

階層クラスタリング手法を用いた
コモディティ・フロー法における配分比率の安定性に対する
産業連関表の行部門統合に係る影響の検証

赤木 茅

May 2018



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

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.

目次

1.	はじめに.....	2
1.1.	背景.....	2
1.2.	コモディティ・フロー法及び部門設計の考え方.....	4
2.	概念整理.....	4
2.1.	定義：配分比率の安定性.....	4
2.2.	複数の需要項目に配分される要因.....	5
2.3.	非単用財が推計精度に与える影響.....	5
3.	分析手法.....	6
3.1.	手法概要.....	6
3.2.	使用データ.....	7
3.3.	コモ法分類の産業連関表分類への統合.....	7
3.4.	階層クラスタリング.....	8
3.4.1.	統合分類候補の選出.....	9
3.4.2.	統合方法.....	11
3.4.3.	部門統合による誤差額の計測.....	11
3.4.4.	乖離指標.....	13
4.	分析結果.....	14
5.	考察とまとめ.....	17
5.1.	乖離総額.....	17
5.2.	行部門の配分総額に占める乖離額の割合.....	17
5.3.	需要項目別にみた乖離額の割合.....	17
5.4.	産業別にみた統合部門の特徴.....	18
5.5.	まとめ.....	18
5.6.	留保された課題及び今後の展開.....	18
6.	参考文献.....	20
7.	付録.....	21

階層クラスタリング手法を用いた
コモディティ・フロー法における配分比率の安定性に対する
産業連関表の行部門統合に係る影響の検証¹

赤木 茅

-----要旨-----

現在政府では、統計改革の一環として、GDP統計の基盤となる産業連関表の供給・使用表（SUT）体系への移行に向けて、SUTの基本構成（部門数、部門構成等）の大枠について検討が行われている。我が国の国民経済計算におけるコモディティ・フロー法の中間年次推計では、原則としてその配分比率に産業連関表の値（基準年値）を用いているため、複数の需要先に配分される部門の需要先別配分比率が中間年において実際には変動している場合、中間年推計値に誤差が生じる。GDP統計の精度向上のためには、SUTの基本構成の検討においてコモディティ・フロー法の産出の配分ができる限り安定するように設計し、基準年値と中間年推計値の間の配分比率の変動を小さくすることが重要である。

本稿では、SUTの基本構成の大枠の検討に資するよう、産業連関表の行部門に着目し、現行の産業連関表データにおいて、生じる誤差の最も少ない行部門から順に統合する階層クラスタリングの手法を利用して、産業連関表の行部門数の設定の違いがコモディティ・フロー法における各需要項目への品目別配分比率の安定性にどのような影響を及ぼすかについて検証を行なった。その結果、行部門数の減少に伴って、誤差額が拡大し、一定以上の部門数を下回ると、推計精度が大きく低下することや、各部門の誤差の発生に関する特徴が確認された。なお、実際の部門設定においては、個別の分類の特性等を十分考慮する必要がある。本稿の分析は、機械的ルールに基づいて全体の傾向を把握するものであり、直接的に部門の設定を提言するものではない。

1. はじめに

1.1. 背景

現在、政府では「統計改革推進会議最終取りまとめ」（平成 29 年 5 月 19 日 統計改革推進会議決定）、「経済財政運営と改革の基本方針 2017」（平成 29 年 6 月 9 日閣議決定）及び「公的統計の整備に関する基本的な計画」（平成 30 年 3 月 6 日閣議決定）に基づき、「生産面を中心に見直した GDP 統計への整備」として、GDP 統計の基盤となる産業連関表（以下、概念としての産業連関表を「IO」、日本の

ものを「JIO」と呼称する。）の供給・使用表（SUT : Supply and Use Tables）体系への移行に向けて、政府一体となって取組を進めている。

「統計改革推進会議最終取りまとめ」に記されている通り、2025 年を対象年次とする JIO（2029 年度公表予定）から SUT 体系に移行するとともに、2030 年度に予定している国民経済計算の基準改定において、全産業の直接推計による中間年 SUT の構築が予定されており、統計委員会の下に設置されている国民経済計算体系的整備部会 SUT タスクフォース会合において、SUT の基本構成（部門数、部門構成）等に係る具体的な検討が行わ

¹ 本稿は国民経済計算体系的整備部会 SUT タスクフォース会合において資料として提出された研究発表を元に作成されている。同会合において、宮川努座長（学習院大学経済学部教授）はじめ出席者の皆様から大変有益なコメントを頂いた。また、西崎文平所長、長谷川秀司総括政策研究官、木滝秀彰国民経済計算部国民生産課課長、梶村麻衣子元研究官、山岸圭輔総務部総務課課長補佐をはじめとした内閣府経済社会総合研究所の職員より、主に実務の観点からお知恵を拝借頂いた。ここに記して御礼申し上げる。なお、当然ながら、残された誤りは全て筆者の責任によるものである。

れている。

国際連合において採択された国民経済計算の国際的なフレームワークである「System of National Account 1993」²において提唱された体系である SUT では、品目³がどの産業から生産あるいは輸入され（供給表）、それらがどのようにして中間投入及び、最終需要に分配されるか（使用表）を記述する統計表からなる。

国民経済計算体系（SNA：System of National Account）は、生産側、支出側、分配側それぞれの GDP 推計値が一致する、いわゆる三面等価を理論的には保障している。しかし、実際の推計に当たっては、基礎統計の精度・時点や、推計手法の違いなどによって、それぞれの推計値の間に乖離が生じることがあり、このような乖離は一般に「統計上の不突合」として記録するか、その乖離を何らかの手法を用いて調整している（以下「バランシング」という）。

我が国の国民経済計算（以下、「JSNA」と呼称する。）では、支出側推計にコモディティ・フロー法（以下、「コモ法」と呼称する。）を、生産側推計に付加価値法を用いているが、この両者の推計値にも不突合が発生している。SUT は、生産物の品目ごとの供給と使用をバランシングしつつ、基礎統計データから最小限の加工で SNA の生産勘定を推計するフレームワークを提供する機能を有しており、実際に平成 23 年基準改定以後、この SUT の枠組みを用いたバランシング手法の導入によって、JSNA における生産側 GDP と支出側 GDP の間の「統計上の不突合」の縮減がなされた。

現在、中間年の JSNA 推計では、IO の取引基本表（X 表、商品×商品）及び、産業別生産

物（商品）産出表（V 表）を基礎として、JSNA への概念・分類の組換えや数学的仮定（商品技術仮定）の適用、及び、毎年の出荷額・産業別の費用構造等の情報を用いた延長年推計等により、V 表、U 表という形で中間年 SUT と実質的に同じ内容のものが推計されている。

しかし、現行の JSNA では基準年次において、経済センサスなどの基礎統計を用いてまず JIO の取引基本表を作り、その後、数学的仮定を置いたうえで取引基本表を U 表に変換する作業を挟むなどの手法を用いており、直接的に SUT 体系を用いていない。このことも一因として、現行の JSNA の推計においては、推計手順の複雑化、それに伴う精度の低下（誤差の拡大）や、JIO と JSNA の概念上の相違の存在などが問題視されている。

清水、他（2008）は、GDP 統計の精度低下要因として、基幹統計の一つである産業連関表と一次統計調査、更に JSNA との統計概念の整合性、及び各種統計の推計手法の間の整合性の問題を指摘している。今般の統計改革では、JIO の SUT 体系への移行に伴い、従来型の取引基本表経由の間接的な推計ではなく、基礎統計から直接 SUT を作成し、そこから GDP 統計を推計することが可能となり、推計過程における基礎統計から JSNA までの概念の一貫性と、推計精度の向上が見込まれている。

我が国の SUT 体系への移行に当たっては、SUT の基本構成（具体的には、生産物・産業の概念、表章部門の考え方、部門の改廃ルール、部門数）の検討を始めとして、様々な課題が存在するが、本稿では、SUT の基本構成のうち、部門の設計に焦点を絞り、特に行部門の部門数の設定がコモ法推計の精度にどのよ

² (Inter-Secretariat Working Group on National Accounts, 1993)

³ 日本の産業連関表の取引基本表は商品×商品のマトリックス、SUT は生産物（商品）×産業のマトリックスとして構成されているが、本稿はコモディティ・フロー法行部門の構成単位を品目としているため、対象を「品目」で統一する。本稿で行われる分析の対象はコモ法における品目を JIO の商品と対応付けたものであり、各品目の呼称は全て「平成 23 年（2011 年）産業連関表（確報）部門分類コード表」における基本分類に則っている。なお、「行部門」とは、産業連関表を横の係数の並びを示す、産業連関表レベルでの概念である。

うな影響を与えるのかについて定量的に検証する。

1.2. コモディティ・フロー法及び部門設計の考え方

コモ法は、定義された品目の分類ごとに、それらの国内総供給の、中間投入、最終需要（最終消費及び総資本形成）への配分を明らかにする形で支出側の国内総生産を推計する一連の手法、言い換えれば、IOの行方向の算出である。

コモ法は、品目の配分先⁴を算出するため、ある品目がただ一つの配分先に需要されるもの⁵であれば、その推計は容易かつ、少なくともいずれの部門に配分されるかに関しては正確なものとなる。したがって、部門の設定では各品目のデータの入手可能性、用途の類似性、代替可能性などを鑑みつつ、商品ごとの需要先が可能な限り一意に定まるようにコモ法による財貨・サービスの品目が細分化されることが望ましいと考えられる。

熊谷（2012）では、JSNAの課題の1つである、生産側推計と支出側推計の統計上の不突合の発生理由として、コモ法における中間年の配分比率を基準年のそれに固定していること、即ち配分比率の中間年の変動を捕捉できていないことを挙げている。各生産物の生産額を正確に把握できれば、中間投入額や最終需要額の正確な把握につながることから、SUT体系への移行において可能な限り配分比率の変動の少ない部門設定をすることは、JSNAの推計精度に関して大きな意味を持つといえる。

ただし、部門分類の詳細化は、GDP統計

の精度向上という観点からは望ましいものであるが、徒に部門を細かくすればよいというものでもない。分類の詳細化は、調査項目や調査対象の増加につながり、SUT作成のベースとなる基礎統計における実査及び報告者の負担増を招く可能性がある。

