

産業・市区町村別の域内総生産額の推定

菊池 悠斗¹・高山 雄貴²・村上 大輔³

¹学生非会員 金沢大学 理工研究域 (〒 920-1192 石川県金沢市角間町)
E-mail: yuto0416@stu.kanazawa-u.ac.jp

²正会員 金沢大学 理工研究域 (〒 920-1192 石川県金沢市角間町)
E-mail: ytakayama@se.kanazawa-u.ac.jp

³正会員 統計数理研究所 データ科学研究系 (〒 190-8562 東京都立川市緑町)
E-mail: dmuraka@ism.ac.jp

本研究では、市区町村産業連関表の作成の根幹を成す市区町村別生産額の推計を行う。推計手法として、複数の学習結果のアンサンブル学習を使用し、各市区町村の空間データや市町村間の空間相関に考慮した学習により市区町村別生産額を推計する。アンサンブル学習の重み決定には 4 種類の手法を使用し、Non-Survey 手法(石川, 小池, 上田 (2001)¹⁾)による推計結果との比較検証を行う。

Key Words : *Input Output Table, spatial statistics, spatial autocorrelation, kriging*

1. はじめに

産業連関表は、国や都道府県、市区町村ごとに作成され、産業連関分析に用いられる統計資料である。1 年間に行われた地域の財や産業の産業間取引を表し、地域にもたらされる経済効果を分析することが可能となる²⁾。近年、日本の多くの自治体では、財政状況が厳しく、限られた財政を有効的に活用していく必要があり、そのために産業連関表は益々必要となる統計資料である。しかし、産業連関表は既存の統計調査 (e.g. 国勢調査, 経済センサス) をもとにして作成される二次統計であり、産業連関表の作成には人手やノウハウが必要となる。そのため、財政が厳しく、かつ作成に関するノウハウを有する人材のいない市区町村では、現状として産業連関表の作成が進まず³⁾、作成されているものの多くは政令指定都市をはじめとした大都市に限られている。

そこで、統計モデルを用いた細分化により、都道府県産業連関表から市区町村産業連関表の作成を行う。本研究では、その第一歩として、産業連関表作成における根幹を成す産業別生産額の推計を行う。細分化手法は、複数の学習結果に基づくアンサンブル学習を使用する。複数の学習結果に基づくアンサンブル学習を行うことで、1 つの学習結果に偏った推計を避け、推計精度の向上に繋がるとされている。また、加重平均はアンサンブル学習の基礎的な手法であり、個々の学習結果が有する性質が異なる場合、高精度の学習が可能であるとされている (Zhou, 2008⁴⁾) ため、本研究では加重平均によるアンサンブル学習を用いる。また、アンサンブ

ル学習の学習方法の 1 つとして、空間統計学の一般的な手法であるクリギングを使用し、空間相関を考慮している。産業の発達には、地域内環境だけではなく、周辺の地域環境 (e.g. 人口, 自然環境, 埋蔵資源) の影響を受ける。そのため、空間相関を考慮することで推計の精度向上が見込まれる。

本稿の構成は、以下の通りである。第 2 章では、本研究で用いる産業別生産額の推計手法を説明する。続く第 3 章では、推計手法を用いた具体的な産業別生産額の推計結果について説明する。その上で、既往研究である Non-Survey 手法 (石川, 小池, 上田 (2001)) による細分化結果との比較を行い、本研究で得られた結果の考察を行う。第 4 章では、本稿の総括を行う。

2. 産業別生産額推計手法

本研究では、都道府県産業連関表における産業別生産額の細分化方法として、複数の学習結果の加重平均によるアンサンブル学習を行う。産業別生産額の推計手順は、次の通りである：

1. 被説明変数 (都道府県値), 説明変数の設定
2. 説明変数による被説明変数 (都道府県値) の按分
3. 説明変数の選定及び重みの決定
4. 加重平均による市町村産業別生産額の推計

以上の手順により、都道府県別生産額 Y から同都道府県内の i 番目の市区町村別生産額 y_i を推定する。具体的には、手順 1 では K 個の説明変数を設定する。手順 2 では各説明変数を按分比とする K 個の按分モデルを考える。 k 個目の按分モデルは、下式で市区町村 i にお

