

『企業活動基本調査』パネル・データの作成・利用について：
経済分析への応用とデータ整備の課題*

松浦寿幸[†] 清田耕造[‡]

要旨

本論文は、近年、経済政策の研究に頻繁に利用されている経済産業省『企業活動基本調査』個票データによるパネル・データの有用性を検討するとともに、利用上の注意点、統計整備における課題を整理したものである。『企業活動基本調査』の個票データは永久企業番号を利用することで比較的簡単にパネル・データ化できることが知られている。しかし、調査項目の中には、業種分類や調査項目が時系列で微妙に変わっているものがあり、機械的に調査項目を接続してしまうと、調査項目の変化によって生じる現象を企業活動の変化として誤って捉えてしまう可能性があることがわかった。ただし、個々の調査項目の時系列的な継続性を吟味した上で各年度のデータをリンクしていくと、ほとんどの変数は、非常に安定的な動きを見せていることも明らかになった。この結果は、『企業活動基本調査』から作成されたパネル・データの信頼性の高さを確認するものであり、政策研究にも耐えうることを示唆するものである。

* 本論文は、経済産業研究所におけるプロジェクト「マイクロデータ開発研究」の研究成果の一部である。本論文を作成するにあたって、元橋一之氏（経済産業研究所、東京大学先端科学技術研究センター）、高橋睦春氏（経済産業研究所）からは有益なコメントを得た。また、経済産業省調査統計部企業統計室から多大な協力を頂戴した。深く謝意を表したい。なお本論文の内容や意見、および残る全ての誤りは、筆者に帰するものであり、経済産業研究所の公式見解を示すものではない。

[†] 経済産業研究所計量分析・データ室(E-mail: matsuura-toshiyuki@rieti.go.jp)

[‡] 経済産業研究所ファカルティフェロー、横浜国立大学経営学部(E-mail: kiyota@ynu.ac.jp)

1. はじめに¹

近年、経済産業省『企業活動基本調査』の個票データ（以下、METI データ）を利用した研究が活発になっている。METI データは日本初の企業レベルの大規模な統計データとして、これまでのマクロ、産業データではわからなかった様々な事実を次々に明らかにしている。この METI データの最大の特徴は企業レベルでのパネル化が可能になっている点だが、パネル化そのものは各研究者によって行われてきた。このため、パネル化に伴う問題や難しさはこれまでほとんど議論されて来なかった。

しかし、実際には、単に個票データを時系列に追跡するだけでパネル・データが完成するわけではない。たとえば、一見継続的に調査されているように見える調査項目も、調査年によって若干定義が変わっていることがある。したがって、どのような項目がどのように調査されているかを把握せず、機械的に調査項目を接続してしまうと、調査項目の変化によって生じる現象を企業活動の変化として誤って捉えてしまう可能性がある。また、これまでの METI データを利用した研究の中には、類似のテーマの研究であっても異なる分析結果が紹介されている場合があった。その違いは、分析手法や分析対象の違いだけでなく、パネル・データの作成手順の違いに起因している可能性もあるのである。METI データへのニーズが近年急速に高まっていることを踏まえると、今ここで METI データのパネル化にあたっての注意点、問題点を議論しておくことには、ある程度の意義があると考えられる。

本論文では、特にデータの利用者としての立場から、METI データを元にパネル・データを作成する上での注意点、問題点を議論する。本論文を通じて、METI データの特徴を明らかにするとともに、今後の統計整備に関する政策的課題を提示する。本論文の以下の構成は次の通り。次節では METI データの元になっている『企業活動基本調査』の特徴と意義を解説する。第三節では調査項目の時系列的な追跡状況を議論し、第四節では METI データで把握されているサンプル数（企業数）を確認する。第五節では経済分析で頻りに利用される主要変数の特性を解説し、最終節で本論文を締めくくる。

2. 『企業活動基本調査』の概要と特徴²

『企業活動基本調査』は企業活動の実態を明らかにするため H4 年度（1991 年度対象）³からはじめられた調査であり、H7 年度（1994 年度対象）から毎年調査が行われている。調査

¹ 本論文に用いられた個票データの使用に際しては、統計法に基づいた統計の目的外使用の許可を得ている。

² 本節の議論は木村・清田（2002）に依っている。

³ 「企業活動基本調査」は、毎年 6 月に前年度の経済活動状況を調査している。そのため、各年度の調査報告書では、原則として、その前年度の経済活動状況が報告されている。本稿では、特に断りが無い限り、調査実施時点（すなわち、報告書のタイトルの年次）を和暦で示し、調査対象となっている年次を「調査対象年度」と呼び西暦で示すものとする。