菅（2009）において言及されているように、JIO作成のための特別調査の回収率は近年急速に低下しており、また、報告者負担の軽減は、同調査の回収率の向上に寄与するとされる⁶。この点に関しては、「公的統計の整備に関する基本的な報告」（平成30年3月6日閣議決定）においても、『統計ニーズに対応するために、報告者に過度な負担を強いることは、統計調査への協力意識の低下、ひいては統計調査の結果精度にも影響を及ぼす』と示唆されており、政府も報告者負担の軽減に努めている。

本稿は、部門数の設計の推計精度への影響を分析の対象とするが、実際に部門数を設定するには、精度のほかに実査及び報告者の負担の両面を考慮して設計されなければならない。

2. 概念整理

2.1. 定義：配分比率の安定性

通常、IOに関わる「安定性」という概念が対象とするところは、IOのレオンチェフ投入係数行列の安定性、すなわち「シンメトリック産業連関表」における中間投入部分の列部門構造の安定性であり、コモ法における行部門構造に係る「配分比率の安定性」という概

⁴ コモ法において、財貨・サービスの商品は需要先別（中間投入向け、家計消費向け、固定資本向けのいずれか）に配分される。今回の検証では、加えて、中間投入のうち、建設向け中間投入を別途配分先として分離している。

⁵ 例えば「原油」は家計が消費することは考えにくく、100%中間投入に配分される。

⁶ 菅（2009）で分析の対象となっているのはIOの投入額の内訳を配分する情報を提供する、「平成17年産業連関表作成のための特別調査」のうち「投入調査」である。本稿で扱われる配分比率に関して関連が深いのは、産業連関表産出額の内訳を配分する情報を提供する「製品需給調査」であり、「平成17年産業連関表作成のための特別調査」及び「平成23年（2011年）産業連関構造調査」の双方が関連する。したがって、各調査が直接対応するものではないが、産業連関表作成に係る調査一般における報告者負担軽減の重要性を示す意味で参考とした。

念は一般的ではない。これは、IO の利用という観点において代表的な波及効果分析の精度に、投入係数の安定性が大きく関わってくるのが理由であるが、JSNA の推計においては、配分額推計の精度にも重要な意味がある。

そこで、本章では、コモ法における「配分比率の安定性」概念を定義し、その分析の意味を確認する。

2.2. 複数の需要項目に配分される要因

先に述べたように、JSNA の推計精度の向上という観点からすると、コモ法において設定される品目は、可能な限り単一の需要先に配分されることが望ましい。コモ法は、品目ごとに国内総供給額を把握し、それを流通段階ごとに各配分先へ配分し、その配分額を算出する手法であり、国内総供給額は、最終的には中間投入、最終需要、固定資本形成のいずれかに配分される（ただし、中間投入から建設向け中間投入を分離可能であり、本稿では分離して扱っている）。したがって、ある品目が100%単一の配分先へ配分される場合、最終段階における誤差が生まれる余地がなくなるため推計誤差が小さくなると考えられる。

コモ法において、品目が複数の需要先に配分される要因には、主に以下2つのパターンが考えられる。

第一に、単純に複数の需要先に配分される品目（以下、「混用財」と呼称する。）であり、これは例えば、「かんしょ、食パン、醤油」などの農産物や食品が考えられる。具体的には、かんしょを一般家庭が食事での利用として家計消費に需要し、食品加工業者や外食業者が中間投入として需要するといったケースが想定される（家計消費と中間投入が混在）。その他にも、上記のとおり、中間投入、家計消費

双方に配分される品目は多数存在する（衣服や家具などの既製品、燃料など）。また同様に、ボイラやポンプなどの機械やソフトウェアなど、中間投入と資本形成双方にまたがる品目⁷や、PCや自動車など、家計消費と固定資本形成双方に配分される品目も存在する。

これらの混用財は、例えば「事業者向けかんしょ」、「家計向けかんしょ」など概念的に単一の需要先に配分される品目（以下、「単用財」と呼称する。）として区分することも可能だが、実際のところ各農家、卸売、小売業者などがそれらを個別に分離し記録しておらず、調査によって個別に把握することが困難である場合が多い。

品目が複数の需要項目に配分される第2のパターンとして考えられるのは、プロダクトミックスである。プロダクトミックスとは、呼称としての部門名に、複数の実態の異なる品目がまとめられていることをいう。一般に、「その他の○○」「ほかに分類されない○○」などのバスケット項目が代表的だが、そのほかにも「衛生陶器と、陶磁器食器」など用途の異なる品目が「陶磁器」という区分でまとめられるなど、分類が粗い場合に起こる。

混用財と、プロダクトミックスの違いは、分類の定義や詳細度に依存するため、明確に区分することは難しいが、用途の代替可能性が判断における1つの指標となる。例えば、先の例の「家計向けかんしょ」を「事業者向けかんしょ」で代替することは可能だが、「陶磁器食器」を「衛生陶器」で代替することは考えにくい。

2.3. 非単用財が推計精度に与える影響

以降、「配分比率」を、ある品目の国内総供

⁷ IOにおいて単独で資本形成になるボイラが、船に組み込まれるようなケースを迂回といい、機械組み込み、建設迂回、土木迂回（据付迂回）、造船迂回、自衛隊が購入した武器等、などの限定的なケースに限って資本財が中間投入に計上される。

給に占める「中間投入⁸、建設向け中間投入、家計消費、固定資本形成」それぞれへの配分額の割合と定義する。また、ある品目の「配分比率が安定的である」ということを、その品目の各配分先への配分比率が経年で変化しないこととする。

特定の品目にプロダクトミックスがある、あるいは混用財である場合、その配分比率はそれぞれの年の生産比、需要構造にしたがって変化するため、経年の安定性は低下する。一方、品目の需要項目が1つに限定されている場合、原理的にその品目の配分比率は完全に安定することとなる⁹。

この変化、特に混用財の配分比率の変化は自然な現象であり、それ自体は問題となるものではないが、JSNAにおけるコモ法では、中間投入（建設向け中間投入を含む、分離して推計は行われていない点に留意）、家計消費、固定資本形成の4つの需要項目に対する配分比率は、把握する情報が存在しないことから、原則として基準年固定であり、中間年次においてはSUTバランシングによる変動等はあるものの基本的に同一の値を用いている。したがって、基準年の配分比率の固定係数と、実際には変動している中間年次の配分比率の間に乖離が生じた場合、即ち配分比率が安定していない場合には中間年の推計値に誤差が生じる。

上記の観点から、「配分比率の安定性」は、「配分比率の基準年固定係数と中間年の値の乖離から生じる、各需要項目への配分額の誤

差の多寡」として統計の精度に影響を及ぼす。

行部門の設定において、既存の部門数を減らした場合、複数の行部門にまたがっていた品目が1つの行部門に統合され、1つの行部門に属する品目の種類は増加する。このとき、1つの行部門に統合された複数の品目の配分比率の構成がそれぞれ異なっていた場合、プロダクトミックスの影響によって、中間年次の配分額に誤差が生じることとなる。

以下では、現行のJIOの基本分類による取引基本表の行部門数から、一定の客観的なルールに則って部門数を減らした場合（すなわち各部門を統合した場合）、どの程度プロダクトミックスの影響によってコモ法における誤差が生じるのかについて、定量的な分析を行う¹⁰。

3. 分析手法

3.1. 手法概要

本稿ではIOの行部門統合による品目別配分比率の安定性への影響を「凝集型階層クラスタリングの手法を用いた誤差推移シミュレーション」により検証する。

具体的には、「平成23年（2011年）産業連関表」の取引基本表について、基本分類（行：518部門）から、統合小分類（行：108分類）まで、統合の影響が最も少ない組み合わせの統合を階層的に行い、各段階において、部門統合による品目別配分額の乖離（実額、割合）を推計する。

⁸ 以降、中間投入は、建築向け中間投入を除いた値を意味する。なお、政府におけるコモディティ・フロー法では、建設向け中間投入を分離推計しているわけではない。本稿では、データにおいて分離が可能なこと、資本形成との関連性が予想されることなどから分離して推計している。

⁹ 例として、100%中間消費の財Aおよび、100%家計消費の財Bを一つの項目として扱った財Cを想定しプロダクトミックスの影響を考える。財Cの中間消費への配分比率（中間消費/家計消費）は財Aの産出額Pa0、財Bの産出額Pb0の比（Pa0/Pb0）となる。翌年のそれぞれの産出額がpa1>pa0, pb1=pb0となった場合その配分比率は（Pa1/Pb1）>（pa0/pb0）となる。このように各財の産出額は毎年変動するため、プロダクトミックスのある財Cの配分比率は不安定となる。一方、混用財の場合、各配分先の需要は経年で変化するため、プロダクトミックスがなくなると、その配分比率は安定しない。

¹⁰ なお、実際のコモ推計は、JIOの取引基本表の基本分類ではなく、より細かい分類数で推計を行っていることから、実際のコモ推計そのものへの影響とは異なる。本稿では、SUTの部門構成を対象とするため、JIOの部門構成に検証の対象を限定した。

具体的な手順は以下に従う。

- 1) コモ分類の「産業連関表基本分類」への統合
- 2) 階層クラスタリング
 - 2-1) 統合分類候補選出
 - 2-2) 統合による品目別配分額の安定性計測
- 3) 結果の検証

なお、本章で記述される各定義及び表現は、含まれるすべての節にて共通のものとする。

3.2. 使用データ

平成 23 年基準コモ法データ

- ・ 国内総供給額(遡及年補正あり 詳細後述)
- ・ 需要項目別配分比率・配分額
(中間投入(建設向け中間投入を除く), 家計消費, 固定資本形成, 建設向け中間投入)
※遡及年補正あり
- ・ 8 桁分類(約 2000 品目),
- ・ 名目暦年: 1995~2015 年, 生産者価格表示¹¹

平成 23 年基準産業連関表 部門分類 基本分類

3.3. コモ法分類の産業連関表分類への統合

現行の JSNA におけるコモ法では、その配分比率の大半を JIO の取引基本表の配分比率に基づいて推計している。したがって、JIO の取引基本表行部門が統合された場合、同様の影響がコモ法にも及ぶこととなる。ここでは、SUT の部門構成の検討に資するよう、JIO の

基本分類約 500 とコモ法 8 桁分類(約 2000)

を JSNA の基準に従い対応付けを行い、一定の客観的なルールに基づき、対応付け実施後のコモ法における部門を一つに統合する。その際、各行部門の総供給額及び配分額は、統合前の部門の単純和となっている。

