接続できなかった件数) (複数対応のケース件数)	2,678,201 4,808,985 74,963
8	1991 データ	2,965,958	(1) 1994 データに付与済みの 1996 コンスタント中の 1994-ID (接続できなかった件数) (複数対応のケース件数)	2,535,275 492,985 62,302

注 1) 表中では本文中の言葉を次のように略記した.

「漢字リストテープ」を「漢字リスト」, 「データテープ」を「データ」, 「現行事業所 ID」を「現 ID」, 「回顧事業所 ID」を「回顧 ID」, 「1994 事業所 ID」を「1994-ID」, 「1991 事業所 ID」を「1991-ID」.

注 2) 表頭に「接合に用いた事業所 ID」とした欄の事業所 ID は、「接合したファイル」の欄のファイルに含まれる事業所 ID と照合する相手のものを示し、1 つ前の接合作業により得られたファイルに含まれる事業所 ID を示す.

表 11. 2004 年時点から遡ることのできる事業所のレコード数: 事業所・企業統計調査パネルデータ.

調査年	民営の法人事業所 (休・廃業など含む)		うち本所・支所の別が判明		うち経営組織が判明	
	事業所数	割合	事業所数	割合	事業所数	割合
2004	4,130,413	100.0%	2,836,717	100.0%	2,836,717	100.0%
2001	3,091,575	74.8%	2,399,581	84.6%	2,399,581	84.6%
1999	2,512,423	60.8%	2,012,167	70.9%	2,012,167	70.9%
1996	2,151,370	52.1%	1,749,279	61.7%	1,749,279	61.7%
1994	1,866,907	45.2%	1,536,675	54.2%
1991	1,551,755	37.6%	1,286,092	45.3%

注 1) 表頭の「割合」とは 2004 年調査時点での事業所のレコード数に対する割合.

注 2) 1994 年調査では経営組織を調査していないので、経営組織が判明した事業所数は 1996 年を越えて遡ることはできない.

資本金階級

法人名簿の規模コードと事業所企業の資本金規模コードは区分が合致しないため、法人名簿の調査時資本金額を用いて事業所企業の資本金階級に統一した。

2) 比較対象レコードの絞り込み

名寄せ処理を効率化するために、あらかじめ候補となる事業所を絞り込んでおく。

- 本支の別が単独事業所または本所事業所
- 経営組織が会社法人
- 産業分類コードが金融・保険業(大分類 = K)を除く

3) 重複調査法人の除去

事業所企業との多重リンクを避けるために、法人名簿に存在する重複法人をあらかじめ除去しておく。法人名簿における活動状況区分コード別の回答状況を調査したところ、回答があった活動状況区分は活動法人と移転法人に限られていた。このため、それ以外の活動区分(重複・休廃業、等)は照合対象から外した。それでもなお法人名および所在地が一致する法人については、当該年次の回答率やパネルの長さを選択基準とした。

5.2 照合作業

所在地に関する照合条件を表 12、全体の照合順序とその条件を表 13 に示すように、厳しい照合条件から徐々に緩めて照合を行った。

なお、法人名簿は 2002 年以前は半角カタカナだった影響か、名称および住所が全角カタカナで入っているものがあるため、事業所名のフリガナとも照合する必要がある。名称の照合については、事業所企業の正式名称には、(株)などの会社形態、店名、部門名などが企業名に加えて入っていることが多いため、部分一致(事業所名称が法人名称を包含)するかどうかや、事業所の通称名との一致についても調べた。

所在地の照合については、法人名簿の所在地は市区郡名からビル名まで単一項目に入っており、都道府県名は基本的には含まれない。市区町村名や番地等のデリミタとして空白が入っている場合もあるが、規則が統一されていないため、事業所・企業統計の所在地情報(都道府県、市区町村、町丁字番地号、ビル名等)の各項目に対応するように分割することはできなかった。このため、所在地の照合では法人名簿の所在地において事業所企業の町丁字番地号やビル名が含まれているかどうかで判定を行った。また郵便番号も照合対象とした。