ける生産額を推定する：

$$\hat{y}_{i,k} = \frac{g_{i,k}}{\sum_{i=1}^N g_{i,k}} Y \quad (1)$$

(1) 式は市区町村毎の説明変数 $g_{i,k}$ に応じた比例配分により市区町村別の推定値 $\hat{y}_{i,k}$ を求める。手順 4 では上で得られる各推定値の加重平均により最終的な推定値 \hat{y}_i を求める：

$$\hat{y}_i = \sum_{k=1}^K w_k \hat{y}_{i,k} \quad (2)$$

重み w_k は手順 3 であらかじめ決めるものとする。以下では、各ステップについて説明する。

(1) 被説明変数、説明変数となる空間データの設定

はじめに被説明変数及び説明変数となる空間データを設定する。

被説明変数には、2015 年に作成された各都道府県産業連関表より産業別生産額を設定する。

説明変数には、統計調査 (e.g. 国勢調査, 経済センサス) の空間データを使用する。なお、説明変数として用いるデータは、都道府県データの市区町村データへの按分に用いるため、全市区町村で集計されているデータを使用する。

(2) 説明変数による被説明変数 (都道府県値) の按分

次に、説明変数を用いて被説明変数となる都道府県産業連関表の産業別生産額を按分する。

本手法では、説明変数の比例配分 (例えば, Eicher and Brewer, 2001⁵⁾) によって、都道府県産業連関表の産業別生産額を市区町村単位の値へと細分化を行う。なお、比例配分に用いる説明変数は、以下の 2 種類に分類される。

1. 説明変数として設定した空間データによる比例配分
2. area-to-point kriging (例えば, Yoo and Kyriakidis, 2006⁶⁾) による比例配分

後者では、都道府県別生産額 Y の背後に下式を仮定した上で、市区町村別生産額 y_i の期待 2 乗誤差 $E[(y_i - \hat{y}_i)^2]$ を最小化することで重み $g_{i,k}$ (ひいては予測量 $\hat{y}_{i,k} = \frac{g_{i,k}}{\sum_{i=1}^N g_{i,k}} Y$) を求める：

$$Y = \sum_{i=1}^N y_i \quad (3)$$

$$y_i = \mu + s_i + \varepsilon_i$$

$$s_i \sim N(0, \tau^2 c(d_{i,j}; r))$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

上式は、市区町村別生産額の和が都道府県別の生産額に一致すること、および市区町村別生産額 y_i が [平均 μ] + [空間相関パターン s_i (分散 : τ^2)] + [ノイズ ε_i (分散 : σ^2)] で説明されることを表す。各要因の影響の強さは

分散パラメータ (τ^2, σ^2) を推定することで調整される。また「距離が近いほど観測値が類似する」という空間相関を捉えるために、 s_i の共分散は市区町村の重心点間の直線距離 $d_{i,j}$ の指数減衰関数 $c(d_{i,j}; r) = \exp(-d_{i,j}/r)$ で与えた。 r は空間相関の及ぶ距離を調整するパラメータである。上式は最尤法で推定する。

(3) 説明変数の選定及び重みの決定

3 段階目では、説明変数を選定し、それぞれの説明変数に対する重みを決定する。本研究では、3 種類の方法により、説明変数を選定、重みの決定を行う。本研究で用いる説明変数及び重みの決定方法は、相関係数による決定、LASSO 回帰、非負条件付き最小二乗法 (非負条件付き Ridge 回帰) による決定である。以下では、それぞれの決定方法について説明する。

a) 相関係数による決定

1 つ目の方法として、相関係数によって重みを決定する。説明変数 i の重み w_i は相関係数 ρ_i を用いて、以下の (4) 式により行う：

$$w_k = \frac{\rho_k^2}{\sum_{k=1}^K \rho_k^2} \quad (4)$$

しかし、相関係数によって重みを決定する場合、必ずしも全てのデータが相関性の高いものであるとは限らないため、推計の精度が低下する可能性がある。そこで、以下の 2 種類の方法により、強い相関性を有するデータのみを説明変数として決定する。