なお、品目の配分比率に IO のデータを用いていないコモ品目及び、輸出割合が高く、マイナスの配分額を持つ部門は対象から除外した。

今回使用したコモ法データは、遡及年補正、即ち、過去の中間年次推計値に関して、判明している後年の基準年値を用いた補正をかけたものである。ただし、JSNA の確報及び四半期速報時には次の基準年値を知ることができないため、JSNA に与える影響を分析するために、遡及年補正された値を用いることは適当ではない。

そこで、5 年毎の基準年次において、配分比率の当回基準年値と次回基準年値の差額を基準年の間隔で割った値の等差級数を各中間年次から引く処理を行い補正の効果を打ち消した。

具体的には以下の通りである。

I を整数全体の集合とする。

データに含まれる全年度を

$$TY = [1995, 1996, \dots, 2011] \subset I \quad (3.3-1)$$

基準年を

$$BY = [1995, 2000, 2005, 2011] \subset I \quad (3.3-2)$$

2)

中間年次を、

$$MY = TY - BY \quad (3.3-3)$$

と、それぞれ全順序集合で表す。

このときある基準年に対応する次の基準年を求める関数、Next: $BY \rightarrow BY$ を以下のように

¹¹元データは 2011~2015 年のデータだが、後述の遡及年補正の打消しを行うにあたり、2012 年~2015 年の期間には次の基準年値が存在しないため打消しができないことから 2011 年以降の値を使用していない。

また、通常コモ法における需要項目は購入者価格表示されるため、本稿の生産者価格表示での分析とは結果が多少異なる。

定義する.

$$\text{Next}[y] = \begin{cases} y + 5, & y \neq 2011 \\ y + 6, & y = 2005 \\ y, & y = 2011 \end{cases} \quad (3.3-4)$$

また, 各中間年次に対応する基準年次を求める関数, $\text{getBase}: \text{MY} \rightarrow \text{BY}$ を以下のように定義する.

$$\text{getBase}[y] = \begin{cases} y, & y \in \text{BY} \\ b \text{ s.t. } (y, b) \in \text{BM}, & y \in \text{MY} \end{cases} \quad (3.3-5)$$

where

BM =

{(1996, 1995), (1997, 1995), (1998, 1995),
(1999, 1995), (2001, 2000), (2002, 2000),
(2003, 2000), (2004, 2000), (2006, 2005),
(2007, 2005), (2008, 2005), (2009, 2006)}

需要項目の集合を $\text{Dep} = \{i, ic, h, f\}$ ¹² として, ある需要項目 $x \in \text{Dep}$ に関するある品目の時系列配分額を以下のような全順序集合であらわす. なお, R^+ は非負の実数全体の集合を表す.

$$\text{corrected}[x] = [v_i | i \in \text{TY}, v_i \in R^+] \quad (3.3-6)$$

このとき, 非補正值は以下ようになる,

$$\begin{aligned} \text{uncorrected}[x] &= [\text{revise}[v_i] | v_i \in \text{corrected}[x]] \\ &\text{where,} \end{aligned} \quad (3.3-7)$$

$$\text{getByIndex}[i][S] = v_i \text{ s.t. } v_i \in S$$

$$\begin{aligned} \text{toBeMinus}[i] &= \\ & (i - \text{getBase}[i]) * \\ & \left(\frac{1}{\text{Next}[\text{getBase}[i]] - \text{getBase}[i]} \right) * \\ & (\text{getByIndex}[\text{getBase}[i]][\text{corrected}] \\ & - \text{getByIndex}[\text{Next}[\text{getBase}[i]]][\text{corrected}]) \end{aligned}$$

$$\text{revise}[v_i] = \begin{cases} v_i, & i \in \text{BY} \\ v_i - \text{toBeMinus}[i], & i \in \text{MY} \end{cases}$$

以降, この統合された JIO 基本分類(約 500 分類, 以下「統合基準分類」.) を検証の対象とする.

3. 4. 階層クラスタリング

本稿では, 統合の過程における段階的な変化に分析の関心があるため, 凝集型階層クラスタリング¹³の手法を用いて, 段階的に部門の統合を行い, その各段階で発生する誤差額を計算するシミュレーションを行う.

階層クラスタリングは以下のアルゴリズムに則って行う.

¹² それぞれ, intermediate demand, intermediate demand for construction, household consumption, fixed capital formation, の頭文字.

¹³ 階層クラスタリングは特定の空間上の各点を, 何らかの距離(類似度)指標の近い組み合わせから階層的にクラスタリングする手法一般を指す. 階層クラスタリングに関する説明は, Jain (1999) 5.1 章などに簡潔にまとめられている. 階層クラスタリングにおいては距離の取り方が設計の重要な要素であり, ウォード法や, 重心法などいくつかの代表的な手法が存在する. 本稿では, 対象のクラスタが統合された行部門を意味し, それぞれに時系列値を持つという特殊性から距離の取り方においても, 後に説明される独自の方法をとっている. 通常階層クラスタリングでは, 何らかの基準を用いて統合を途中で停止することで, 元データを K 個のクラスタにクラスタリングする. 従って, そのような基準を用いない場合, 全てのデータ点は最終的に単一のクラスタに統合されることとなり, その過程を表すデンドログラムを得ることが目的となることが多い. 本稿は, 統合の過程を分析することを目的としており, 本稿における分析では, 後に説明される事前に定められた約 100 のクラスタに分割された時点で統合は停止するため, 最終的な分類結果自体は意味を持たない. 距離指標に関しては付録にて論じている.

- 1) 部門の全体集合 $S[t]$, (初回 $t = 0$) から統合候補 α, β を選出
- 2) 統合候補 α, β を統合し, 新たな部門 $\widehat{\alpha\beta}$ を作成¹⁴.
- 3) $S[t + 1] = (S[t] - \{\alpha, \beta\}) \cup \{\widehat{\alpha\beta}\}$ として更新する.
- 4) 統合が不可能になるまで 1)~3)を繰り返し各 t における誤差額を推計する.

以下, 各手順に関して詳細を論じる.

3.4.1. 統合分類候補の選出

本節では, 上記 1) 全体集合 S から統合候補 α, β の選出について述べる.

1つに統合される2つの行分類は統合の影響を最小限に止め, かつ現実的な統合案を策定するために, 以下の3点の基準に則り選出する.

- ① 概念的な類似性,
- ② 定量的な類似性,
- ③ 統合後総供給量のスケール,

① 概念的な類似性

統合された品目の組み合わせの概念的な類似性を担保するため, 統合基準分類に属する行分類のうち, 「産業連関表統合小分類 (108 部門)」の各部門に属する品目内の組み合わせを統合対象候補とする. 言い換えれば, IO の統合小分類の各分類に属する品目同士の統合は行われるが, 統合小分類を超えての統合は行われないということである. これは, 用途の類似性, 代替可能性等の基準で区分された上位分類の中で統合候補を選出す

ることで, 概念的に全く異なる行部門を統合することを避けることを目的としており, これによって, 配分比率が似ているが, 用途が全く異なるような部門間の統合がなされることを一定程度防いでいる.

② 定量的な類似性

上記①の条件を踏まえ, 以下, 部門間の類似性について整理する. 部門が統合された際に, それらの部門が双方とも同様の需要項目に配分される単用財であった場合, 部門間の統合による推計誤差は生じない. したがって, 可能な限り, 統合される部門間の配分比率構造は類似していることが望ましい.

本稿では, 行部門同士の類似度として, 基準年の平均配分比率ベクトルのユークリッド距離を用いている¹⁵. ユークリッド距離とは, ユークリッド空間上の距離尺度であり, ここでは, 距離が近いほど, 配分比率構造が似ていることを示す.

定義された全ての品目の集合を S とする. y 年における品目 $s \in S$ を以下のように定義する.

$$s = \{ \{y, v_i, v_{ic}, v_h, v_f\} | y \in TY, \\ v_i \in \text{uncorrected}[i], \\ v_{ic} \in \text{uncorrected}[ic], \\ v_h \in \text{uncorrected}[h], \\ v_f \in \text{uncorrected}[f] \} \quad (3.4.1-1)$$

$\{y, v_i, v_{ic}, v_h, v_f\} \in s$ は y 年における各需要項目に対する配分額を表現しており, $v_i \sim v_f$ は各需要項目 $i \sim f$ の配分額を表す.

ある行部門 s に関する y 年の各需要項目に

¹⁴ $\widehat{\alpha\beta}$ は, 以後, α と β が統合され, 1つの部門になった新しい部門を表すために用いる.

¹⁵ ユークリッド距離は IO の分析においてよく用いられる指標であるが, クラスタリングにおいて一般的な指標の一つにコサイン類似度なども存在する. 付録において, それぞれの指標の比較と意味に対する考察を行っている. 通常, ユークリッド距離を用いる際には, 正規化及び, 同方向のベクトルの扱いへの対応が必要とされるが本稿の分析では, ベクトルが比率を表す (0~1の値を持ち, 和が1になる) という制約によってユークリッド距離とコサイン距離の間で結果にほぼ差はない.

関する配分額は,

$$\text{Proj}[s][y] = \{y, v_i, v_{ic}, v_h, v_f\}, \\ \text{s.t. } \{y, v_i, v_{ic}, v_h, v_f\} \in s \quad (3.4.1-2)$$

によって表される.

需要項目 $x \in \text{Dep}$ の配分額は,

$$\text{Alloc}[s][y][x] = v_x, \text{s.t. } v_x \in \text{Proj}[s][y]$$

と表現できる.

また, ある行部門の y 年における総供給額は,

$$\text{Total}[s][y] = \sum_{x \in \text{Dep}} \text{Alloc}[s][y] \quad (3.4.1-3)$$

となる.

ある部門 $\alpha, \beta \in S$ の各需要項目に対する代表配分比率をそれぞれ,

$$\vec{R}_\alpha = (a_i, a_{ic}, a_h, a_f) \quad (3.4.1-4)$$

$$\vec{R}_\beta = (b_i, b_{ic}, b_h, b_f) \quad (3.4.1-5)$$

Where

$$\text{BTotal}[x] = \sum_{y \in \text{BY}} \text{Total}[x][y]$$

$$\text{BVal}[x][i] = \sum_{y \in \text{BY}} \text{Alloc}[x][y][i]$$

$$\alpha_i = \frac{\text{BVal}[\alpha][i]}{\text{BTotal}[\alpha]}, i \in \text{Dep}$$

$$\beta_i = \frac{\text{BVal}[\beta][i]}{\text{BTotal}[\beta]}, i \in \text{Dep}$$

とすると,

$$\text{Euc}[\alpha \beta] \text{ は,}$$

$$\text{Euc}[\alpha \beta]$$

$$= \sqrt{\sum_{x \in \text{Dep}, \alpha_x \in \vec{R}_\alpha, \beta_x \in \vec{R}_\beta} (\alpha_x - \beta_x)^2} \quad (3.4.1-6)$$

となる.