の範囲は、H13 年度調査（2000 年度対象）の場合、日本標準産業分類の大分類 D（鉱業）F（製造業）G（電気・ガス・熱供給・水道業）⁴、I（卸売・小売業、飲食店）⁵、J（金融・保険業）のうち小分類 663（クレジットカード業、割賦金融業）L（サービス業）の一部⁶、に属する事業所を有する企業のうち、従業員 50 人以上、かつ資本金または出資金 3000 万円以上の会社である⁷。この『企業活動基本調査』の個票データが METI データである。

METI データの特徴として、以下の四点が挙げられる。第一は、永久企業番号を有している点である。永久企業番号とはそれぞれの企業に固有に設けられる背番号のことであり、この番号を利用することで各企業を時系列的に追跡できるようになっている⁸。このように複数の経済主体の情報を時系列的に整備したデータ（クロス・セクション×時系列データ）はパネル・データ（またはロングチュージナル・データ）と呼ばれている。パネル・データを利用することで、政策変更前後での企業活動の変化を捉えられるようになるだけでなく、計量経済学により精緻な分析も可能になる。

第二は、企業を単位として情報を収集している点である。これまでにわが国で整備されてきたデータの多くは事業所ベースの統計調査であった。事業所ベースの統計調査の場合、立地と生産活動を細かく把握できるという利点があるが、たとえば、研究開発や海外進出など企業レベルの意思決定を伴う企業活動についての情報は得られないという弱点もある。生産・雇用・販売を含めた企業内組織の変革や企業間関係の再編成、企業活動のグローバル化の動機やインパクトを分析するには、事業所レベルの統計ではなく企業レベルのデータが不可欠となってくる。『企業活動基本調査』は企業を調査対象としているため、これまでの事業所統計ではわからなかった企業活動そのものを捉えることが可能になっているのである。

第三は、サンプルとなる企業の多さと調査項目の豊富さである。これまでの日本の企業レベルの分析では、有価証券報告書を用いた財務データによる分析がしばしば試みられてきたが、標本が上場大企業に限定されるなどの点で制約があった。企業活動基本調査は従業員 50 人以上、かつ資本金または出資金 3000 万円以上の企業がカバーされている（調査対象となる産業については後述）。また、調査項目としては、部門別の従業者数や 3 桁レベルの品目別売上高、輸出・輸入の状況、企業間取引状況、子会社・関連会社の保有状況などがあり、財務情報以外の企業活動についても詳細に捉えられるようになっている。

⁴ 中分類 37（熱供給業）及び 38（水道業）を除く。

⁵ 中分類 61（その他の飲食店）を除く。

⁶ 調査対象に含まれるサービス業については、H13 年調査報告書の「利用上の注意」の別表を参照のこと。

⁷ なお、調査の範囲のうち、G（電気・ガス・熱供給・水道業）J（金融・保険業）のうち小分類 663（クレジットカード業、割賦金融業）L（サービス業）の一部が、経済産業省企業活動基本調査規則の改正により 2001 年より追加された。

⁸ なお、個々の企業は全て番号で管理されているため、個票データの利用者は個別企業を特定することなく分析が可能となっている。データの利用者は企業の名称までは特定できない。

第四は、調査対象でカバーされている産業の多さである。調査対象となる企業は製造業だけでなく、卸・小売業やサービス業なども含まれている。世界各国で整備されている多くの統計は製造業を対象としたものであり、製造業以外の産業まで広くカバーして継続的に調査されている統計はほとんど存在しない。製造業だけでなく非製造業をもカバーしている点で、『企業活動基本調査』は世界でも有数の企業統計となっている。

1990年代は、日本経済にとって「失われた10年」と呼ばれている。いわゆる日本的経済システムとして称賛されてきたものに対する評価が一変し、企業レベルでも企業内組織の改革や企業間関係が大きな課題となってきた。企業内組織変革（リストラクチャリング）、長期的企業間関係の見直し、企業活動のグローバル化の進展など企業を取り巻く内外の環境が激変する中、1990年代の政策を分析・評価していくためには、企業レベルのデータにもとづく綿密な分析が不可欠となってくる。『企業活動基本調査』は、今後の日本経済の進むべき方向を検討する上で、極めて重要な統計データセットとなっていると言える。

3. 調査項目・産業分類の時系列的推移

前節でみたように、METI データはパネル・データとして利用できるという大きな利点を持っているが、調査項目の中には1)時系列で同一項目として継続的に調査されているものだけでなく、2)継続的に調査されていないもの、また3)継続的に調査されているが定義に変更があるものがある。このため、例えば三つ目の項目を同じ変数として利用すると、定義の変化による数字の変化を企業活動の変化として誤って捉えてしまう可能性がある。このような問題を避けるためには、調査年によってどのような変数が利用できるのか、またどのような変数が継続的に捉えられるのかを正確に把握しておく必要がある。