法人企業統計では、法人番号 40000 未満の管理法人を対象に 1 年間で年次調査および四半期調査(4 回)の計 5 回の調査が行われており、それゆえ 5 つの調査名簿がある。加えて、本店住所と連絡先住所の 2 種類があり、調査年次の途中で移転や法人名称の変更があることを考慮するため、それぞれの組み合わせでも照合できるように配慮する必要があった。

プログラムによる接続結果の件数を表 14 に示す。

最終的に、緩い照合条件下で接続したものについては目視による確認を行い、誤った接続については修正を行った。目視確認の結果については後述する。

6. 結果および評価

ここまで法人企業統計調査と事業所・企業統計調査のパネルデータの作成の概要と、プログラムによる完全照合の手法と結果について述べてきたが、照合や判定処理を行ったプログラムの有効性を評価するためには、処理結果を人の目でチェックすることが不可欠である。今回の作業では法人企業統計の管理法人パネルの作成および法人企業統計と事業所・企業統計の完全

表 12. 所在地の一致条件.

所在地 Lv.	ステップ	所在地の一致度合い
1	a	町丁番地号およびビル名が一致する
	b	町丁番地号が一致する
	c	郵便番号（7桁）が一致する
2	d	市区町村名が一致する
	e	郵便番号（5桁）が先頭一致する
	f	郵便番号（3桁）が先頭一致する

表 13. 法人企業統計調査と事業所・企業統計調査の照合条件.

照合 Lv.	所在地 Lv.	ステップ	名称および資本金規模の一致度合い
1	1	a,b,c	法人名＝事業所名&資本金規模が一致
2	1	a,b,c	法人名＝事業所名
3	1	a,b,c	法人名＝事業所ヨミ&資本金規模が一致
4	1	a,b,c	法人名＝事業所ヨミ
5	1	a,b,c	法人名末尾なし&法人名㊦事業所名&資本金規模が一致
6	1	a,b,c	法人名末尾なし&法人名㊦事業所名
7	1	a,b,c	法人名末尾なし&法人名㊦事業所ヨミ&資本金規模が一致
8	1	a,b,c	法人名末尾なし&法人名㊦事業所ヨミ
9	1	a,b,c	法人名＝事業所通称名, 資本金規模が一致
10	1	a,b,c	法人名＝事業所通称名
11	1	a,b,c	法人名末尾なし&法人名㊦事業所通称名&資本金規模が一致
12	1	a,b,c	法人名末尾なし&法人名㊦事業所通称名
13	2	d,e,f	法人名＝事業所名&資本金規模が一致
14	2	d,e,f	法人名＝事業所名
15	2	d,e,f	法人名＝事業所ヨミ&資本金規模が一致
16	2	d,e,f	法人名＝事業所ヨミ
17	2	d,e,f	法人名末尾なし&法人名㊦事業所名&資本金規模が一致
18	2	d,e,f	法人名末尾なし&法人名㊦事業所名
19	2	d,e,f	法人名末尾なし&法人名㊦事業所ヨミ&資本金規模が一致
20	2	d,e,f	法人名末尾なし&法人名㊦事業所ヨミ
21	2	d,e,f	法人名＝事業所通称名, 資本金規模が一致
22	2	d,e,f	法人名＝事業所通称名
23	2	d,e,f	法人名末尾なし&法人名㊦事業所通称名&資本金規模が一致
24	2	d,e,f	法人名末尾なし&法人名㊦事業所通称名

照合において文字列照合による接続処理を行っているため、それぞれにおける判定条件を緩和した場合の処理結果について、名称・所在地を含む照合項目をリストアップして目視による確認作業を行う必要があった。

目視作業では、プログラムの照合に使用しなかった属性項目やインターネットの情報などを用いて判定作業を行っており、プログラムによる判定が疑わしいと目視判断した場合は誤判定とした。最終的にプログラムによる処理結果に目視作業による修正を反映させた接続情報を元に企業データベースを構築した。

以下では目視確認によるプログラム処理結果の訂正作業を通じてプログラムの有効性を評価

表 14. 2004 年時点における法人企業統計と事業所企業統計の 1 対 1 接続の結果件数表 (目視修正前, 低階層法人を含む).