③ 統合後総供給量のスケール

誤差額は, 配分比率の差と総供給額の積によって求められるため, 仮に統合された2つの部門の配分構造が大きく異なっても, 双方の総供給額が0に近い場合, 誤差の影響は僅かしか生じない. 同様に, 配分構造差が僅かな行部門同士でも, その総供給額が非常に大きい場合, 生じる誤差は大きくなる.

したがって, 統合の対象となる部門選定に当たっては, 可能な限り総供給額の小さな行部門から統合していくのが好ましいといえる.

「平成27年(2015年)産業連関表作成基本要綱」¹⁶には, 平成27年表における部門分類の基準として, 「国内生産額又は総需要額の大きさ」を掲げており, その中において「既に設けられている部門について, 列の国内生産額及び行の総需要額が減少して, 各々1000億円未満となった場合には, 原則として類似の既存部門に統合する。」と記載されており, 部門の国内生産額及び総需要額の桁数が部門設定の1つの目安となっていることが示されている. そこで以下のとおり, 行部門 α, β の基準年総供給額平均の和の常用対数値を1つの基準として用いた.

$$\text{Scale}[\{\alpha, \beta\}] \\ = \log_{10}(\text{BTotalAve}[\alpha] + \text{BTotalAve}[\beta]) \quad (3.4.1-7)$$

Where

$$\text{BTotalAve}[x] = \frac{\sum_{y \in \text{BY}} \text{Total}[x][y]}{|\text{BY}|}$$

¹⁶ 『平成27年(2015年)産業連関表作成基本要綱 第2部 産業連関表の作成手順及び作業内容』85頁より引用

以上の「②定量的な類似性」,「③統合後総供給量のスケール」の基準を統合し, 行部門の α, β の統合指標を,

$$\begin{aligned} \text{Sim}[\{\alpha, \beta\}] \\ = \text{Euc}[\alpha, \beta] * \text{Scale}[\{\alpha, \beta\}] \end{aligned} \quad (3.4.1-8)$$

とすると $\text{Sim}[\{\alpha, \beta\}]$ は, 配分率の構成が似ており, かつ, 総供給額が小さいほど小さな値をとる. 統合後に生じる誤差額は, スケール及び, 配分比率の差異の度合によって決まるため, この指標は統合の結果発生する誤差額を模している.

以上すべてを統合し, 同一の統合中分類に属する部門の全ての組み合わせにおいて, $\text{Sim}[\alpha, \beta]$ を計測し, 同値が最も小さい組み合わせを部門統合の対象として選定する.

3.4.2. 統合方法

次に, 以下に定めるように部門を統合する.
 $x, y \in S$ が同様の「産業連関表基本分類小分類」に属するとき $x \sim y$ と表現すると, 各基本小分類は商集合 $S/\sim$ に属する.

このとき, 品目 $\omega \in \Omega \in S/\sim$ を

$$\begin{aligned} \omega = \{ \{y, v_i, v_{ic}, v_h, v_f\} \mid y \in Y \\ , v_i \in \text{uncorrected}[i] \\ , v_{ic} \in \text{uncorrected}[ic] \\ , v_h \in \text{uncorrected}[h] \\ , v_f \in \text{uncorrected}[f] \} \end{aligned} \quad (3.4.2-1)$$

とする.

2つの部門 $\alpha, \beta \in \Omega \in S$ が統合された新たな部門 $\widehat{\alpha\beta}$ を考える.

$\widehat{\alpha\beta}$ の配分額は, 基準年次においては, α, β それぞれの配分額の単純和, 中間年次においては, 対応する基準年次の配分率を, 2つの部門の総供給額の和に乗じたものとなる.

統合後の配分比率を求める関数
 $\text{hatAlloc} : \Omega \times TY \times \text{Dep} \rightarrow R^+$

を以下のように定義する.

$$\begin{aligned} \text{hatAlloc}[\{\alpha, \beta\}][y][x] \\ = \begin{cases} \sum_{\omega \in \{\alpha, \beta\}} \text{Alloc}[\omega][y][x] , & y \in BY \\ \frac{\sum_{\omega \in \{\alpha, \beta\}} \text{Alloc}[\omega][\text{getBase}[y]][x]}{\sum_{\omega \in \{\alpha, \beta\}} \text{Total}[\omega][\text{getBase}[y]]} * \\ \sum_{\omega \in \{\alpha, \beta\}} \text{Total}[\omega][y] , & y \in MY \end{cases} \end{aligned} \quad (3.4.2-2)$$

このとき,

$$\begin{aligned} \widehat{\alpha\beta} &= \text{merged}[\{\alpha, \beta\}] \\ &= \{ (y \\ &\quad , \text{hatAlloc}[\alpha, \beta][y][i] \\ &\quad , \text{hatAlloc}[\alpha, \beta][y][ic] \\ &\quad , \text{hatAlloc}[\alpha, \beta][y][h] \\ &\quad , \text{hatAlloc}[\alpha, \beta][y][f]) \\ &\quad \mid y \in TY \\ &\quad , i, ic, h, f \in \text{Dep} \} \end{aligned} \quad (3.4.2-3)$$

となる.

3.4.3. 部門統合による誤差額の計測

誤差額を以下のように定義する.

$\alpha, \beta \in \Omega \in S/\sim$ が統合された部門 $\widehat{\alpha\beta}$ の配分額に含まれる y 年の誤差は, α, β がともに1度も統合されていない行部門である場合は, 下記に示すように各部門の差額の絶対値の和となる.(誤差の原因は配分比率の乖離であるため, 絶対値をとらない誤差額の総和は, 増加分, 減少分が打ち消しあい, 常に0となる. すなわち, コモ法における総供給額にこの誤差額は影響を及ぼさず, 各需要項目の値にのみ影響を与える.)

$$\begin{aligned} \text{Diff}[\{\alpha, \beta\}][y] &= \\ \sum_{x \in \text{Dep}} & \left| \left(\sum_{\omega \in \{\alpha, \beta\}} \text{Alloc}[\omega][y][x] \right) - \text{Alloc}[\widehat{\alpha\beta}][y][x] \right| \\ \text{Diff}[\phi][y] &= 0 \end{aligned} \quad (3.4.3-1)$$

ただし, $\{\alpha, \beta\}$ が既に統合された部門を含む場合, その誤差に, 過去に統合された際に生じている誤差の影響を加える必要がある。

正確な計算のためには, $S[0]$ から $\widehat{\alpha\beta}$ が生成されるまでの全統合を追跡し, その差額を求めるべきであるが, 本稿では各段階で生じた誤差の累計を, 過去の累積誤差として, 最新段階で生じた誤差に加えている。

本稿で行われる階層クラスタリングでは, 概念的類似性の制約から, 統合基本分類小分類より少ない部門数への統合は行うことができない。したがって, 上記の部門統合及び配分額の誤差の計算を繰り返すと, 部門数は, 初期の基本分類に基づく約 500 部門から, 統合部門小分類に基づく約 100 部門まで推移することとなる。

初期の全行部門の集合を $S[0]$, 最初に統合される部門のペアを α, β とすると,

$S[1] = (S[0] - \{\alpha, \beta\}) \cup \{\widehat{\alpha\beta}\}$ となる。

$|S[t]| = n$ とすると, $|S[t+1]| = n-1$ となるため, 統合は, 約 400 回行われることとなる。

$S[t], t \in I, 0 \leq t < |S/\sim|$ において統合される部門統合候補のペアは, 以下のように定義される。

$$\begin{aligned} \text{canBeMerged}[S[t]] \\ = \{ \{\alpha, \beta\} \mid \alpha, \beta \in \Omega, \Omega \in S[t]/\sim, \alpha \neq \beta \} \end{aligned} \quad (3.4.3-2)$$

このとき, $S[t]$ において統合される, 部門のペアは,

$$\begin{aligned} \text{Pair}[S[t]] &= \{ \alpha, \beta \} \\ \text{s.t} \\ \{ \alpha, \beta \} &\in \text{canBeMerged}[S[t]], \\ \forall \{ x, y \} &\in \text{canBeMerged}[S[t]], \\ \text{Sim}[\{ \alpha, \beta \}] + \varepsilon &> \text{Sim}[\{ x, y \}] \end{aligned} \quad (3.4.3-3)$$

となる¹⁷。

また, $S[t]$ までに過去に統合された, 全てのペアの集合は,

$$\text{pairs}[S[t]] = \bigcup_{x=0}^t \{ \text{Pair}[S[x]] \} \quad (3.4.3-4)$$

となる。

ある部門 α が過去に統合してきた, 部門の集合は以下のようにあらわされる。

$$\begin{aligned} \text{includedAll}[\alpha][t]^{18} \\ = \text{includedPair}[\alpha][t] \cup \\ \bigcup_{r \in \text{includedPair}[\alpha][t]} \text{includedAll}[r][t-1] \\ \text{includedAll}[x][0] = \{ \phi \} \end{aligned} \quad (3.4.3-5)$$

$$\begin{aligned} \text{includedPair}[\alpha] = \\ \begin{cases} \{ \{ p, q \} \mid \{ p, q \} \in \text{pairs}[S[t]] \wedge \text{merged}[\{ p, q \}] = \alpha \} \\ \{ \phi \}, \end{cases} \quad \text{otherwise} \end{aligned}$$

このとき, 品目 α, β を統合した部門に含まれる $S[t]$ における y 年の誤差額は,

¹⁷ この時 ε は候補を一つに絞る為の調整項であり, プログラム上では, 値が同じ場合, 後に処理された方が優先されるよう定めている。

¹⁸ includedAll は再帰関数であり, $t=0$ から計算が行われる。

$$\begin{aligned} \text{diffSum}[\widehat{\alpha\beta}][t][y] \\ = \sum_{x \in X} \text{diff}[x][y] \end{aligned}$$

(3.4.3-6)