1. 相関係数の閾値を設定する場合

1 つ目の方法として、説明変数の閾値を設定する場合である。相関係数の閾値を設定することで、相関性の高い空間データのみを説明変数とする。例えば、閾値を 0.90 としたとき、相関係数が 0.90 以上のデータを説明変数とする。また、相関係数が 0.90 以上となるデータが含まれない場合は、全てのデータを説明変数として加重平均を行い、市区町村データへ細分化する。

2. 相関係数のランクを設定する場合

2 つ目の方法として、説明変数として設定するランクを設定する場合である。例えば、説明変数のランクを 3 としたとき、相関係数の大きな値を有するデータから順番に 3 種類を説明変数として設定する。相関係数の閾値により設定した場合、閾値を満たすデータが含まれていない可能性があった。一方で説明変数のランクにより設定する場合、データの中で相対的に相関係数が大きな変数を設定することができる。

b) LASSO 回帰

2 つ目の方法として、LASSO 回帰によって説明変数の重みを決定する。LASSO 回帰により推計された説明

変数 k のパラメータ β_k を用いて、重み w_k は次の (5) 式により算出される：

$$w_k = \frac{\beta_k}{\sum_{k=1}^N \beta_k} \quad (5)$$

c) 非負条件付き最小二乗法 (非負条件付き Ridge 回帰)

3つ目の方法として、非負条件付き最小二乗法を用いた重みの決定を行う。最小二乗法により得られた各説明変数 k に対するパラメータ β_k を用いて、LASSO 回帰と同様に (5) 式により各説明変数に対する重み w_i が得られる。しかし、最小二乗法の場合、LASSO 回帰とは異なり、全ての説明変数を用いた回帰を行う。そのため、説明変数として複数の空間データを投入していることから、多重共線性により過学習となる可能性が高い。そこで、Ridge パラメータ λ を設定することで、非負条件付き Ridge 回帰として正則化を行い、パラメータ β_k の算出を行っている。

(4) 加重平均による市町村産業別生産額の推計

最終段階として、以上で算出してきた各説明変数による按分結果 $\hat{y}_{i,k}$ 及び重み w_k を用いることで、(2) 式により産業別生産額 y を算出する。

なお、本手法では、細分化手法の汎化性能を向上するために、クロスバリデーションによる産業別生産額を行う。具体的な手順は以下の通りである。なお、以下の手順ではグループ 1 の推計を行う手順を示している。

1. 都道府県をグループ 1~4 に分類する。本研究では、都道府県コード順に各グループ 11 又は 12 都道府県ずつになるように設定している。
2. グループ 2,3,4 の都道府県データを市区町村データに細分化する。
3. グループ 2,3,4 の細分化結果から得られた各説明変数に対する重み w_i を出力する。
4. 手順 3 で得られた各説明変数に対する重み w_i を用いて、グループ 1 の都道府県データを市区町村データに細分化する。
5. 手順 4 で行ったグループ 1 の細分化結果を最終的なグループ 1 の市区町村データとする。
6. 手順 2 に戻り、残るグループ 2,3,4 の市区町村データの推計を実施し、完了する。

3. 産業別生産額の推計

本章では、実データを用いて、第 2 章で説明した推計手法を実施した結果を示すとともに、この結果と Non-Survey 手法による推計結果との比較を行う。

(1) 投入データ

本研究では、説明変数として投入する空間データを表 1 を設定した。また、細分化の対象となる都道府県産業連関表は 2015 年に作成されたもの (奈良県では、2015 年産業連関表が公開されていないため、2011 年に作成されたもの) を使用した。なお、対象とする産業分類は表 2 の 13 部門分類を使用する。都道府県によって、わずかに産業区分が異なるため、予め同一の 13 部門分類に統一する必要がある。また、説明変数として投入する「従業員数」、「事業所数」及び「付加価値額」については各産業に対応するデータを使用する。なお、データが国勢調査区分での提供となっているため、表 2 に従って、各部門の説明変数とする。