表1は1991年から2000年までの調査項目をまとめたものである。観測期間を通じて継続的に追跡できるものとそうでないものがあることがわかる。観測期間を通じて同じ定義で継続的に追跡できる項目としては、例えば有形固定資産額（No.95）、常時従業者数合計（No.62）などがある。一方、無形固定資産額（No.98）は1991年のみでしか利用できない。また、男女別のパート従業者数は2000年では利用できない。

=== 表1 変数項目時系列比較表 ===

さらに注意が必要な変数として、継続的に追跡できるものの、その定義に変更がある変数が挙げられる。具体例は補論で詳しく解説するが、ここではその例を幾つか紹介しよう。

まず経済変数として重要な設備投資額は、METI データでは有形固定資産当期取得額としてまとめられている。この有形固定資産当期取得額（設備投資額、No.119）は、1994年のみ有形固定資産増加額に減価償却費を加えたものとして定義されている。このため、継続的に追跡できる項目とはみなせない。また、本社以外の部門別事業所数、従業者数のうち、

商業事業部門の数値(No.34, No.35)は、1994年以降、飲食店事業部門が独立した項目となったため、1991年の数値と定義が異なる。このような変数の定義には十分に注意する必要がある。

また、調査項目と同様に産業分類についても年によって変更されていることがあるため、機械的に産業番号を接続すると、異なる産業を同一産業とみなしてしまう可能性が出てくる⁹。表2はMETIデータの産業分類を時系列にまとめたものである。一般機械器具卸売業(No.73)、自動車卸売業(No.74)、電気機械器具卸売業(No.75)、その他機械器具卸売業(No.76)は、1991年のみ、「機械器具卸売業」として集約して格付けされていることがわかる。また飲食店は、1997年から一般飲食店(No.90)とその他飲食店(No.91)に細分化されている。さらに、2000年からサービス業の分類が増えており、例えば、冠婚葬祭業(No.114)、ゴルフ場(No.118)、遊園地(No.120)などが新設されている¹⁰。調査項目だけでなく、産業分類を利用するにあたって注意が必要であることがわかるだろう。

=== 表2 産業分類の時系列対応表 ===

4. 企業数の変化

『企業活動基本調査』に限らず、企業・事業所を対象とした統計データにおいては、企業数や事業所数は時系列を通じて変動している。この原因として、参入・退出や合併などが考えられるが、その一方で、その統計の調査方法に依存する部分もあると考えられる。『企業活動基本調査』の場合は従業員50人以上、かつ資本金もしくは出資金3000万円以上の企業を対象としたサンプル調査であるため、同一の企業であっても年によって調査対象範囲に入ったり、対象外になったりする場合がある。その他、調査実施にあたって、非協力的な企業もあり、これらの企業は結果的に調査対象から抜け落ちてしまう¹¹。このため、METIデータでどのくらいの企業がカバーされているのか、そして実際にどの程度サンプルが変動しているのかを確認しておくことには、METIデータの性質を把握する上で必要不可欠な分析と考えられる。そこで、調査対象となる企業数の変化を確認する。

表3は調査対象となる企業数をまとめたものである。表3の「継続企業」は前年度も今年度もデータが利用できる企業である。「新規企業」は前年度のサンプルには存在しないが、今年度サンプルに加わった企業であり、「消滅企業」は前年度のサンプルには存在するが今年度のサンプルから消滅した企業を指す。標本数は、毎年24000程度となっている。各企

⁹ 第1回調査(1991年度対象)から第2回調査(1994年度対象)にかけての産業分類の変更は、平成5年の日本標準産業分類の第10回改訂に伴うものである。その対応関係についての詳細は、H7年調査報告書の「利用上の注意」を参照のこと。

¹⁰ この他、有機化学工業製品製造業(No.26)は、1991年のみ、化学繊維製造業(No.27)を含む点に注意が必要である。

¹¹ 企業規模の変化による規模上がり・規模下がりに加えて、企業の事業転換による転入・転出が発生することにも注意を要する。

業を永久企業番号で時系列に接続し、前年度のサンプルと比較すると、毎年 10%程度の企業がサンプルから消滅している一方で、ほぼ同数の企業が新規に追加されているアン・バランス・パネル・データになっていることがわかる。

=== 表 3 METI データベースにおける標本数 ===

なお、計量経済分析では、しばしば、パネル・データの抜け落ち、新規追加を含むデータセットに対して特別な処置を求める場合がある。そこで、METI データにおけるサンプルの抜け落ち、新規追加のパターンを整理しておこう。