照合 Lv.	ステップ						total
	所在地 Lv.=1			所在地 Lv.=2			
	a	b	c	d	e	f	
1	769	28329	3150	.	.	.	32,248
2	64	1393	251	.	.	.	1,708
3	18	312	516	.	.	.	846
4	2	32	144	.	.	.	178
5	114	1343	161	.	.	.	1,618
6	13	93	21	.	.	.	127
7	13	120	154	.	.	.	287
8	0	13	30	.	.	.	43
9	2	39	3	.	.	.	44
10	0	7	3	.	.	.	10
11	0	3	1	.	.	.	4
12	0	0	0	.	.	.	0
13	.	.	.	924	61	82	1,067
14	.	.	.	204	22	36	262
15	.	.	.	19	54	44	117
16	.	.	.	26	24	24	74
17	.	.	.	51	7	6	64
18	.	.	.	8	0	2	10
19	.	.	.	5	11	12	28
20	.	.	.	1	3	2	6
21	.	.	.	2	1	2	5
22	.	.	.	12	2	1	15
23	.	.	.	0	0	0	0
24	.	.	.	1	0	0	1
total	995	31,684	4,434	1,253	185	211	38,762

する。また、作成した企業データベースを用いて、事業所・企業統計調査(2001年)の名寄せ済みデータ(本所事業所とその支所事業所, 親子企業)と同年の法人企業統計調査との接続を試みた結果についても述べる。

6.1 法人企業統計の管理法人パネルデータ

法人企業統計の管理法人(資本金1億円以上)のパネルデータをプログラムで作成した結果については、照合条件を緩和して継続法人と判定したもの(判定コードが5~8)と、新規法人と判定したもの(判定コード=9)について、目視による処理結果の確認作業を行った。プログラムによる判定結果の目視訂正件数を表15に示す。

2002年以前の半角文字情報から2003年以降の全角漢字情報に切り替わるところで判定精度がかなり悪くなることを予想していたが、2003年(前年の2002年との比較)よりも2004年(前年の2003年との比較)の方がプログラムによる判定結果が悪く、訂正件数が多いことが分かる。これは、2003年の法人名簿データが2002年の半角文字を全角文字に変換しただけのものが多く含まれていたため、2004年以降に漢字データへ移行したケースが目立ったことが理由としてあげられる。

表 15. 法人企業統計管理法人パネルにおける目視作業結果.

調査年	接続 (判定コード=5～8)			非接続 (判定コード=9)		
	訂正前	目視訂正	訂正率	訂正前	目視訂正	訂正率
1983	0	0	-	0	0	-
1984	3	2	67%	19	3	16%
1985	5	4	80%	16	1	6%
1986	4	4	100%	12	7	58%
1987	9	8	89%	21	4	19%
1988	14	13	93%	28	3	11%
1989	19	14	74%	20	5	25%
1990	14	11	79%	17	5	29%
1991	17	14	82%	17	1	6%
1992	22	15	68%	16	2	13%
1993	24	17	71%	13	2	15%
1994	33	25	76%	25	3	12%
1995	17	14	82%	26	4	15%
1996	25	22	88%	34	7	21%
1997	24	21	88%	23	1	4%
1998	21	15	71%	20	3	15%
1999	30	21	70%	16	2	13%
2000	34	29	85%	25	0	0%
2001	38	34	89%	25	3	12%
2002	64	61	95%	1,954	0	0%
2003	3,338	326	10%	2,408	175	7%
2004	9,477	92	1%	3,104	1,495	48%
2005	208	61	29%	2,557	15	1%
合計	13,440	823	6%	10,396	1,741	17%

注) 非接続 (判定コード=9) の訂正前件数では、前年の調査名簿において当該法人番号が欠番だったケースは除いている。

以上の作業によって、1983～2005年の23年分の調査名簿データを用いて、各管理法人にユニークな識別IDである法人パネルIDを付与できた。パネルの接続年数の分布を調べた結果を表16に示す。ここでのパネル長は接続年数を意味しており、パネル長=23となっている約14%の法人企業が調査名簿に23年間掲載されていることがわかる。

