

標本調査データからの小地域情報の抽出可能性

— 都道府県別業況DIの推定をめぐる検証 —

坂田幸繁*

要旨

標本調査データからの小地域推定の方法は、標本調査法の論理に即したデザインベースの推定法から、モデルベースのアプローチへとその重点を移行しつつある。そのような方法転換を統計的に評価するために、小企業を対象に実施された景況調査の個票データを仮想母集団として、そこからの抽出標本による推定実験を行った。その結果、モデルベースの推定量が平均平方誤差の観点では有効であること、その効果は補助情報に大きく依存することなどを確認した。

キーワード

小地域推定, 標本調査, リサンプリング, 地域業況DI, 変量効果モデル

1. はじめに — 課題設定 —

完全失業率の都道府県別推定値, 一人当たり平均所得の市区町村別推定値, 貧困世帯の就学児童数の学区別推定値など, これらの数字の社会的意義に多くを語る必要はないであろう。最初のもは日本, 他の2つは政策プログラムにリンクした米国の例をイメージしている¹⁾。このようなタイプの数字が任意地域, 任意時点に対して獲得できるとすれば, その情報価値は極めて高い。

しかし今日の統計作成体系を前提とすれば, このような推定値の獲得には標本調査統計のデータリソースに大きく頼らざるを得ない²⁾。いま少し問題を限定して, 標本調査の調査時点は所与としたときの任意地域の推定問題に焦点を当てることにしよう。その場合でも, 確率抽出標本を通例とすれば, 目標地域に対

して要求精度を満たす十分な標本サイズが確保されている保証はない。極端な場合, ターゲットとする該当地域に属す要素が標本に含まれていない。いわゆる小地域推定の問題に逢着する³⁾。本稿では, 推定目標となる変数を含む標本調査データが個票レベルで利用できるという想定のもとに, 小地域推定の方法を実際の景況調査の個票データを素材に評価することにしたい。もちろん, 当該調査データの他に利用可能な補助的情報があれば, 推定に積極的に利用すべきである。このような点にも留意しつつ, 小地域推定の方法展開を標本調査データからの地域情報の抽出可能性という視点から吟味することにしたい。

具体的には, 推定精度が直接には担保されない地域に対する業況DI (Diffusion Index) の推定問題を例証として, 小地域推定がどのように可能であるのか, その条件は何であるのか, 抽出実験によるシミュレートから検証することにしたい。そのために次節では, い

* 中央大学経済学部

〒192-0393 東京都八王子市東中野742-1

わゆる小地域推定の方法論の要点を整理した上で、3節で使用データとその評価方法について説明し、4節以降で推定法とともにそのパフォーマンスの結果概要を示しながら、上記課題にアプローチしていくことにする。実際のデータにはさまざまな制約（すべて無作為標本というわけではない、あるいは非協力、無回答、記入誤差の存在など）があるが、論旨を明確にするために、ここでは無作為抽出標本という面に焦点を絞って検討を加える。

2. 標本調査からの地域母数の推定

2.1 直接推定 (direct estimation)

対象とする地域 i に属する標本サイズがある程度大きいならば、通常確率標本からの推定図式が利用できる。地域といっても部分母集団の推定と同じである⁴⁾。いま変数 y (例えば所得や売上など) についての地域合計を例に推定問題を考えよう⁵⁾。

調査母集団を U 、母集団の大きさを N 、それを構成する要素を $j(j=1, 2, \dots, N)$ 、母集団のある地域、すなわち関心のある部分母集団を U_i とおけば、 j 番目の要素がもつ変数 y の値 $y_j(j=1, \dots, N)$ を用いて、地域合計は $Y_i = \sum_{j \in U_i} y_{ij} = \sum_{j \in U_i} y_j$ と書ける。ただし y_{ij} は、 j が当該地域に属す母集団要素であれば y_j 、そうでなければ0の値をとるように定義しておく。いま母集団 U に対して抽出率 f の無作為抽出標本 s (サイズ n) が得られており、そのうち当該地域に属する標本 (地域標本と呼ぶ) を s_i と表すことにしよう。他に利用できる補助情報 (企業数や世帯数、資本金や就業者数など) が何もなければ、地域標本を抽出ウェイト w_j (抽出率の逆数) で膨らませて地域合計の推定値 $\hat{Y}_i = \sum_{j \in s_i} w_j y_{ij} = \sum_{j \in s_i} w_j y_j$ とするしかない⁶⁾。そしてその分散は母集団に関する合計推定の場合と同じく、 $y_{ij}^* = w_j y_{ij}$ に関する分散 $s_{y_i}^2$ に有限母集団修正項を付加して $v(\hat{Y}_i) = (1-f)ns_{y_i}^2$ とすればよい。