Where

$$X = \text{includedAll}[\widehat{\alpha\beta}][t]$$

となる。

また、全年度に渡るこの部門の誤差額は、以下の通りとなる、

$$\text{diffAllY}[\widehat{\alpha\beta}] = \sum_{y \in TY} \text{diffSum}[\widehat{\alpha\beta}][y] \quad (3.4.3-7)$$

3.4.4. 乖離指標

クラスタリングの各階層段階において、以下のとおり、乖離指標（実額、割合）を推計する。

① 乖離総額（年平均）；

誤差の累積額を年数で割った値¹⁹。

$$\begin{aligned} \text{totalDiff}[S[t]] \\ = \sum_{x \in S[t]} \text{diffAllY}[x] / |TY| \end{aligned} \quad (3.4.4-1)$$

② 行部門総額に対する乖離額の割合（年平均）；

各行部門 $\omega \in S[t]$ に関して、行部門の総供給額に占める累計誤差の平均。

$$\begin{aligned} \text{diffRCol}[S[t]] \\ = \frac{\sum_{\omega \in S[t]} (\text{diffAllY}[\text{includedPair}[\omega]] / \text{Total}[\omega])}{|S[t]| * |TY|} \end{aligned} \quad (3.4.4-2)$$

③ 各需要項目別・配分総額に対する乖離額の割合（年平均）；

各需要項目 $x \in \text{Dep}$ に関しての配分総額に占める乖離額の割合。

統合した際の特定の需要項目における非累積誤差を

$$\begin{aligned} \text{diffDep}[\{\alpha, \beta\}][x][y] &= \left| \left(\sum_{\omega \in \{\alpha, \beta\}} \text{Alloc}[\omega][y][x] \right) - \text{Alloc}[\widehat{\alpha\beta}][y][x] \right| \\ \text{diffDep}[\phi][x][y] &= 0 \end{aligned} \quad (3.4.4-3)$$

とすると、累積誤差は、

$$\begin{aligned} \text{diffSumDep}[\{\alpha, \beta\}][x][t][y] \\ = \sum_{x \in X} \text{diffDep}[\{\alpha, \beta\}][x][y] \\ \text{diffSumDep}[\phi][x][t][y] = 0 \end{aligned} \quad (3.4.4-4)$$

Where

$$X = \text{includedAll}[\text{merged}[\{\alpha, \beta\}][t]]$$

全年度値は、

$$\begin{aligned} \text{diffDepAllY}[\{\alpha, \beta\}][x] \\ = \sum_{y \in TY} \text{diffSumDep}[\{\alpha, \beta\}][y] \end{aligned} \quad (3.4.4-5)$$

となる。

¹⁹ 基準年には誤差が生じないため、この値は誤差が生じている年の平均誤差額よりは少なめの値となる。基準年の誤差を除いた値を求めるのが望ましいが、取り除く際に計算がより複雑になること、傾向やスケールは変わらないことなどから、基準年を含めて除している。以降、全ての指標においても同様の対処をしており、したがって目安として13/15 程値が小さくなっている。

このとき、求める値は、以下で表される、

$$\begin{aligned} & \text{diffRDep}[S[t]][x] \\ &= \frac{\sum_{\omega \in S[t]} \text{diffDepAllY}[\{\text{includedPair}[\omega]\}][x]}{\sum_{y \in TY} \sum_{\omega \in S[t]} \text{Alloc}[\omega][y][x]} \end{aligned} \quad (3.4.4-6)$$

4. 分析結果

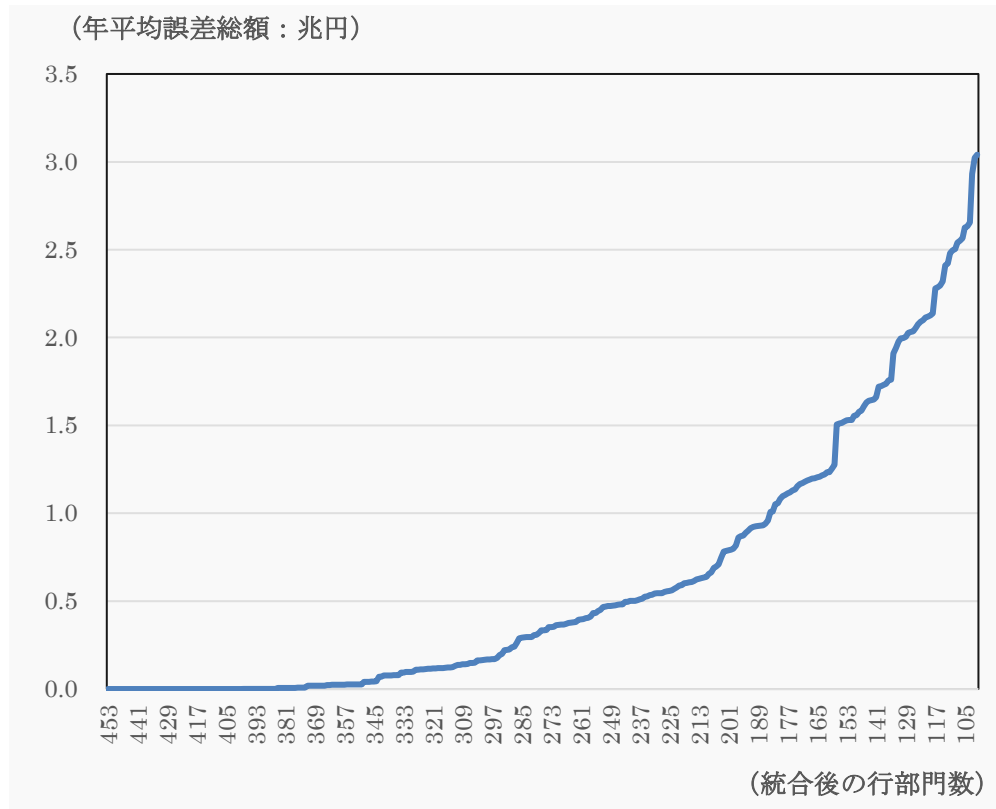
上記定義に従い、階層クラスタリング用プ

ログラムを作成・実行し、その結果として定義された各指標を以下図 1～3 にプロットした。

また、図 4 において、図 1 における各 $S[t]$ において、統合された部門を 1～3 次産業に分類し、色別に表示している。また、凡例として各段階において、統合された代表的な部門名を例示した²⁰。

結果に対する考察は、次章にて行う。

図 1 乖離総額（年平均）



²⁰ ただし、2つの部門を統合した部門の呼称は、プログラム上で偶然先に計算されたものを採用しており、あくまで一時的なものであることに留意が必要。階層アルゴリズムの後半に登場するほど同一の名称であっても、複数の部門が統合された部門となっている。