6.2 法人企業統計調査と事業所・企業統計との接続結果

法人企業統計調査と事業所企業統計調査の2004年調査データを用いたプログラムによる完全照合の結果については、表14、条件については表12、表13を参照されたい。

照合条件を緩めに設定した照合結果のうち、Lv.5以降については、名称及び所在地情報等の属性の組み合わせをリストアップし、プログラムによる判定結果を目視にて確認する作業を行った結果を表17に示す。照合レベルLv.5～12では誤判定は無かったが、所在地の照合条件を緩和したLv.13以降において約2割の誤判定が生じる結果となった。しかし、条件緩和によって残り約8割の接続結果を拾うことができたことは評価すべきと考えられる。

事業所・企業統計調査の漢字リストテープは調査名簿ではなく回答データであるため、接続できた事業所データからは調査票の回答項目の情報を得ることができる。一方、法人企業統計の照合データは調査名簿なので、回答(調査票)データが存在するとは限らず、回答項目の情報

表 16. パネルの長さ(接続年数)別法人数の分布.

パネル長	度数	%	累積度数	累積%
1	13,940	19.06	13,940	19.06
2	6,983	9.55	20,923	28.61
3	4,845	6.63	25,768	35.23
4	4,377	5.99	30,145	41.22
5	3,883	5.31	34,028	46.53
6	2,849	3.90	36,877	50.43
7	2,507	3.43	39,384	53.85
8	2,382	3.26	41,766	57.11
9	2,000	2.73	43,766	59.85
10	1,769	2.42	45,535	62.26
11	1,638	2.24	47,173	64.50
12	1,578	2.16	48,751	66.66
13	1,820	2.49	50,571	69.15
14	1,905	2.60	52,476	71.76
15	1,754	2.40	54,230	74.15
16	1,569	2.15	55,799	76.30
17	1,573	2.15	57,372	78.45
18	1,212	1.66	58,584	80.11
19	1,209	1.65	59,793	81.76
20	1,001	1.37	60,794	83.13
21	1,383	1.89	62,177	85.02
22	803	1.10	62,980	86.12
23	10,152	13.88	73,132	100.00

※数値はパネル数 (=ユニークな法人数)

を得るためには調査票データとの接続が必要となる。

相互の接続結果について、法人企業統計の調査名簿(2004年)における資本金別および業種別の法人数の分布をそれぞれ表 18 に示す。この内で年報調査票の回答データが存在する法人数の分布を表 19 に併せて示す。法人企業統計の調査名簿のレベルでは全体として5~6割の法人企業が接続できており、回答データである調査票(年報)データのレベルでは7割前後の接続率であった。

6.3 法人企業統計と事業所・企業統計名寄せデータとの接続

法人企業統計調査と事業所・企業統計調査の接続は、調査年次と照合する名称や所在地の文字データの都合から2004年時点のみの照合となった。2004年調査は残念ながら事業所・企業統計調査は簡易調査であり、本所事業所とその支所事業所や親子関係にある企業について名寄せを行うために必要な名称・所在地に関する設問がないので、名寄せ集計自体を行うことができない。

図3に今回の照合状況(予定を含む)を示すが、本所・支所事業所や親子企業の名寄せ済みデータが存在する近傍の事業所・企業統計調査は大規模調査年である2001年である。しかし、法人企業統計の調査名簿データが漢字ではないため、完全照合による接続は困難であるが、パネルデータの接続と2004年の相互接続を経由すれば、疑似的に2001年での相互接続が可能である。

この手法を用いて、2001年時点での相互接続を試みた。事業所・企業統計の2001年調査に

表 17. 2004 年時点における法人企業統計と事業所企業統計の 1 対 1 リンケージの目視訂正件数表.

level	step	目視修正前	目視訂正	訂正率
13	d	924	70	8%
13	e	61	12	20%
13	f	82	13	16%
14	d	204	137	67%
14	e	22	14	64%
14	f	36	25	69%
15	d	19		0%
15	e	54		0%
15	f	44		0%
16	d	26	26	100%
16	e	24	19	79%
16	f	24	12	50%
17	d	51	6	12%
17	e	7		0%
17	f	6		0%
18	d	8	1	13%
18	f	2	1	50%
19	d	5		0%
19	e	11	3	27%
19	f	12	1	8%
20	d	1		0%
20	e	3	2	67%
20	f	2		0%
21	d	2	2	100%
21	e	1	1	100%
21	f	2	1	50%
22	d	12	12	100%
22	e	2	2	100%
22	f	1	1	100%
24	d	1	1	100%
合計		1,649	362	22%