表 4 は、1991 年から 2000 年までの個票に基づく、回答・非回答パターンをまとめたものである。この期間中の全サンプル企業数は 41,451 となっている。このうち、バランス・パネルを構成する企業数、つまり観測期間を通じて継続的に追跡できる企業数は 12,680 社となっている。バランス・パネルを構成しない企業のうち、途中の中断無く 2 回以上連続してデータの得られる企業数は 14,227 であり、期間中 1 回だけ回答している企業数は 7,976 である。そして、これらに該当しない企業数（一度、回答を中断し、その後再開した企業）は 6,568 となっている。

=== 表 4 回答・非回答のパターン別企業数 ===

また、これらの企業は産業間をまたいで活動していることも確認できる。表 5 は連続する二期間についての企業の所属する産業を製造業と非製造業の二分類に分けてまとめたものである。表頭は $t-1$ 年企業が属していた産業、表側には t 年に企業が属する産業をあらわしており、表中の数値は、その企業数をあらわしている。同一企業が同一産業に属するのならば、非対角要素はゼロに近くなる必要がある。この表より、製造業と非製造業という大雑把なくくりでも、毎年 300 社程度が産業を移動していることがわかる。

=== 表 5 産業移動マトリックス ===

このように METI データでは多くの企業が産業を移動している。このような状況を踏まえると、分析にあたっては、『企業活動基本調査』の産業格付けルール¹²を理解しておくことが重要であろう。企業活動基本調査では、企業の産業は売上高の最大構成品目に基づいて格付けされている。具体的には、企業活動基本調査では、企業の売上高を、(1)鉱産品の販売、(2)製造品の販売、(3)製造品の加工賃収入、(4)卸売・小売、飲食店の売上、(5)サービス事業収入、(6)その他の事業収入に分けて、それぞれ詳細に調べている。これらを大分類ご

¹² 詳細は、『企業活動基本調査』各年度版「利用上の注意」を参照のこと。

とに合算し、最も売上高の大きいもので大分類を決定し、その大分類の中において、売上高の最も高い品目で産業（小分類）を決定している。このため事業の多角化を行っている企業のある部門の売上が変化し、最大構成比を占める部門が変わった場合、当該企業の産業分類は変更になるのである。

5. 主要経済変数の動き

本節では経済分析で利用される主要変数に注目し、実際の統計データが時系列的にどのように変化しているかを確認しておこう。具体的には企業の生産性を考える上で重要なアウトプット、インプット、およびコストの情報、さらに経済分析でよく利用される設立年を取り上げる。本論文で取り上げるアウトプット、インプット、コストの変数は表 6 にまとめた。表 6 の変数のうち、資本ストック、付加価値額、生産委託額、中間投入額は筆者らによって再計算されたものである¹³。

=== 表 6 時系列特性を検討する調査項目一覧 ===

5.1. アウトプットとインプット、コスト

本節では、アウトプット、インプットとコストについて、次のようなクロス・セクションと時系列の二つの視点から検討する。クロス・セクションの視点では、主要変数の基本統計量を分析する。一方、時系列の視点では各主要変数の変化率の分布に注目し、その特性を明らかにする。

以下の表 7、表 8 では、企業活動基本調査における主要変数の平均、変動係数、標準偏差、確率四分位の 25% 値と 75% 値を、それぞれ水準と変化率について計算したものである。表 9 は、本来非ゼロの値をとるべき変数で、ゼロ値を回答している企業数をまとめたものである。以下、主要変数の動きを解説する。

¹³ 資本ストックは、西村・中島・清田 (2003) にならい、以下のように推計した。

$$K_{it} = \begin{cases} K_{it-1} + (A_{it} - A_{it-1}) / PI_t & \text{if } (A_{it} - A_{it-1}) > 0 \\ K_{it-1} + (A_{it} - A_{it-1}) & \text{if } (A_{it} - A_{it-1}) \leq 0 \end{cases}$$

ここで、 K_{it} は t 期における i 企業の資本ストック、 A_{it} は、有形固定資産残高である。 PI_t は、投資財デフレーターであり、国民経済計算（内閣府）「形態別の総資本形成」の有形固定資産のデフレーターを用いた。 $(A_{it} - A_{it-1}) > 0$ は、純投資と見なし、実質化し前期の資本ストックに加算するが、 $(A_{it} - A_{it-1}) \leq 0$ は、減価償却のみで純投資はゼロであると考え、実質化せずに、前期の資本ストックから差し引いている。企業活動基本調査の調査項目に設備投資額や減価償却についての項目があるが、前述のように、いずれも時系列で比較した場合、動きが安定的ではなく、一般的に利用されている恒久的棚卸法を利用することが難しい。このため、今回はこのような簡便法を利用した。また、付加価値、生産委託額、中間投入額については、表 6 を参照のこと。