補助情報として、例えば地域のサイズ (部

分母集団の大きさ) N_i が既知であれば、その地域の母平均推定量としての標本平均に地域サイズを掛ける方法がある。この場合、補助情報がなく標本だけで上記のように推定する方法より推定誤差は減少する⁷⁾。より一般的に、補助情報があれば比推定や回帰推定など、不偏性や一致性をある程度維持しながら、補助情報による制約 (正しい関係情報) の分だけ推定誤差を減少できる可能性が高まる。補助情報を利用したときの推定量を \hat{Y}_i^* とおけば、この場合も地域標本だけを用いる線形推定量として、 $\hat{Y}_i^* = \sum_{j \in s_i} w_{ij}^* y_{ij} = \sum_{j \in s_i} w_{ij}^* y_j$ と表すことができる。地域標本か否かで再定義されたデータ y_{ij} に対して、利用可能な補助情報を制約条件とする修正ウェイト w_{ij}^* で膨らませるだけである。補助情報の有無にかかわらず、地域標本サイズがその精度を規制することにより変わりはなく、あたかも部分母集団を標本調査の対象としたかのように地域標本だけを利用した推定法である。いわば標本調査の論理に忠実な標準的な推定方式であり、これを直接推定と呼ぶ。

2.2 地域以外の標本も利用した推計

— 間接推定 (indirect estimation) —

サンプルの追加といった調査法の問題は別として、与えられた標本の下で、かつ補助情報を利用して推定精度が実用的でない状況であれば、当該地域に属さない標本の助けを借りる方法が考えられる。このようなアプローチを間接推定と呼ぶ。

ターゲットとなる地域に属していなくとも、その近辺にある標本データは推定に利用できるかもしれない。実質的な考え方は同じであるが、推定目標とした地域とその周辺を含むエリアや類似した地点集合とは y の変動に関して違いがないと仮定すればよい。このような考え方に基づく間接推定の単純な形態として合成推定量 (synthetic estimator) が構成される。 $\hat{Y}_i = \sum_{j \in s_i} w_j y_j$ と表すことができ、 $j \notin s_i$

に対して $w_j > 0$ となる要素 j が少なくとも一つ存在することが特徴である。実際の地域推計によくみられるように、データが入手できない項目について当該地域平均(部分母集団)を全国平均(母集団)と同じと仮定する処理法はこれに該当する⁸⁾。

仮定がある程度成立していると信じることにすれば、当該地域に対する推定量の分散は形式的な標本サイズの増加により低下する。さらに補助情報が利用できれば、仮定の下で分散はさらに小さくできるかもしれない。しかしながら、そのような仮定の妥当性が保証されているわけではなく、そこから生じるバイアスは不可避である。当然、真値との距離尺度である平均平方誤差(MSE)を大きくする危険性がある。そこで、先の直接推定量をこのような合成推定量で調整し、具体的には両者の加重平均として推定量を定める方法が考えられる。

一般的に直接推定量を $\hat{\theta}_i$ 、合成推定量を $\tilde{\theta}_i$ とおけば、その加重平均 $\phi_i \hat{\theta}_i + (1 - \phi_i) \tilde{\theta}_i$ を推定量として採用すればよい。これを複合推定量(composite estimator)と呼ぶ。ウェイト ϕ_i はMSEを最小化するように決めてやればよい。さらにはこのようなタイプの推定量の延長上にJames-Stein推定量を位置づけることができる⁹⁾。MSEを評価関数とすると、それは直接推定量より優れており、また合成推定量(あるいはその代替量)が真の値に近付くときMSEを最小化することが知られている。さらに加重平均式が示唆するように、これは経験最良線形不偏予測(EBLUP)モデルや経験バイズアプローチとも密接に関係する¹⁰⁾。

2.3 モデルベースの推定へ

標本設計に忠実な直接推定では推定量の分散が大きく、(暗黙の)仮定に基づく間接推定ではバイアスが生じる。そこで加重平均をとることで両者をバランスさせMSEを低下

させる。このような方法展開は、本来母集団分布について仮定をおかず関心対象の母数を推定するという標本調査の枠組みの下では、自ずと限界がある。とくに間接推定量については地域母数の変動(差異)について複雑な関係を導入できない。