表-1 投入空間データ

データ	出典
産業連関表	2015 年都道府県産業連関表
就業者数	2015 年国勢調査
事業所数 付加価値額	2016 年経済センサス -基礎調査-
総人口	2015 年国勢調査
製造品出荷額等	2015 年工業統計調査 2016 年経済センサス -活動調査-
商業年間商品販売額 (卸売業+小売業)	

表-2 産業部門分類

産業連関表 産業区分	国勢調査大分類
01 農林漁業	A 農業, 林業, B 漁業
02 鉱業	C 鉱業, 採石業, 砂利採取業
03 製造業	E 製造業
04 建設	D 建設業
05 電力・ガス・水道	F 電力・ガス・熱供給・水道業
06 商業	I 卸売業, 小売
07 金融・保険	J 金融業, 保険業
08 不動産	K 不動産業, 物品賃貸業
09 運輸・郵便	H 運輸業, 郵便業
10 情報通信	G 情報通信業
11 公務	S 公務 (他に分類されるものを除く)
12 サービス	L 学術研究, 専門・技術サービス業, M 宿泊業, 飲食サービス業, N 生活関連サービス業, 娯楽業, O 教育, 学習支援業, P 医療, 福祉, Q 複合サービス事業, R サービス業 (他に分類されないもの)
13 分類不明	T 分類不能の産業

(2) 推計精度の評価方法

推計によって得られた結果の精度評価方法について説明する。推計結果の精度評価は、実際に作成されている市区町村の産業連関表における実績値との比較により行う。比較対象となる市町は表 3 の通りである。なお、2015 年作成の産業連関表のみの場合、市区町村の

規模が政令指定都市のような大都市となり、かつサンプル数が少なくなる。そのため、本研究では、2015 年前後 4 年以内に作成された市区町村の産業連関表を使用することで、比較対象市区町村のサンプル数を増やしている。また、精度の評価方法は、%RMSE により、実績生産額と推計生産額の間が生じた誤差の大きさにより評価を行う。なお、 $x_{i,k}$ は実績市区町村別生産額、 n は比較対象市区町村数 (今回は 28)、 \bar{x} は、実績市区町村別生産額の平均値としたときに、%RMSE は下式で求める：

$$RMSE = \sqrt{\frac{(x_{i,k} - \hat{y}_{i,k})^2}{n}}$$

$$\%RMSE = \frac{RMSE}{\bar{x}} \times 100$$

表-3 市区町村産業連関表 比較対象

作成年次	市区町村
2017	兵庫県豊岡市
2016	岐阜県高山市
2015	北海道札幌市, 北海道釧路市, 埼玉県さいたま市, 神奈川県横浜市, 神奈川県川崎市, 神奈川県相模原市, 岐阜県郡上市, 静岡県静岡市, 愛知県名古屋, 京都府京都市, 兵庫県神戸市, 岡山県岡山市, 広島県広島市, 福岡県福岡市, 熊本県熊本市, 沖縄県那覇市
2014	長野県塩尻市, 京都府宮津市, 岡山県笠岡市
2012	山口県周南市, 愛媛県松山市, 愛媛県新居浜市
2011	北海道別海町, 千葉県千葉市, 静岡県三島市, 大阪府大阪市

(3) 産業別生産額の推計結果

第 2 章で説明した相関係数による決定 (閾値による決定, ランクによる決定), 非負条件付き最小二乗法 (非負条件付き Ridge 回帰) による決定, 及び LASSO 回帰による決定の全 4 パターンの推計を行った。なお、今回は相関係数の閾値を 0.90, ランクを 3 として相関性の高いデータに着目して推計を行った。また、非負条件付き最小二乗法では、Ridge パラメータ $\lambda = 0.1$ とし、非負条件付き Ridge 回帰として推計を行った。さらに、比較対象として Non-Survey 手法による推計 (「就業者数」による都道府県別生産額の按分による推計) も行った。全 5 通りの推計結果は、表 4 に示す。

得られた結果を考察する。まず、相関係数による説明変数と重みを決定した場合に得られた結果を述べる。1 つ目は、閾値を 0.90 としたときに生じた誤差が 7 部門において、他手法に比べて小さくなっており、精度の良い推計が行われている点である。この要因として、まず相関性の高いデータのみを説明変数として設定し、推計を行うことができたことが挙げられる。例えば、今回の説明変数の中に、製造業に特化した「製造品出荷額