==== 表 7 記述統計 ====
==== 表 8 変化率の記述統計量 ====
==== 表 9 欠損値の有無 ====

[アウトプット]

売上高

METI データでは、企業のアウトプットを捉える方法として売上高がある。売上高の平均は 23026¹⁴であり、時系列的にも安定している。標準偏差は全体では 224849 であるがこちらは時系列的には小さくなる傾向にある。また、変化率に注目すると、全体の平均値は 0.4% だが、景気変動を反映して毎年の平均値は大きく変動している。しかし、標準偏差は 0.185 で安定していることがわかる。

付加価値

売上高以外の指標でアウトプットとしてしばしば用いられるのは付加価値である。『企業活動基本調査報告書』の定義に基づくと、付加価値は、

$$\text{営業利益} + \text{給与総額} + \text{租税公課} + \text{減価償却費} + \text{賃借料}$$

と定義される。ただし、平成 3 年度と平成 6 年度に限り、粗付加価値、すなわち、

$$\text{売上高} + \text{棚卸資産増減} - \text{営業費用} + \text{給与総額} + \text{減価償却費}$$

で定義されている¹⁵。今回のように、パネル・データによる分析を行う際は、変数の定義に一貫性を持たせる必要がある。そこで、ここでは、以下のように付加価値を定義してその特性について検討を行った。

$$\text{売上高} - \text{営業費用} + \text{給与総額} + \text{減価償却費} + \text{賃借料}$$

また、後述するように、減価償却費の変化に多少不自然な箇所があるので、減価償却費を利用しない形で付加価値を定義するために、西村・中島・清田(2003)に倣い、これに加えて、以下の定義でも検討を行った。

$$\text{売上高} - \text{仕入高}$$

付加価値 と には、それらの基本統計量に若干の違いがある。例えば、全体をプールした平均は付加価値 のほうが大きい。この理由は、 の場合、間接経費（人件費や賃借料、減価償却費等、付加価値に計上されるべき費用項目以外の項目）が含まれるからである。しかし変動係数を見ると、両者の差は小さくなることが確認できる。

¹⁴ 各変数の記述統計量の単位はすべて 100 万円である。

¹⁵ 詳しくは、「企業活動基本調査報告書」各年度版の「利用上の注意」を参照のこと。

一方、成長率をみると、全体をプールしたサンプルによる平均値は付加価値 では 0.8% であるが、付加価値 は-0.8%と符号が逆になっている。変動係数も両者の間に違いがある。しかし、確率四分位の 25%値、75%値のレベルではほとんど同じ動きをしており、定義そのものの違いと言うよりは、むしろ最大値、最小値などの違いがこれらの変動を導いていると考えられる。

[インプット]

一方、インプットとしては資本（有形固定資産）、中間投入（仕入高）がある。その特性は以下のようにまとめられる。

有形固定資産

全体の平均値は 4920 であり、標準偏差は 45639、変動係数は 9.28。時系列でみると、2000 年に調査対象企業が増えたこともあり、2000 年のみ平均、標準偏差ともに大きくなっている。一方、変化率は平均値で見ると、1998 年以降小さくなっているが、標準偏差は比較的安定している。

仕入高

全体の平均値は、15718 であり、標準偏差は 205006、変動係数は 13.0。平均値は比較的安定しているが、標準偏差は年々小さくなっている。一方、変化率は、平均が-0.4%、標準偏差が 0.578 となっている。標準偏差は比較的安定しているものの、平均は年によりかなりぶれがみられる。

[コスト]

減価償却費

水準でみると、データ全体をプールしたときの平均値は 517、標準偏差は 5396、変動係数が 10.4 であり、年毎に見ても、さほど大きな変動はみられない。しかし、変化率で見ると、全体の平均と標準偏差がそれぞれ 0.27 と 0.65 なのに対して、1995 年の平均値と標準偏差は 0.31 と 1.05 と大きな値になっている。確率四分位 75%値で見ても、他の年度の倍以上となっているので、小数の外れ値の影響ではない。時系列で比較するには十分注意が必要である。

賃金率

賃金率は、調査項目からは得られない項目であるが、給与総額と従業員数から平均賃金を定義することで、しばしば経済分析で利用されている。ここでは、給与総額 / 従業者数で定義した賃金 と分母を連続する 2 時点の従業員数の平均を用いた賃金 の二種類の指標を作成し、その特性を検討する。賃金 は、平均賃金の分子が前年度の 1 年間のフロー