図 2 行部門総額に対する乖離額の割合（年平均）

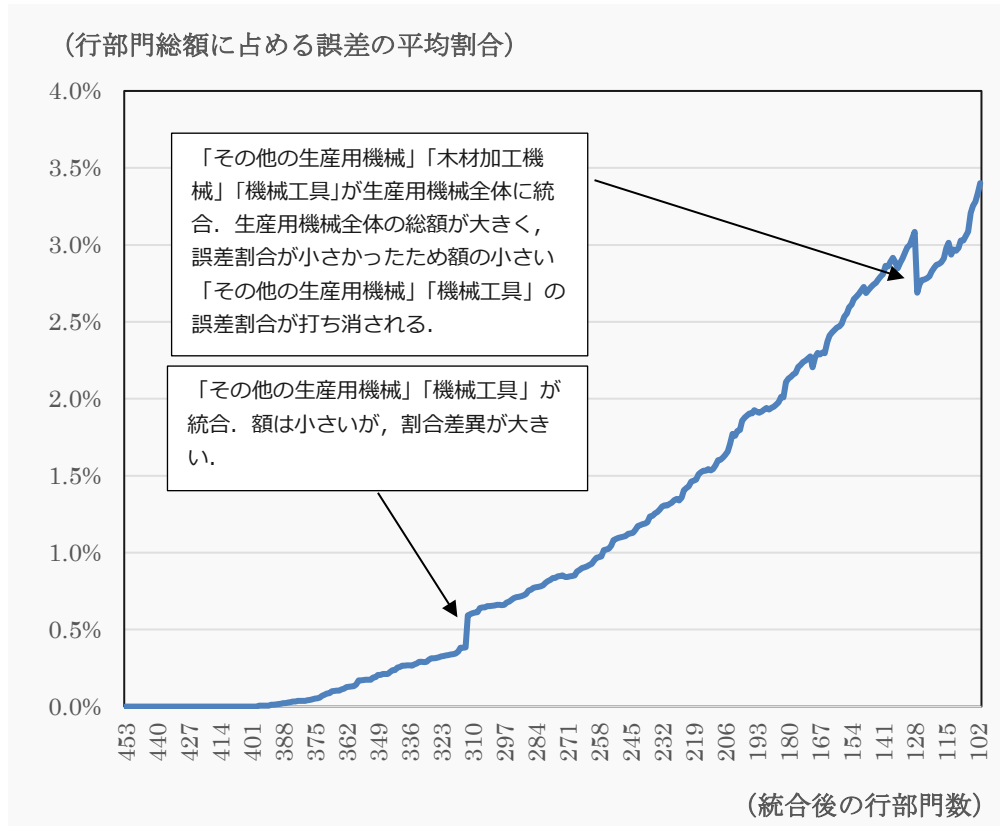


図 3 各需要項目別・配分額総額に対する乖離額の割合（年平均）

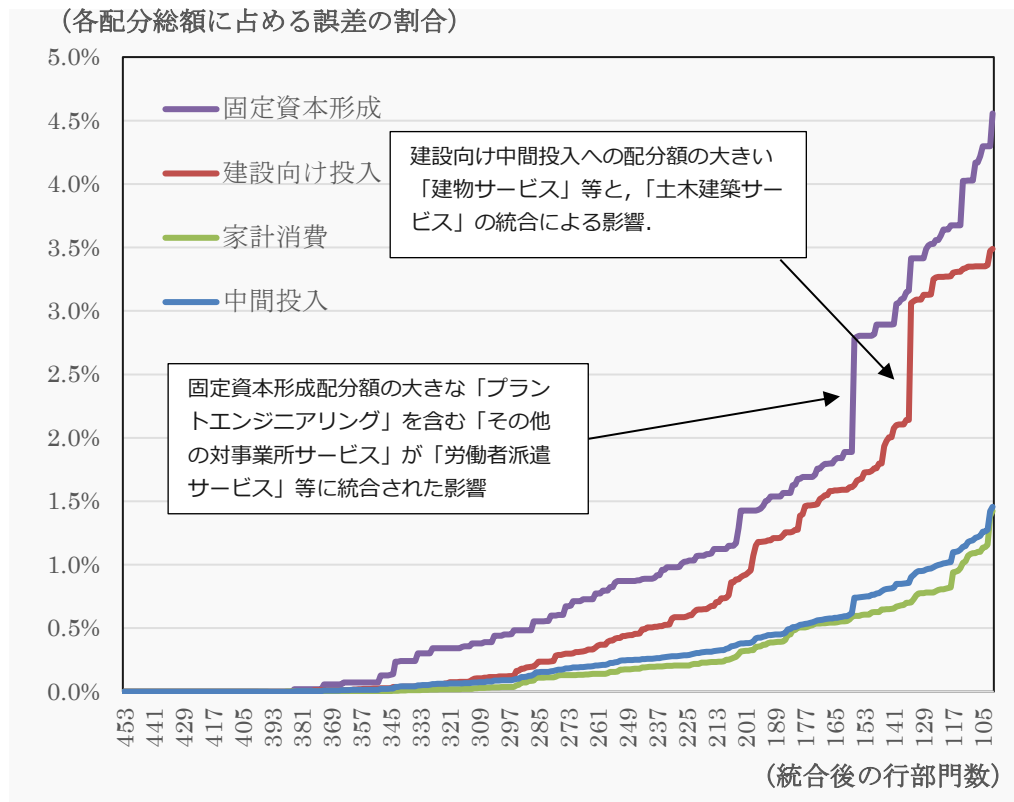
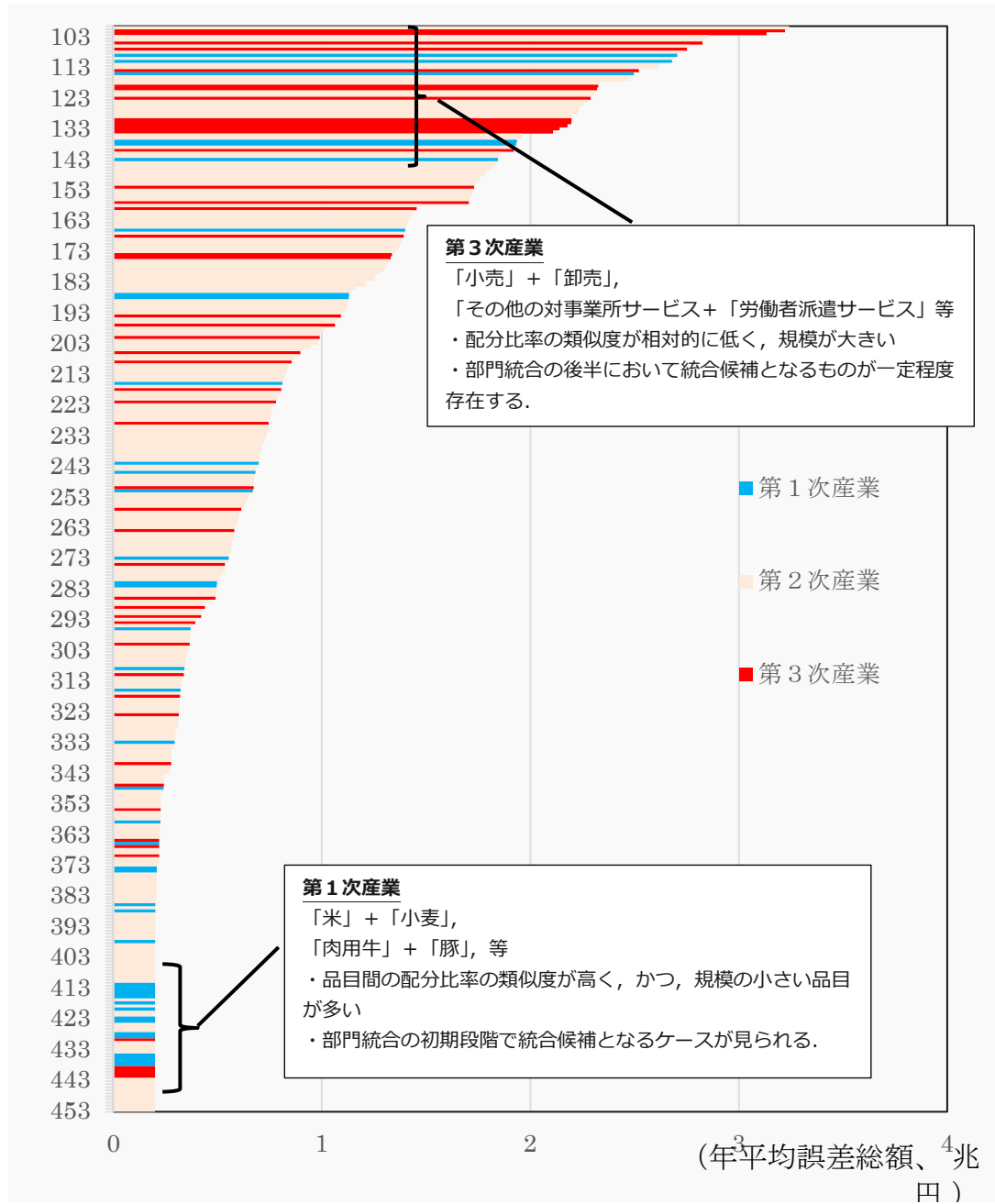


図 4 図 1 における統合された産業ダイアグラム



5. 考察とまとめ

5.1. 乖離総額

図1より、現行の「産業連関表基本分類」の行部門（約500）から部門統合することにより、統合された部門品目内のプロダクトミックスによる推計誤差が拡大し、品目別配分額乖離（実額、割合）は一定して拡大傾向にあることが分かる。

なお、3.4.2節で説明されたように、この誤差額はあくまで需要項目ごとに配分される額の中での乖離であるため、総供給額そのものは全く増減していない点に留意が必要である。

また、統合の初期段階、およそ部門数が400に至るまでの統合段階では、実額、割合ともにほとんど誤差が発生しないことも見て取れる。

当然のことながら、IOの行部門統合の影響は、コモ法の配分比率推計のみに限定されるものではなく、投入係数の安定性や、実査における報告者負担、総供給額の精度等との比較の上で総合的に判断される事項である。

5.2. 行部門の配分額総額に占める乖離額の割合

図2より、部門統合を進めると、行部門の配分額総額に対する乖離額（合計）は約250部門で1%、約180部門で2%を超え、100部門まで統合を進めると約3.5%まで増加する。これは、各部門の配分額が統合の仕方によっては平均で約3.5%程度変動することを意味しており、部門数の設定によって各行の推計精度が大きく低下することが分かる。

5.3. 需要項目別にみた乖離額の割合

図3 需要項目別（中間投入、家計消費、固定資本形成、建設向け中間投入）の配分額の

乖離の割合をみると、「建設向け中間投入と固定資本形成」、「中間投入と家計消費」が連動する傾向があり、混用財（あるいはプロダクトミックスされた品目）においては、この組み合わせの双方を配分先に持つ品目が多いことが分かる。

中間投入や家計消費と比較して、建設向け中間投入及び固定資本形成の乖離額の割合の増加傾向は大きく、約100部門まで統合を進めた場合、建設向け中間投入は約3.5%、固定資本形成は約4.5%の割合で推計誤差が含まれることとなる。

また、この2つの需要項目においては、それぞれ乖離額の割合が大きく上昇する統合のタイミングが存在することが分かる²¹。

固定資本形成では部門数173となる統合段階において、「労働者派遣サービス、法務・財務・会計サービス」が統合された部門と、「その他の対事業所サービス」が統合された際にその乖離額の割合が大きく上昇している。「その他の対事業所サービス」には「プラントエンジニアリング」が含まれており、プラントエンジニアリングは固定資本形成への単用財であり、かつその総供給額が大きいことからこのような変動が生じている。

また建設向け中間投入においては、部門数135に統合する段階において、「建物サービス、法務・財務・会計サービス、労働者派遣サービス、その他の対事業所サービス」からなる部門と「土木建築サービス」が統合されている。

土木建築サービスと対応するコモ法の部門は、建設向け中間投入における総供給額の規模が最大の部門であり、同部門の統合の影響を受け、建設向け中間投入の乖離額の割合が大きく上昇している。

このように、乖離額だけではなく、割合でみることで、統合すると特定の需要項目の配分額推計精度に大きく影響を及ぼす品目を明らかにすることも可能であり、そのような部

²¹本稿におけるクラスタリングでは、実額評価での乖離の増加がある程度滑らかな上昇関数になるように評価関数を設計し統合を行っているため、各種割合においては、このような揺れ動きが観察されることは正常である。

門は、部門設計の際に統合せず、可能な限り単独の部門として設置することが好ましいといえる。

5. 4. 産業別にみた統合部門の特徴

図 4 は、図 1 における各統合段階において、第 1, 2, 3 次産業の部門が統合されたかを示すダイアグラムである。品目別にみると、「米」「小麦」「肉用牛」「豚」など、第 1 次産業に属する品目は、品目間の配分比率の類似度が高く、かつ、規模が小さな品目が多く、階層アルゴリズムにおける初期段階で統合候補となるケースが見られる。

第 2 次産業は、品目数自体が多いこともあり、統合段階に全体的に広く分布している。

一方、サービス産業（第 3 次産業）は、配分比率の類似度が相対的に低く、かつ、規模の大きい品目（「卸売」「小売」「その他の対事業所サービス」「労働者派遣サービス」など）が後半に集中して統合される傾向が見える。

このように、産業毎に統合によって発生する誤差には一定の傾向が存在する。部門の検討に当たっては、類似度が高く、かつ規模の小さな品目から統合候補とするべきであり、その該当品目の多い、1 次、2 次産業を優先的に統合することが望ましいと考えられる。

5. 5. まとめ

これまでの検証結果に対する考察を以下にまとめた。

第一に、現行の IO 基本分類（約 500 分類）から部門統合（部門数の削減）を行うことにより、品目別の配分額の乖離（実額、割合）は拡大傾向がみられた。基準年・中間年 SUT の部門構成の検討に当たっては、現行の IO の部門数を保ちつつ、可能な限りプロダクトミックスを避けた単独の需要項目に配分されるような行部門の設定が望ましい。