注) 目視修正前はプログラムによる接続件数
目視訂正はプログラムの誤判定件数
照合レベル 12 以下については目視訂正件数ゼロ

における本所・支所事業所および親子企業名寄せ済みデータは同調査年の漢字リストテープおよび確報データテープと共通の事業所 ID を持つため、2004 年調査データとは回顧事業所 ID を用いれば直接接続することができる。

ただし、ここでは 2 つの留意すべき問題がある。1 つは 2001 年調査と 2004 年調査の経年による事業所の属性変化である。例えば、2004 年調査を用いた法人企業統計と事業所・企業統計調査の接続作業において、事業所企業の照合対象は単独事業所と本所事業所に限定し、支所事業所を対象としなかったが、回顧事業所 ID で 2001 年に遡って接続すると、当該事業所が支所事業所であったケースが存在した。(2004 年時点で法人企業統計の調査名簿と接続した本所事業所または単独事業所で、2001 年調査への回顧事業所 ID を持つ事業所は 9,045 件あり、そのうち 212 件は 2001 年調査では支所事業所であった。)

同一企業のまま当該事業所の位置付けが変わったのであればその本所事業所と法人企業を接続すればよいが、異なる企業に変わった場合はそのまま接続できないことになり、これらの判断をプログラムで機械的に行うことは難しい。

もう 1 つは補完事業所の存在である。これは、名寄せ処理の過程で判明した回答データの無い重要な事業所についての補完があり、2004 調査から回顧事業所 ID を用いてこれらの補完事

表 18. 2004 年の相互リンケージ結果— 調査名簿におけるリンク数および割合：資本金規模別(上)、業種別(下).

規模コード	管理法人			低階層法人(年報)			低階層法人(季報)		
	全体	リンク	割合	全体	リンク	割合	全体	リンク	割合
200万円未満	0	0		2,322	775	33%	0	0	
200万円以上,300万円未満	0	0		1,332	574	43%	0	0	
300万円以上,500万円未満	2	0	0%	2,446	977	40%	0	0	
500万円以上,1000万円未満	0	0		2,368	1,217	51%	0	0	
1000万円以上,2000万円未満	125	61	49%	4,141	2,183	53%	4,062	2,130	52%
2000万円以上,5000万円未満	129	67	52%	4,379	2,746	63%	4,344	2,751	63%
5000万円以上,1億円未満	383	235	61%	3,020	1,794	59%	3,009	1,876	62%
1億円以上,10億円未満	32,182	16,860	52%	57	33	58%	42	32	76%
10億円以上	6,290	4,145	66%	4	4	100%	3	3	100%
合計	39,111	21,368	55%	20,069	10,303	51%	11,460	6,792	59%

業種(13分類)	管理法人			低階層法人(年報)			低階層法人(季報)		
	全体	リンク	割合	全体	リンク	割合	全体	リンク	割合
農林水産業	156	67	43%	737	331	45%	429	215	50%
鉱業	168	61	36%	254	143	56%	175	97	55%
建設業	2,207	1,531	69%	2,563	1,615	63%	1,663	1,211	73%
製造業	10,141	6,782	67%	5,578	3,407	61%	3,290	2,237	68%
卸売・小売業	8,734	4,438	51%	3,694	2,053	56%	1,964	1,207	61%
飲食店	448	240	54%	494	212	43%	202	83	41%
不動産業	4,478	1,882	42%	1,471	427	29%	767	252	33%
情報通信業	3,401	1,813	53%	515	134	26%	257	111	43%
運輸	1,842	1,211	66%	947	475	50%	599	361	60%
電気業	99	41	41%	98	15	15%	52	6	12%
ガス・熱供給・水道業	208	146	70%	101	64	63%	78	56	72%
宿泊業	889	527	59%	272	137	50%	158	82	52%
サービス業	6,340	2,629	41%	3,345	1,290	39%	1,826	874	48%
合計	39,111	21,368	55%	20,069	10,303	51%	11,460	6,792	59%