James-Stein推定量はこのようなアプローチの限界点に位置するが、むしろそれが提起するのは、地域母数に関する仮説を明示的にモデルとして表現することでより良い推定量が得られる可能性である。そうであれば、例えば下記のように、地域データ y_{ij} を補助変数による回帰効果 $\beta_0 + \beta_1 x_{ij}$ 、地域別変量効果 τ_i 、および個体変動 ε_{ij} の線形和として考える素朴な混合効果モデルの利用も、小地域推定の方法として自然に受け入れることができよう。モデルの良さがMSEなどの推定値により適切に評価できれば、多様なモデルを比較考量した柔軟なアプローチが可能となる。

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \tau_i + \varepsilon_{ij}$$

$$\tau_i \sim N(0, \sigma_\tau^2), \varepsilon_{ij} \sim N(0, \sigma^2)$$

標本調査データによる地域母数の推定図式について、このようなモデルベースへの転換(調査データはモデルからの実現値)の正当性が認められるか否か、これが小地域推定の方法論的な分水嶺となる。

3. データと評価方法

3.1 使用データとその特性

検討のためのデータリソースとして、「全国小企業動向調査」(日本政策金融公庫総合研究所、調査時は国民生活金融公庫)の2004年7-9月調査(以下、本調査)の個票データを利用する。SSJデータアーカイブ(Social Science Japan Data Archive)¹¹⁾からの提供を受けたものである。この調査個票データを母集団と想定し、そこからのリサンプリング・データによって都道府県別DI値を様々な方法で推定し、評価しようという試みである。

本調査は、いわゆる景況調査の一種であり、国民生活金融公庫取引先（当時）の従業者30人未満の全国の小企業を対象に四半期毎に実施される。全国事業所の業種別構成比等を参考に公庫取引先企業から下記のようにサンプルが選定・調査されている¹²⁾。

標本数：10,617企業，有効回答数6,979企業
業（回答率65.7%）

対象地域：沖縄県を除く全国
（都道府県別変数を提供）

抽出：総務省「事業所・企業統計調査」における事業所数，国民生活金融公庫の年間貸出件数，実績等を勘案の上，総合研究所がサンプル数を決定し，公庫支店がサンプル選定を行う。

業種，取扱品目，従業者数をフェース項目として，売上高，受注額，採算水準，資金繰りなど，景況調査特有の質問項目がアンケート形式で調べられる。この中で本稿の関心対象は，業況判断（当期実績，来期見通し）項目である。その結果から業況判断DIが作成され，それは景気指標のひとつとして実勢把握や予測に多用される。2004年7-9月調査の質問文と結果（%）は次のようであった。長期停滞から若干回復傾向がみられた時期であり，選択肢間のばらつきがそれまでに比べ増加しており，都道府県間DIの格差も相対的に拡大している時期のデータである。

[当期実績] 7月から9月のあなたの企業の業況はどうか。(回答結果%)

1) かなり良い 2.0 2) やや良い 16.3
3) 良くも悪くもない 26.4 4) やや悪い 35.2
5) かなり悪い 20.0

[来期見通し] 10月から12月のあなたの企業の業況はどうなるでしょうか。(回答結果%)

1) かなり良い 1.0 2) やや良い 17.0
3) 良くも悪くもない 32.2 4) やや悪い 33.2
5) かなり悪い 16.6

3.2 推定法の評価方法

推定法評価のために，次のような手順での推定実験を反復実行した。

- ① 上記調査のデータセットを母集団（すなわち真値は既知）として，都道府県を層化変数に抽出率 $f=1/5$ の非復元層化単純無作為抽出で推定用のサンプル(再抽出標本)を作成する。
- ② 再抽出標本に対して，特定の推定法（後述）を適用し，来期見通しに関する都道府県別DI値を推定する。
- ③ 条件（利用可能な補助変数の有無）や推定法式を変更して，②を再実行する。
- ④ 上記について100回の抽出と推定を繰り返し，その結果から推定量に関するMSEの分布などの特性を測定する。

当期実績と来期見通しに関する欠測データは除外しているため，仮想母集団の実際のサイズは $N=$ 約5500である。そこから46都道府県について標本を再抽出し，推定量を計算した。そのうち以下では，真値の水準や回答の分布パターンが特徴的と思われる北海道，福島，東京，山梨，奈良，島根，香川，大分の推定結果を示している。仮想母集団の特性（真値）とそれらの地域の抽出標本サイズは表1に示している。再抽出された推定実験用の都道府県別標本サイズ n としては，東京都

表1 調査データ（仮想母集団）と再抽出標本サイズ

都道府県	N	n	真値	(参考)地域ブロックn
北海道	235	47	-25.1	(北海道・東北) 152
福島県	95	19	-21.1	(北海道・東北) 152
東京都	540	108	-18.6	(関東) 289
山梨県	23	5	-43.5	(中部) 202
奈良県	44	9	-39.8	(近畿) 184
島根県	57	11	-18.4	(中国) 91
香川県	44	9	-42.0	(四国) 45
大分県	67	13	-28.4	(九州) 126
全県平均	5458	1089	-23.7	-