表-4 推計結果 (手法別%RMSE)

産業 (13 部門)	相関係数		非負条件付き 最小二乗法
	閾値 (0.90)	ランク (= 3)	
農林漁業	49.816	50.682	45.198
鉱業	150.799	78.302	106.184
製造業	13.543	37.180	141.487
建設	44.062	48.045	34.439
電力・ガス・水道	62.957	77.437	80.308
商業	64.601	54.957	68.062
金融・保険	19.860	23.702	29.636
不動産	19.167	21.652	24.214
運輸・郵便	24.619	26.457	64.025
情報通信	28.805	32.049	35.895
公務	34.421	42.645	63.853
サービス	17.356	26.737	24.810
分類不明	25.201	31.616	82.537

産業	LASSO 回帰	Non-Survey
農林漁業	45.198	44.550
鉱業	93.943	175.205
製造業	151.569	75.305
建設	42.154	35.987
電力・ガス・水道	80.721	93.739
商業	39.872	65.710
金融・保険	31.331	19.159
不動産	32.479	34.438
運輸・郵便	34.104	34.792
情報通信	36.329	29.636
公務	57.001	27.300
サービス	21.930	21.734
分類不明	137.264	38.53

等」を設定しており、相関性の高いデータのみに着目したことで、他の説明変数の影響を排除することができ、高精度の結果に繋がったと考えられる。2 つ目は、第三次産業を中心に高精度に推定を行うことが出来ている点である。これは、説明変数として、「就業者数」や「事業所数」、「総人口」といった人口ベースのデータを投入しているためであると考えられる。第一次産業では農林漁業のような各地域の自然資源に依存し、第二次産業ではモノが中心となった産業である。一方で、第三次産業では接客のようなサービス業、高度な技術を有するヒトが資本となる産業である。そのため、人口ベースのデータを多く設定しているため、第三次産業を中心に精度が高くなっていると考えられる。また、今後、自然資源等に関するデータを説明変数として設定することで他の産業も精度が向上していくと予想できる。

Non-Survey 手法と比較したとき、多くの産業で閾値及びランクのいずれにおいても誤差が小さくなっている。また、Non-Survey 手法が高精度である場合においても、精度に大きな差がないため、相関係数により説明変数と重みを決定する方法は、産業別生産額の推計に効果的であることがわかる。

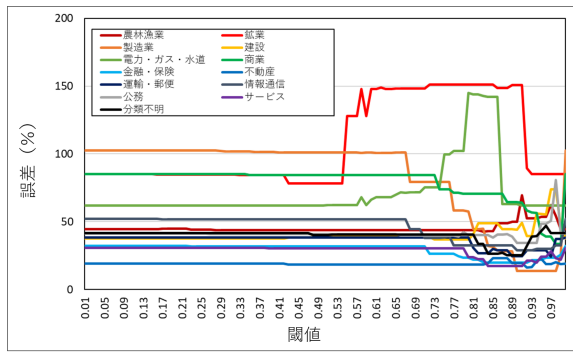


図-1 閾値を 0.01 間隔で変えたときの推計誤差

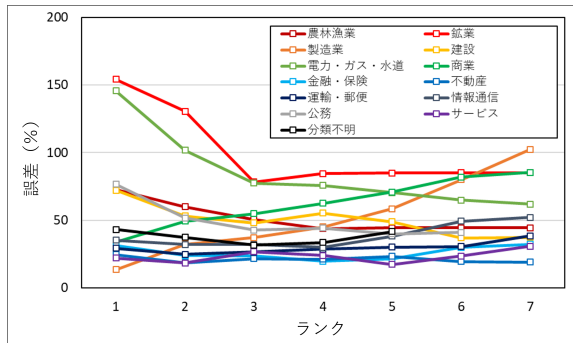


図-2 ランクを 1 間隔で変えたときの推計誤差

一方、相関係数による方法の課題として、閾値とランクをどのように設定するべきか、が挙げられる。図1と図2はいずれも閾値を0.01ずつ、ランクを1ずつ変えたときの誤差の挙動を示したものである。この2つの図からわかるように、産業により誤差の挙動は大きく異なっており、基準値を変えることで大きく精度が変化する。