の値であるのに対して、分母が直近のストック値であるという、調査対象期間の不一致を補正することを意図した定義である。

賃金率水準でみたときの、全体の平均値は賃金 が 4.73、賃金 は 4.82 であり、時系列的にも安定している。標準偏差は、前者が 2.25、後者が 1.74 と前者がかなり大きく、変動係数でも賃金 が 0.475、賃金 が 0.362 と、賃金 の散らばりが大きいことがわかる。変化率で見ても、賃金 の平均値 5.13 は賃金 の平均値 3.854 を大きく上回っているほか、標準偏差も若干賃金 のほうが大きい。

図 1 は、分布の特性を調べるため、2000 年の賃金率で賃金 (wage 1)と賃金 (wage 2)の Box Plot (箱ひげ図)を作成したものである。箱ひげ図の「箱」(中央の四角形)は確率四分位の 25%値から 75%値の領域を示す。この箱から上下に伸びている「ひげ」(四角形の上下に位置する線)は、箱の高さの 1.5 倍以内にあるデータの上限と下限を示している。この範囲に入らないデータは外れ値として「ひげ」の上下に丸印で示されている。この図から賃金 の分布のほうが、賃金 の分布に比べて、右裾が厚くなっていることがわかる。これは、賃金 では年度末の従業員数を用いているため、年度内に従業員数が大きく変化した企業では賃金率が過大、もしくは過小に評価されていることによって分布の裾が厚くなっていると考えられる。この対処法としては、例えば 2 時点連続でデータが利用可能な場合は賃金率 を用い、可能でない場合は賃金率 を用いる方法が考えられるだろう。

設備投資

水準、および変化率で見たときの、標準偏差はそれぞれ、7189 と 0.365 であり、この数値は年毎にサンプルを分けて計算した場合もさほど変わらないので比較的安定しているかのように見える。しかし、欠損値の有無を調べた表 9 をみると、毎年 5000 社近くの企業が設備投資額を 0 と回答しており、この点には注意が必要である。なお、設備投資を 0 と回答している企業の多くは非製造業であり、たとえば、2000 年のデータでみると、4791 社のうち、3237 社が非製造業であった。小売業やサービス業では、機械や建物は自社保有せず、リースや賃貸による場合が多いことを反映しているのかもしれない。

5.2. 企業年齢

企業年齢はアウトプットやインプットのような生産関数に関わる情報とは異なるが、企業特性を示す変数として多くの研究で注目されてきた。METI データでは、企業年齢そのものの情報が得られないため、これまでの分析の多くは、企業年齢を(調査対象年 - 設立年)と定義することで対応している。しかし、METI データを使って単純に企業年齢を計算しても、同一企業の企業年齢を隣接する年次で比較したときに、非連続な動きをとるケースが出てくる場合がある。以下ではこの問題について少し詳しく解説しよう¹⁶。

¹⁶ この問題に加えて、単純に(調査対象年 - 設立年)で企業年齢を定義すると、企業年齢がマイナスになってしまう問題がある。この問題は、調査対象年と調査時点のズレから生

=== 表 10 設立年次別企業数 ===

表 10 は、調査年別に設立年別企業数を調べたものである。設立年次別の企業数はデータの新規追加や消滅の影響を受けるので時系列的には多少の変更が見られると考えられる。しかし、表 10 の各設立年次の企業数を時系列で比較すると極端な変動がみられる箇所がある。たとえば、1996 年と 1997 年の 1945 年設立企業数を比較すると、1996 年のそれは、2453 社であるのに対して、1997 年では、200 社に激減している。

表 11 は、継続してデータが得られる企業のうち、どの程度の企業が異なる調査年次を回答しているかをチェックした結果である。毎年、多くの企業が設立年次の情報を更新しているが、特に 96 年から 97 年の設立年次変更企業数が目立つことがわかる。このような設立年次の変更があると、本来、毎年 1 ずつ増加すべき企業年齢が非連続に変化することになる。

=== 表 11 設立年次の変更があった企業数 ===

では、なぜこのような問題が発生するのであろうか。その理由としては以下の二つが考えられる。一つは、調査票の設立年次に関する質問形式が変更になったことであり、もう一つは、同一企業であっても担当者によって異なる設立年が報告されていることがあるためである。まずは、前者の理由によるものについて具体的にみていこう。第 1 回調査では、昭和 20 年以前に設立された企業に対して詳しい設立年次の回答を求められていなかった。ところが、第 2 回調査では、平成 4 年 4 月 1 日以降に設立した企業のみで設立年次を回答するように求めている¹⁷。さらに第 3 回調査以降は、毎年、すべての企業に設立年次を回答