業所データに接続することができない。(重要な事業所とは、共通の本所事業所を持つ多数の支所事業所が存在するも名寄せすべき本所事業所の回答データが無い場合や、多数の子会社を持つ事業所企業で自身も子会社であるが名寄せすべき親会社の回答データがない場合のように、重要な事業所企業の回答データがみつからなかった場合に事業所データの補完が行われている。)

実際に、2004年調査における新設事業所(回顧事業所IDを持たない事業所)と2001年調査の名寄せ済みデータにおける補完事業所を接続することを試みたが、これまでに述べてきた同様の手法を用いて事業所名称と所在地の文字情報で照合したところ、結果として、429件の補完事業所(2001年)に対して97件の新設事業所(2004年)と接続することができた。ただし、事業所の経年変化に関しては考慮することができないので、ここでの問題は事業所名称および所在地が2001年から2004年の間で変わらなかった事業所に限定された接続になってしまうことである。

2004年時点の法人企業統計と事業所・企業統計の相互接続情報を用いてそれぞれの統計調査のパネルデータを接続し、さらに補完事業所についても加味した上で、2001年時点の法人企業統計調査と事業所・企業統計調査の名寄せ済みデータとを接続した。法人企業統計の回答データ(年報調査データ、2001年)がある法人企業について、業種別資本金階層別の法人数の分布を

表 19. 2004 年の相互リンケージ結果—年報調査票におけるリンク数および割合：資本金規模別(上), 業種別(下).

資本金階級	管理法人			低階層法人(年報)		
	全体	リンク	割合	全体	リンク	割合
200 万円未満	0	0		991	567	57%
200 万円以上,300 万円未満	0	0		604	397	66%
300 万円以上,500 万円未満	1	0	0%	1,309	732	56%
500 万円以上,1000 万円未満	0	0		1,346	893	66%
1000 万円以上,2000 万円未満	84	43	51%	2,479	1,716	69%
2000 万円以上,5000 万円未満	77	41	53%	3,033	2,289	75%
5000 万円以上,1 億円未満	275	181	66%	2,142	1,568	73%
1 億円以上, 10 億円未満	8,174	5,948	73%	56	33	59%
10 億円以上	5,271	3,892	74%	4	4	100%
合計	13,882	10,105	73%	11,964	8,199	69%

業種(13 分類)	管理法人			低階層法人(年報)		
	全体	リンク	割合	全体	リンク	割合
農林水産業	59	36	61%	453	273	60%
鉱業	88	42	48%	163	117	72%
建設業	751	620	83%	1,694	1,340	79%
製造業	4,661	3,660	79%	3,447	2,682	78%
卸売・小売業	2,806	1,861	66%	2,204	1,634	74%
飲食店	153	112	73%	224	147	66%
不動産業	1,264	823	65%	858	330	38%
情報通信業	1,153	864	75%	248	105	42%
運輸	796	611	77%	602	393	65%
電気業	62	35	56%	41	10	24%
ガス・熱供給・水道業	104	81	78%	94	62	66%
宿泊業	312	239	77%	150	103	69%
サービス業	1,673	1,121	67%	1,786	1,003	56%
合計	13,882	10,105	73%	11,964	8,199	69%

表 20 に示す. 2004 年時点では低階層法人(資本金 1 億円未満)への接続が含まれていたが, 法人企業統計調査のパネルデータでは低階層法人部分は標本調査であるために接続率が非常に悪く, 結果的に管理法人(資本金 1 億円以上)にほぼ限定されてしまうことがわかる.

7. まとめと課題

今回の作業で法人企業統計調査と事業所・企業統計調査それぞれのパネルデータの作成, および 2004 年時点での相互接続を行った. 加えて, 事業所・企業統計調査の大規模調査年である 2001 年時点での相互接続も試みた. 事業所・企業統計の 2001 年調査における本所事業所とその配下にある支所事業所の名寄せ情報と法人企業統計調査の法人名簿との接続結果を用いることで, 法人企業を事業所概念で捉えることが可能になる. また, 同様に親子企業名寄せの結果と法人名簿のリンケージ結果を用いれば, 近年関心が持たれている連結企業に関する調査に対応するための名簿整備にも役立てることができる.