特に、プラントエンジニアリングなど、現行の IO 基本分類では統合されている部門に

関しては、分離して単独の部門を設定とすることで、精度の向上が見込める可能性が高い。また、類似度や総供給額の規模からみて、統合による推計精度低下への影響の強い部門が存在するが、これらについては、部門設定の際に部門を統合することのないよう、十分留意することが好ましい。

今回の検証では、特に第 1・2 次産業において、品目間の配分比率の類似度が高く、部門統合による影響（乖離）が小さい品目が一定程度存在することも判明した。類似の用途であるなど統合の必然性があり、かつ同一の需要項目に配分される分類同士は、部門を統合しても、配分比率の安定性の観点からは支障がないため、基礎統計の実査における企業負担など等を考慮した統合が必要となる場合、これらの部門を優先的に統合の候補とすることが好ましい。

5. 6. 留保された課題及び今後の展開

今回の検証は、一定の仮定の下で全体の影響を見たものであり、評価関数をはじめとするアルゴリズムの設計及び分類基準の変更により結果は変動する。例えば、概念的な類似性の基準として使用されている、「産業連関表統合小分類」を同中分類に変更することにより、結果は変わる。これは概念的な類似性の基準を緩める、すなわち類似しているものの範囲を広めることを意味し、より用途や代替性の観点からは遠い部門を統合することを許すことになる。この場合、より広い範囲から統合候補を探索することとなり、同一の配分先を持つ単用財の組み合わせが増えることから、統合の影響の少ない部門の組み合わせが増加し、各グラフの上昇の傾きが比較的緩やかになることが推定される。概念的な類似性を重視する程度に関する明確な基準は存在しないため、どのような基準を採用するかに関しては議論が必要である。

また、本稿の検証は、あくまで部門統合における推計精度への影響を全体的に把握する

ことを目的として行われたものである。

したがって、部門構成の具体的な検討に当たっては、本稿の機械的な判断はあくまで目安、参考として利用されるものであり、用途の類似性、国際基準への対応、今後の成長可能性など様々な基準に則って、個別品目の特性を考慮した詳細な検討が必要であることは論を俟たない。現実には、本稿の分析結果でも、特定の品目を統合した結果、乖離幅が急激に拡大している場合が複数あることが見て取れ、具体的な品目構成を検証する際に一定の示唆を与えている。

報告者負担の軽減を図るため、GDP 精度から品目の統合が可能とみられる分類など等についても、本稿の検証は、「統合しても配分比率の安定性の観点からは影響の少ない部門」が存在することを指すのみであり、具体的に特定の部門の統合を奨励するものではなく、統合される部門及び部門数の設定は他の影響を鑑みて、個別に検討する必要がある。

また、本稿では、部門の統合に関する検証を行っているが、検証されているのはあくまで現行の分類からの乖離額であり、現行の分類における既にプロダクトミックスを含む部門などに関しては、むしろそれらを分離することが好ましいこともあり、実際の部門数に関しては、現行の分類の分離、統合両面から考える必要がある。

これら、個別の品目に関する検証及び、分離が好ましい品目の特定、検証など等に関しては引き続き研究していきたい。

また、本稿の執筆は、国民経済計算体系的整備部会 SUT タスクフォース会合にてその内容を発表させていただき、宮川座長をはじめとする出席者の方々から、行のみではなく、列も指標に入れて部門を論じるべきである点、時系列の変化を分析の対象に取り入れるべきである点、分散等の指標によって、平均値以外の指標を分析するべきである点など大変有益なご指摘を受けた。

本稿では、扱うことができなかったため、それらも合わせて今後の課題としたい。

6. 参考文献

- Inter-Secretariat Working Group on National Accounts. (1993). System of National Accounts 1993 - 1993 SNA. Brussels/Luxembourg, New York, Paris, Washington, D.C.: Inter-Secretariat Working Group on National Accounts.
- Jain,A.K,et al. (1999). *Data Clustering: A Review*. ACM Computing Srveys, Vol. 31, No. 3.
- United Nations Statistics Division. (1999). Handbook of input-Output Table Compilation and Analysis. New York: United Nations Statistics Division.
- 吉岡徹哉, 鈴木俊光. (2016). 供給・使用表 (SUT)の枠組みを活用した支出側 GDPと生産側 GDPの統合. 内閣府経済社会総合研究所 「季刊国民経済計算」, 第160号.
- 熊谷章太郎. (2012). 年次SUTを用いたバランスシステムの在り方の一考察. 基幹国民経済計算 147号.
- 経済財政諮問会議. (平成28年). 統計改革の基本方針. 経済財政諮問会議. 経済財政諮問会議.
- 経済財政諮問会議. (平成28年). 統計改革の基本方針. 経済財政諮問会議.
- 産業連関部局長会議. (平成29年). 平成27年(2015年)産業連関表作成基本要綱. 参 照 先 :
- http://www.soumu.go.jp/toukei_toukatsu/data/io/youkou_2015.htm
- 菅幹雄. (2009). 産業連関表作成のための特別調査(投入調査)の現状. 環太平洋産業連関学会, 産業連関, Vol 17, No. 3.
- 清水雅彦, 宮川幸三. (2008). 日本の産業連関表について 基本表と供給・使用表の関係. 環太平洋産業連関学会, 産業連関, Vol. 16, No. 3.
- 総務省. (日付不明). 平成23年(2011年)産業連関表(確報),部門分類コード表. 参 照 先 : <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200603&tstat=00001073129&cycle=0>
- 統計改革推進会議. (平成29年5月19日). 第3回統計改革推進会議 資料 統計改革推進会議最終とりまとめ. 第3回統計改革推進会議 (ページ: 7). 統計改革推進会議.
- 内閣府. (平成29年6月9日閣議決定). 経済財政運営と改革の基本方針 2017～人材への投資を通じた生産性向上～(骨太方針). 内閣府.
- 内閣府統計委員会. (平成30年3月6日閣議決定). 「公的統計の整備に関する基本的な計画」第Ⅲ期基本計画.
- 櫻本健. (2012). 日本の国民経済計算体系における供給使用表年次表に関する研究. 内閣府 New ESRI Working Paper, No.26.

7. 付録

本稿では、IO 分析において一般的なユークリッド距離を配分比率の類似度指標として用いているが、クラスタリングにおいてベクトル間のノルムを表現する指標には他にもいくつか一般的なものが存在する。本付録では、その代表的なものとして2つのベクトルの角度を用いるコサイン距離との比較を行い、指標の選択の根拠を示す。

3.4 節において定義した通り、ある部門 $\alpha, \beta \in S$ の各需要項目に対する代表配分比率をそれぞれ、

$$\begin{aligned}\vec{R}_\alpha &= (a_i, a_{ic}, a_h, a_f) \\ \vec{R}_\beta &= (b_i, b_{ic}, b_h, b_f)\end{aligned}$$

Where

$$BTotal[x] = \sum_{y \in BY} Total[x][y]$$

$$BVal[x][i] = \sum_{y \in BY} Alloc[x][y][i]$$

$$\alpha_i = \frac{BVal[\alpha][i]}{BTotal[\alpha]}, i \in Dep$$

$$\beta_i = \frac{BVal[\beta][i]}{BTotal[\beta]}, i \in Dep$$

とすると、ユークリッド距離 $Euc[\alpha \beta]$ は、

$$\begin{aligned}Euc[\alpha \beta] \\ = \sqrt{\sum_{x \in Dep, \alpha_x \in \vec{R}_\alpha, \beta_x \in \vec{R}_\beta} (\alpha_x - \beta_x)^2}\end{aligned}$$

となる。

ここで、コサイン類似度 $Cos[\alpha, \beta]$ は

$$\begin{aligned}Cos[\alpha, \beta] &= \frac{\vec{R}_\alpha \cdot \vec{R}_\beta}{|\vec{R}_\alpha| |\vec{R}_\beta|} \\ &= \frac{\sum_{x \in Dep} a_x b_x}{\sqrt{\sum_{x \in Dep} a_x^2} \cdot \sqrt{\sum_{x \in Dep} b_x^2}}\end{aligned}\quad (7-1)$$

となり、コサイン距離は2つのベクトルの方向が完全に一致したとき、1 をとり、直行するとき 0 をとる。

このとき、コサイン類似度を用いた統合の判断のための指標は、

$$\begin{aligned}Sim[\{\alpha, \beta\}] \\ = (1 - Cos[\alpha, \beta]) * Scale[\{\alpha, \beta\}]\end{aligned}\quad (7-2)$$

となる。

一般論として、距離指標としてユークリッド距離を用いた場合、方向の等しいベクトルのノルムも評価対象となり、数量で表現されているが割合が対象の場合など、方向のみが意味をもつ対象では元データを正規化する必要がある。一方で、コサイン類似度を用いると、ノルムの違いが無視されるため正規化をする必要がなく、最大（最小値）が事前に把握できない場合など正規化が困難なケースにおいて利用されることが多い。

ただし、本稿における分析ではベクトルの要素が比率であるため、各要素が 0~1 の値を取り、ベクトルの和が常に 1 になるという制約を持つ。したがって、本稿の検証では、各ベクトルの分布が 4 次元空間上の 3 次元空間上に制約されており、同一の方向を持つベクトル間の距離は全て 0 となる。またその制約によりコサイン類似度と、ユークリッド距離の間での差異は非常に小さ

いものとなっているため、どちらを用いても結果に大きな差は生じない。

図 6, 7 はそれぞれユークリッド距離とコサイン距離を用いて計算した結果を比較した図であるが、いずれも大きく違うことが分かる。

2 つの指標の違いとしては、コサイン類似度を用いた場合、空間上の端点に位置する 2 つのベクトルのコサイン類似度が、非端点の組み合わせよりも大きくでる傾向があるという特徴がある。