じるエラーに起因している。『企業活動基本調査』では、財務状況や取引活動については、調査時点の前年度の企業の状態を質問しているが、設立年等の企業の基本情報については調査時点である 6 月 1 日現在の情報を記載するように求めている。このため、1 月以降に設立された企業は調査時点の年が設立年として報告される。この場合、調査対象年は調査時点の前年であるので、企業年齢はマイナス 1 になってしまう。表 10 の網掛けになっている部分が調査対象年度と調査年時のズレから生じている問題を示したものである。たとえば、調査対象年次 1995 年の列（調査実施年 1996 年）をみると、1996 年設立の企業が 12 社あることがわかる。この 12 社について、調査年 - 設立年で企業年齢を計算すると、-1 歳になってしまう。そこで、網掛け部分については、企業年齢を 0 とするなどの対処が必要である。

¹⁷ また、第 1 回の調査の設立年の情報を利用する場合、設立年次が和暦として記録されている点は特に注意が必要だろう。具体的には、和暦設立年次が 21 以上、63 以下であれば、昭和期設立企業であるので、1925 を足し、21 以下であれば平成以降に設立された企業なので 1988 を足すという作業を行う。ただし、例外として、設立年次が 0 になっている企業と、設立年次が 9 になっている企業が存在する。これらの企業は、設立年次 0 は 1945 年以前の設立の企業、設立年次 9 の企業は昭和 9 年（1934 年）設立の企業と考えることができる。この点は、新保（1999）の指摘による。

するように質問項目が改められている。そのため、第1回調査で昭和20年以前に設立されたと回答している企業を追跡すると、第2回目以降の調査で設立年次に修正が行われることがあるのである。表11の「変更あり」の企業のうち「昭和20年以前に設立」した企業とは、設立年次に変更があった企業のうち、第1回調査で「昭和20年(1945年)以前に設立した」と回答していた企業の数である。この表から、設立年次が変更になっている企業の半数近くが調査票の質問形式の変更によって設立年次を変更したことがわかる。一方、残りの企業は後者の理由によるものだと考えられるが、その変更パターンはシステムティックなものではないので、その背景を詳しく知ることは難しいのが現状である。

これらの問題に対する対処方法としては、他の企業ベースのデータとの比較が考えられる。しかし、これらの作業は容易ではない。そこで実際の分析にあたっては、とりあえず設立年を固定させるために、設立年を最新の調査に合わせる方法と期間中の最も古い年に合わせるという方法が考えられる。しかし、どちらの方法が望ましいかは一概には言えない¹⁸。

6. 終わりに

本論文では、METIデータの有用性を整理するとともに、パネル・データを作成する上での注意点、問題点を議論した。本論文の貢献のひとつは、METIデータからより精緻なパネル・データを作り上げるひとつの方法を提示していることにある。パネル・データは、企業データベースを永久背番号で単に接続しただけで完成するものではない。個々の調査項目の変更を正しく捉えなければ、調査の変更を企業活動の変化と間違えて捕らえてしまう可能性があるためである。本論文では、調査項目の変化、産業分類の変化を綿密に捉えることで、METIデータの情報を最大限に生かすようなパネル・データの作成法を提示した。

また、本論文では、主要な経済変数をもとに、その動きについても簡単な分析を行った。分析の結果、一部に非常に変化の大きい企業が含まれているものの、平均的にみると多くの変数は、時系列的に（不自然な変化をすることなく）比較的安定して推移していることがわかった。この事実は、METIデータの信頼性の高さを確認する結果と言えるだろう。

なお、一部の非常に変化の大きい企業、変数については、特異値が統計量に深刻な影響をもたらす可能性があるため、経済分析の際は十分に注意を払う必要がある。しかし、統計調査の段階で精度を上げていくことができれば、より信頼性の高いデータベースを整備することができる。これらのデータは政策立案にも直結する重要な情報を提供するため、より綿密な調査が期待される。この点については、研究者が統計分析によって得られた知

¹⁸ ただし、設立年を観測期間の最も新しい年に合わせるという対処法を採用する場合は、企業年齢が負になるという問題が生じる可能性がある。たとえば、1991年から2000年まで継続してデータの得られる企業を考えよう。この企業が1996年調査までは設立年を1991年と回答し、1997年以降1995年に更新したとする。この場合、最新の調査年に合わせて設立年を定義すると、1994年の企業年齢は-1歳、1991年の企業年齢は-4歳となってしまう。