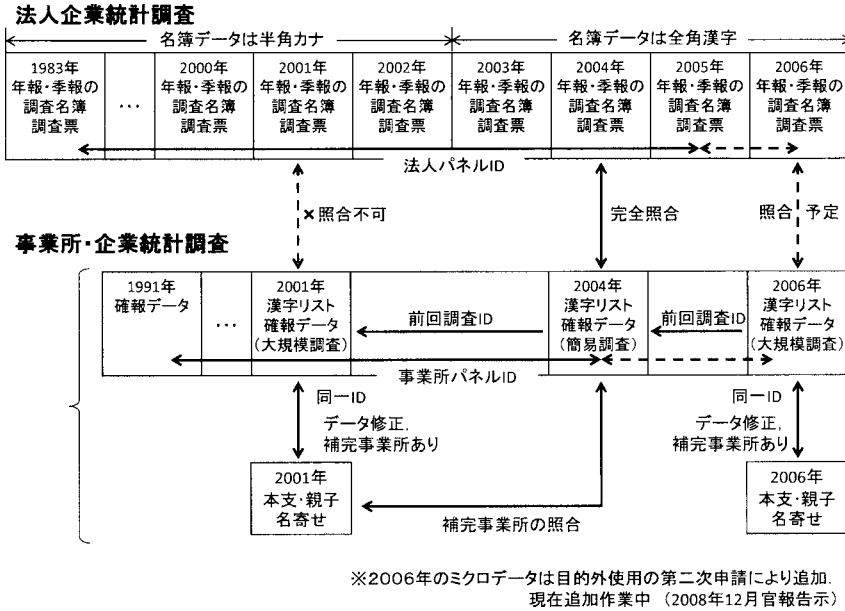


図 3. 企業データベースにおける照合関係。

表 20. 2001 年の相互リンク結果 — 年報調査票 (業種 × 資本金規模)。

業種 (10分類)	資本金規模			合計
	5千万円以上, 1億円未満	1億円以上, 10億円未満	10億円以上	
農林水産業	0	14	8	22
鉱業	0	6	20	26
建設業	0	155	188	343
製造業	0	898	1,403	2,301
卸売・小売業	0	517	449	966
不動産	0	198	223	421
運輸通信	0	204	221	425
電気	0	2	24	26
ガス・水道	0	23	25	48
サービス	1	512	546	1,059
合計	1	2,529	3,107	5,637

ただ、現状では相互の調査年が一致し、かつ、名寄せ作業で使用する項目(名称、住所)が全角情報であるマイクロデータが2004年調査のみであるため、この一時点での完全照合になる。2004年の事業所・企業統計調査は簡易調査であるため、法人企業との接続によるメリットが限定的である。また、2001年時点の相互接続は、2004年の相互接続時点からそれぞれのパネルを経由した間接的・疑似的な接続である。

パネルデータの構築に関する問題点としては、事業所企業統計は悉皆調査であるが、照合で

使用した事業所漢字リストテープは、回答データであるため、未回答を含む事業内容不詳の事業所は名称や所在地の情報が無いので照合できない。事業所・企業統計調査のパネルデータに関しても、実際には存続事業所であったにもかかわらず、回答データが無いまたは移転のために、前後の調査年と接続できないケースがある。

法人企業統計は調査名簿を使用しているので回答の有無に影響されないが、悉皆調査対象法人(資本金10億円以上、実質6億円以上)を除けば標本調査であるため、毎年の標本替えによる影響がある。また、法人企業統計のパネルデータについては、管理法人(資本金1億円以上)についてはある程度の年数を連続して調査対象になっていることに加え、年報・季報の両調査対象にもなっているため、時系列分析にも耐えるデータベースを構築することができた。一方、低階層法人(資本金1億円未満)については、母集団が少ない特定業種の法人を除き、負担軽減の観点から連続する複数調査年で調査対象にはならず、年報・季報の調査についても同一調査年ではいずれか一方のみ調査されるため、パネルデータとしての集計用途には不向きである。