例えば、ベクトル

$(0,50,50,0), (0,49,51,0)$ と

$(0,100,0,0), (0,99,1,0)$ のユークリッド距離は等しいが、

$$\cos[(0,50,50,0), (0,49,51,0)] = 0.9998$$

$$\cos[(0,100,0,0), (0,99,1,0)] = 0.9999$$

となり、コサイン距離では、 $(0,100,0,0), (0,99,1,0)$ がより類似度が高いこととなる。

図 6, 7 においては、この影響で、最終的な値は変わらないものの、幾つかの統合の順序の違いがグラフの増減の位置として現れていることが見て取れる。

図 5 は、ユークリッド距離とコサイン類似度の概念差を理解するための図であり、縦軸、横軸それぞれが配分比率 (0~100%) を示す。なお、本稿の分析で扱われるベクトルは 4 次元だが、簡単のために 2 次元ベクトルについて扱っている。

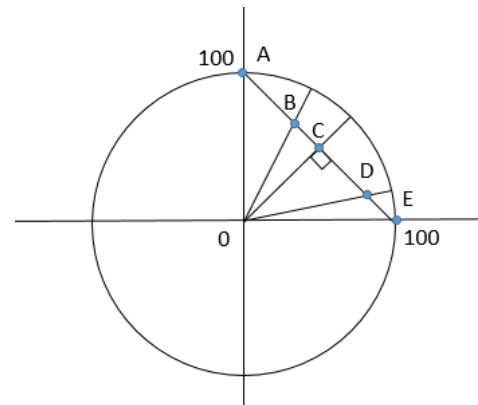


図 5 類似指標概念図

本稿では各ベクトルは第 1 象限において分布し、また、和が 1 になるという制約によって各点の分布は、数直線 AE 上に限定される。なお、3 次元では、各軸の端点を結んだ平面上、4 次元では 3 次元空間上に分布が限定されることとなる。

簡単のため、点 C を線分 AE に対して点 O から垂線を引いた交点とする。このとき、点 C の代表する品目は横軸、縦軸それぞれの項目に 50% ずつ配分されていることになる。この点 C と比較される、「点 C に比較的配分比率の近い点 B」、「100% 横軸に配分される点 E」、「点 E と比較的配分比率の近い点 D」の 4 点を考える。

このとき、ベクトル $\overrightarrow{OB}$, $\overrightarrow{OC}$ のユークリッド距離は $|BC|$ 、コサイン類似度は $\cos(\angle BOC)$ となり、 $\overrightarrow{OD}$, $\overrightarrow{OE}$ に関しても同様である。また、端点、即ちほぼ単一需要先に配分されるベクトルは点 A, E の近傍に分布することになるが、仮に $\angle DOE = \angle BOC$ のとき、即ち、コサイン距離が等しい場合、 $|BC| > |DE|$ 、つまり、端点近傍のユークリッド距離が中央と比較して小さくなることを以下に示す。

ここで、

$$|OA| = |OE| = 1,$$

$$\angle BOC = \angle DOE = \theta,$$

$$|BC| = x, |CD| = y, |DE| = z$$

とする。

このとき、 $|AE| = \sqrt{2}$ であることから、

$$x = \frac{1}{\sqrt{2}} \tan \theta \quad (7-3)$$

$$y = \frac{1}{\sqrt{2}} \tan\left(\frac{\pi}{4} - \theta\right) \quad (7-4)$$

$$z = \frac{1}{\sqrt{2}} - y = \frac{1}{\sqrt{2}} * \frac{2 \tan \theta}{1 + \tan \theta} \quad (7-5)$$

となる、

$$\begin{aligned} x - z &= \frac{1}{\sqrt{2}} \tan \theta - \frac{1}{\sqrt{2}} * \frac{2 \tan \theta}{1 + \tan \theta}, \\ &= \frac{\tan \theta (\tan \theta - 1)}{\sqrt{2} (1 + \tan \theta)} \end{aligned} \quad (7-6)$$

であるから、

$$\tan \theta < 1 \Leftrightarrow \theta < \frac{\pi}{4}$$

において、同じユークリッド距離を持つ2つのベクトルの組み合わせの内、端点近傍に位置する組み合わせのコサイン類似度が中央と比較して低くなるといえる。点Cが線分AEを等分するとき、 $\angle BOC$ が 45° を超えることはないため、点C、点Eにそれぞれ対応する点の組み合わせとしては、全ての場合でこれは成り立つ。点C及び、点Eを仮定しない場合、即ちそれぞれの点が50%

あるいは、100%の配分比率を持たない場合にもこの傾向は成り立つと考えられるが、その場合角度が非常に大きい場合において、これが逆転する可能性もある。ただし、角度がそのように大きくなる統合は、シミュレーションにおける後半にのみ集中することとなる。

これらのことから、コサイン類似度を採用することは、ユークリッド距離と比較して、端点近傍の項目を優遇することを意味する。これは、例えば、衛生用紙や、情報用紙など、中間投入に9割以上、家計消費に僅か配分される品目が、野菜や果物などのように家計消費が7割、中間投入が3割配分される品目よりも優先して統合されることを意味している。用紙の例のように全体のわずかな部分が他の配分先に需要される品目の方が、配分比率が安定している可能性はあるが、一般に端点近傍の項目を優先して統合することが望ましいというコンセンサスは存在しない。

本稿の分析においては、上述の通り、いずれを用いても結果に大きな差は生まれないが、手法が一般的かつ、結果に偏りが無いことから、ユークリッド距離を類似度指標として用いた。

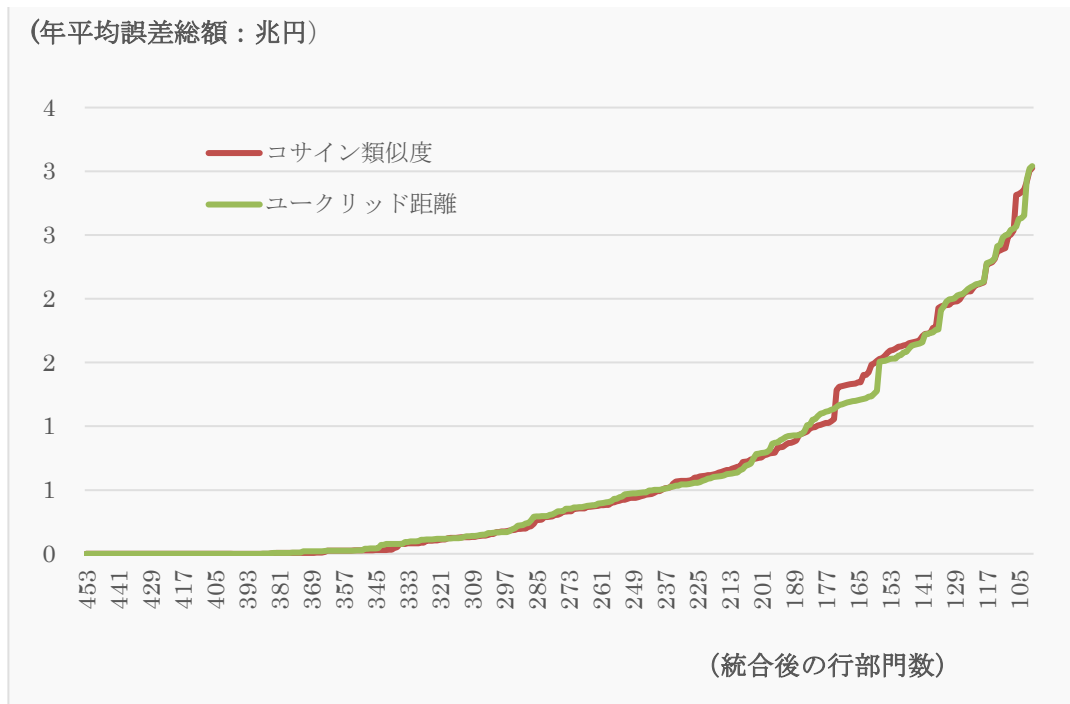


図 6 コサイン距離，ユークリッド距離の比較（年平均誤差総額）

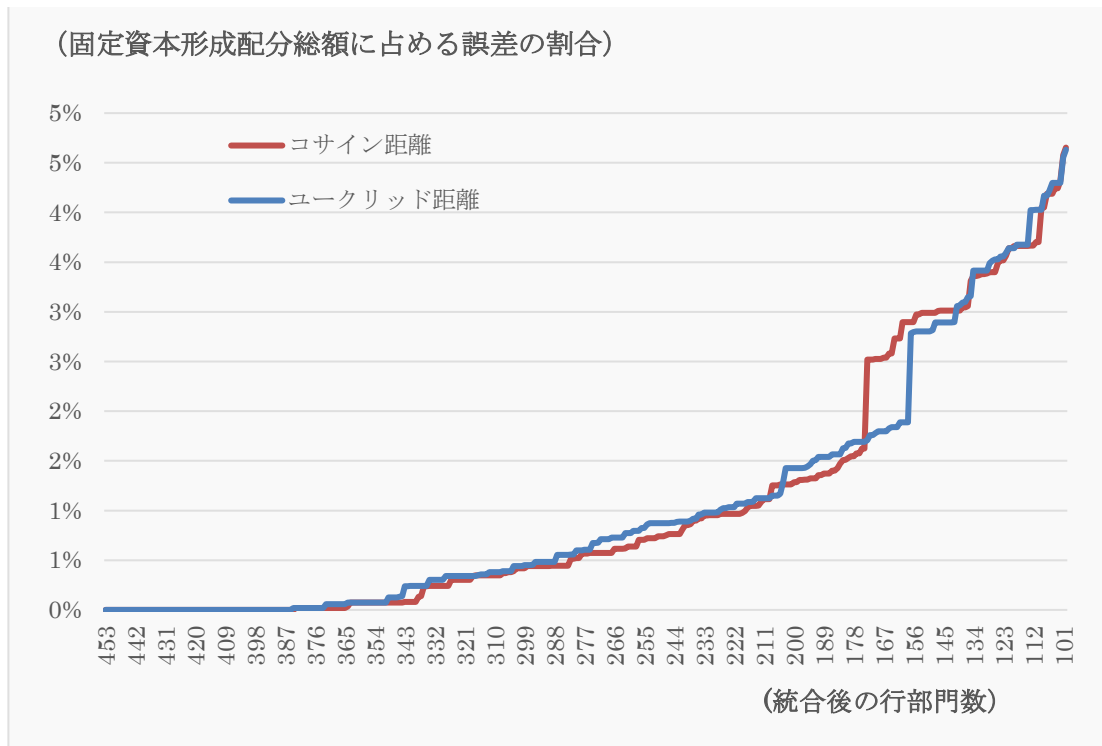


図 7 ユークリッド距離，コサイン類似度の比較（固定資本形成誤差割合）