見を統計調査の実施部局にフィードバックすることが重要な意味を持つと考えられる。

本論文を締めくくるにあたって、今後、METI データの可能性をさらに高めていくために、三つの課題を示しておこう。第一は、統計調査そのものの継続性、拡充である。METI データは企業組織や貿易、雇用といった情報が豊富だが、その一方で有価証券報告書に比べて財務項目が主要項目に限定されているという弱点もある。もしこの点が改善されれば、コーポレート・ガバナンスや企業金融の分析が可能になり、特に非上場企業も含めて調査されている点で、非常に強力な情報源となりうる¹⁹。

また、第三節で見たように、調査項目の中には継続的に追跡されていない情報も存在する。調査項目が継続的に追跡されていれば、企業が年を経てどのように変化していくのかを綿密に捉えることができるだけでなく、計量経済学的に精緻な分析を行えるという利点がある。調査の継続性や拡充は、統計調査のコストや労力との関係もあり簡単ではない。しかし、地道な調査が政策を形作る上で重要な情報を提供していることは、十分に認識する必要があるだろう。

第二は、METI データと他の集計統計との接続である。METI データはそれ自体有用なものだが、他統計と接続することでその利用価値は一層高まる。例えば他省庁で公表されている産業レベルのデータと接続すれば、産業特性をより細かくコントロールすることが可能になる。また、他の集計された統計と抽出サンプルが一致するように METI データを再集計し、接続して分析する方法も考えられる。たとえば、産業特性の変化と賃金構造の分析では、しばしば賃金センサス（厚生労働省）のデータに各種の産業別集計データを接続して利用されるが、『企業活動基本調査』の公表されている集計表では必ずしも企業規模区分や業種分類が一致しているとは限らない。こういった場合、METI データを用いてデータを再集計し分析する方法が考えられる。

また、他統計との接続として最も重要な課題のひとつに、デフレーターを整備が挙げられる。企業の生産活動を時系列で分析していく上ではアウトプットとインプットの実質化が不可欠だが、実質化のためにはデフレーターとの接続が必要になってくるからである。とりわけ、日本経済が 90 年代にデフレに直面していたことを踏まえると、実質的な経済活動を捉えることはある程度の意義があると考えられる。デフレーターを整備し、企業活動を実質的に把握していくようになれば、価格変化の影響を取り除いた企業活動の変化を見極めることができるのである。

第三は、他の個票データとのマッチングである。前述のように『企業活動基本調査』の調査項目は多岐にわたっており様々なテーマでの分析が可能だが、他の企業、事業所レベルの個票データとマッチングすることで、企業の生産、立地、国際化などの分析をより精

¹⁹ なお、統計審議会の諮問第 292 号の答申「企業活動基本調査の改正について」（2003 年 12 月 12 日）によると、『企業活動基本調査』の財務内容については、平成 17 年度調査から法人企業統計調査の調査結果を活用することを課題として提言している。これが実現すれば、経済分析の可能性がさらに広がると期待される。

緻に捉えることができるようになる。これまでにこのような試みを行った例として、『海外事業活動基本調査』個票データと『企業活動基本調査』個票をマッチングした河井（1997）がある。河井（1997）では、企業活動基本調査ではあまり詳細な情報が得られない各企業の海外子会社の生産状況の情報を、海外事業活動基本調査によって補い、その国内事業活動に与える分析を行っている。ただし、このような分析には、複数の名簿情報を用いて企業名称等を手がかりに数万社のコンバーターを作成するという作業を要するため、統計調査の実施部局との連携や協力が不可欠になってくる。

以上で見てきたように、METI データはうまく活用することで、その利用可能性は無限に広がっているとも言える。もちろん、データの活用にあたっては、その特性を十分に検討し、議論を重ねることが必要である。さらなるデータ特性の理解の深化によって、質の高い学術的、および政策的研究が期待される。

参考文献

- 河井啓希（1997）「日本企業の人事・雇用面におけるリストラと産業構造変換」（財）機械振興協会・経済研究所『企業統計パネル・データを活用した国際化・企業間関係及び雇用創出に関する調査研究』
- 木村福成・清田耕造（2002）「企業活動のグローバル化と企業パフォーマンス：『企業活動基本調査』にもとづく分析」『経済統計研究』、30(2): 1-12。
- 経済産業省経済産業政策局調査統計部『企業活動基本調査報告書』各年度版
- 新保一成（1999）「企業活動基本調査パネル・データ」慶應義塾大学未刊行論文。
- 西村清彦・中島隆信・清田耕造（2003）「失われた 1990 年代、日本産業に何が起こったのか？ - 企業の参入退出と全要素生産性 - 」経済産業研究所ディスカッションペーパー RIETI Discussion Paper Series 03-J-002。