今後の課題としては、統計法に基づく目的外使用の追加申請によって利用可能となった2006年までのマイクロデータを用いて、2006年時点における法人企業統計調査と事業所・企業統計調査の相互接続である。事業所・企業統計調査の2006年調査は大規模調査であり、実現すれば、本所・支所の名寄せデータや企業の親子名寄せデータを法人企業統計の財務情報と直結して利用可能になる。さらに、大規模調査年である2001年と2006年の時点間の事業所・企業統計調査と法人企業統計との接続結果を用いて比較分析することができる。

謝 辞

本研究の成果は、日本学術振興会科学研究費補助金(研究種目：基盤研究(A)、課題番号：17203017、研究課題：法人企業統計調査と事業所・企業統計調査のマイクロデータの統合新統計の編成と解析研究、研究代表者：松田芳郎)のもとで、統計法第15条2項に基づきマイクロデータの目的外使用の許可を指定統計「法人企業統計」(官報第4562号総務省告示第240号)と指定統計「事業所統計」及び「事業所・企業統計」(官報第4648号総務省告示第463号)について頂いたことにより得られました。ここに関係各位に対し深く感謝しお礼申し上げます。とりわけ、研究代表者でもある松田芳郎青森公立大学教授には丁寧にご指導頂きました。また、プログラム処理による照合結果の目視確認作業では(財)統計情報研究開発センターの中川雅義参与と石井達男普及部長には大変お世話になりました。さらに当科学研究費の全体研究集会(2007年10月18日開催)で報告を行い研究会参加者からは貴重な助言をいただき改稿の機会を得ました。すべてのコメントを本稿に生かすことはできませんでしたが、ここに重ねて深く感謝しお礼申し上げます。なお本稿におけるすべての誤謬は筆者に帰するものです。

参 考 文 献

- 松田芳郎(1991). 『企業構造の統計的測定方法』, 岩波書店, 東京.
- Ruggles, R. and Ruggles, N. D. (1999). The analysis of longitudinal establishment data, *Macro- and Microdata Analyses and Their Integration* (eds. N. D. Ruggles and R. Ruggles), 235-267, Edward Elgar Publishing, Northampton, Massachusetts. (First appeared as a presentation at the Bureau of the Census and NSF Conference on Longitudinal Establishment Data File and Diversification Study, Alexandria, Virginia, October 17-18, 1984)
- 総務庁統計局(1992). 『平成3年事業所統計調査報告』, 総務庁統計局, 東京.
- 総務庁統計局(1995). 『平成6年事業所名簿整備調査報告』, 総務庁統計局, 東京.

- 総務庁統計局(1998). 『平成8年事業所統計調査報告』, 総務庁統計局, 東京.
- 総務庁統計局(2000). 『平成11年事業所統計調査報告』, 総務庁統計局, 東京.
- 総務省統計局(2003). 『平成13年事業所統計調査報告』, 総務省統計局, 東京.
- 総務省統計局(2006). 『平成16年事業所統計調査報告』, 総務省統計局, 東京.
- 菅 幹雄, 宮川幸三(2008). 『アメリカ経済センサス研究』, 慶應義塾大学出版会, 東京.
- Wallgren, Anders and Wallgren, Britt (2007). *Register-based Statistics Administrative Data for Statistical Purposes*, Wiley Series in Survey Methodology, John Wiley and Sons, England.
- 財務省財務総合政策研究所研修部(2006). 『平成18年度研修部教材(9)法人企業統計実務』, 財務省財務総合政策研究所研修部, 東京.

Integrated Enterprise Database of “Financial Statements Statistics of Corporations by Industry” and “Establishment and Enterprise Census”:
1983–2005

Setsuo Suoh¹, Hiroki Furuzumi² and Tamaki Miyauchi³

¹Library and Academic Information Center, Kobe Gakuentoshi, University of Hyogo

²School of Economics, University of Hyogo

³Faculty of Economics, Keio University

We construct an Enterprise Database from microdata of two sets of statistics: “Financial Statements Statistics of Corporations by Industry” (1983–2005) and “Establishment and Enterprise Census” (1991–2004), permitted by the law of statistics. We make two panel datasets of these statistics and connect them to each other. We explain how to make this panel dataset and integrated enterprise database.