次に、非負条件付き Ridge 回帰及び LASSO 回帰を用いた場合の結果を述べる。まず、相関係数による方法とは異なり、製造業における誤差が極めて大きな値となっており、いずれも約 150% となっている。製造業は相関係数の高いものみに着目したときに精度が高くなったと説明した。図1および図2に着目したとき、いずれも閾値とランクが大きくなるとともに誤差が大きく拡大している。すなわち、製造業の場合、1つの説明変数(今回は「製造品出荷額等」)のみに着目したときに精度が高くなっている一方、非負条件付き最小二乗法と LASSO 回帰では相関係数の大小を考慮していないため、今回誤差が拡大したと考えられる。また、相関係数の場合及び Non-Survey 手法のときに精度の高い第三次産業においても、誤差が大きくなっている。

以上の結果から、説明変数の相関係数に基づいて推計を行うことが効果的であり、また、「就業者数」だけでなく、複数のデータを用いることで推計精度が高まることが明らかになった。一方、課題として、上述した閾値やランクの決定をどのように行うかが課題として

挙げられる。さらに、今回比較対象の市区町村として 28 の産業連関表を用いてきたが、その多くが大都市を中心としたものである。そのため、大都市が有する産業構造に偏った精度検証となっており、特異的な産業構造を有する市区町村での推計精度が不明であり、全市区町村の推計精度をどのように検証していくかがもう 1 つの課題として挙げられる。

4. おわりに

本研究では、複数の学習結果に基づくアンサンブル学習により、空間相関を考慮した市区町村産業連関表における産業別生産額の推計を行った。推計手法では、アンサンブル学習の基礎的手法である加重平均を使用し、加重平均に用いる重みを相関係数による方法、非負条件付き最小二乗法による方法、LASSO 回帰による方法を用いた。そして、既往研究である Non-Survey 手法に基づいた産業別生産額の推計結果との比較を行った。その結果、相関係数による説明変数と重みの決定を行った場合、Non-Survey 手法よりも誤差の少ない推計結果が得られた。また、第三次産業を中心に高精度の推計を行うことが可能となり、より複数のデータを説明変数の候補として細分化を行った場合において、さらに産業別生産額及び産業連関表全体の作成精度をさらに向上することが期待できる。

しかし、産業別生産額の推計において、いくつか課題が見受けられた。具体的には、相関係数により説明変数を決定する場合、どのように閾値やランクを決めていくべきかという課題が明らかになった。また、誤差の推計において、実際の市区町村産業連関表の実績値との比較を行ったが、比較対象の多くが大都市となっているため、規模の小さな自治体や特徴的な産業構造を有する地域の推計精度をどのように計るかが課題となる。以上の課題は、今後すべての市区町村産業連関表を作成していくうえで、重要な研究課題となる。

参考文献

- 1) 石川 良文, 小池 淳司, 上田 孝行: Non-Survey 手法による都市圏産業連関表の作成, 土木学会第 56 回年次学術講演会, 2001.
- 2) 土居 英二, 浅利 一郎, 中野 親徳: はじめよう地域産業連関分析 Execl で初歩から実践まで [改訂版] 基礎編, 日本評論社, 2019
- 3) 石川良文: 日本の地域産業連関表作成の現状と課題, 産業連関, Vol.23, No.1-2, pp.3-17, 2016
- 4) Zhou.Z.-H: Ensemble Learning, *Machine Learning*, pp.101-128, 2008
- 5) Eicher, C. L., and Brewer, C. A.: Dasymetric mapping and areal interpolation: Implementation and evaluation, *Cartography and Geographic Information Science*, Vol.28, No.2, pp125-138, 2001

- 6) Yoo, E. H., and Kyriakidis, P. C.: Area-to-point kriging with inequality-type data, *Journal of geographical systems*, Vol.8, No.4, pp.357-390, 2006

(2023. 3. 6 受付)