注) 真値とは調査データの都道府県別DI値（記述統計値）である。

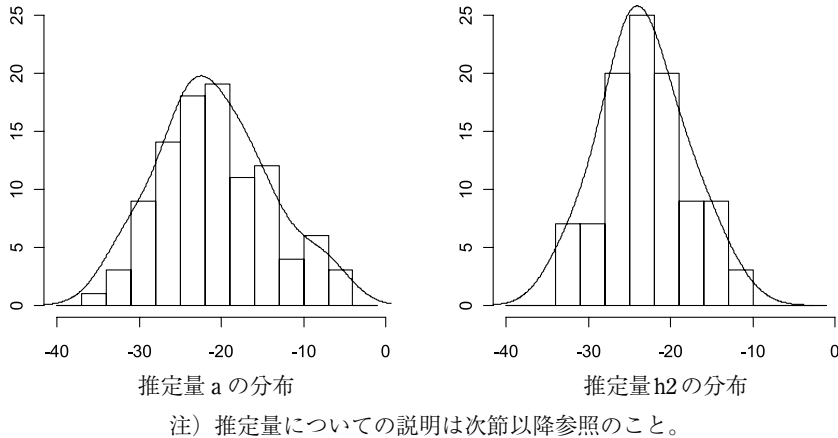


図1 推定値の分布例（福岡：N=211，n=42，DIの真値=-21.1）

で最大100, 最小は山梨県の5サンプルであった。

都道府県別DI値といっても特別な計算を要するものではない。通常、選択肢のうち「良い」に分類される回答比率(%)から「悪い」に相当する回答比率を控除した値がDIである。今回の場合、すでに示した5つの選択肢を「かなり良い」から順に100, 50, 0, -50, -100のポイントを割り当て、その平均をとることと同じである¹³⁾。したがって、都道府県別DI推定は部分母集団(地域)に対する母平均の推定問題として処理すればよい¹⁴⁾。作業結果のイメージを与えるため、図1に2つの推定方式によるDI値の違いを推定量の分布として例示した。このような結果が得られれば、推定量の特性(良し悪し)についての評価が可能となる。

3.3 推定量の候補と補助情報

1節で述べた理論的方法を具体化して、表2のようにa~hまでの推定量を候補として、沖縄を除く都道府県を対象地域(以下では、「県」と略称。i=1~46)にその県別DI値を求めることにした。まず推定量のタイプとしては、直接推定、間接推定、モデルベースの推定法の3種を取り上げ、それをさらに補助

情報の有無で区分している。補助情報がある場合には、標本から抽出した回帰関係を利用した推定(回帰推定)に絞って検討している。また間接推定については、推定対象とした地域(ここでは都道府県)の周辺標本を利用する合成推定量と、それを直接推定量との加重平均として調整する複合推定量とを区別して検証している。

合成推定量としての間接推定には、7つの地域ブロック(I=1~7)を設定した。すなわち北海道・東北、関東、中部、近畿、中国、四国および九州である¹⁵⁾。推定対象の都道府県がそのブロックに属するときには、そのブロック標本をあたかも当該県の標本であるかのように使ってDI値を推定している。

推定の目標にした変数は、既述のように業況判断の県別来期見通し(実際にはその平均としてのDI)であるが、補助情報としては従業員数(対数)と業況判断の当期実績が利

表2 検証した推定量の分類

条件	補助変数	補助変数あり
	なし	(1もしくは2変数)
直接推定	a	b1, b2
間接推定(合成)	c	d1, d2
間接推定(複合)	-	e, f1, f2
モデル	-	g, h1, h2

用できるケースを想定している。それぞれ、県別母平均が既知であり、当然、標本からの推定用データセットにそれらの補助変数が含まれることを前提としている¹⁶⁾。センサスや業務統計など他のデータリソースから比較的容易に入手できる基本情報ではあるが、目標変数との相関があまり期待できない情報の一例として従業者数(対数)を、逆に相関は高いが一般には入手できない情報の一例として当期実績を取り上げている¹⁷⁾。

実際、同調査について、上記のようにカテゴリーを点数化した来期見通しとの相関は、従業者数(対数、以下省略)で $r=0.22$ 、業況判断の当期実績では $r=0.73$ であった。因みに、従業者数と当期実績との相関は $r=0.19$ である。検証作業では、これらの補助変数を同時に使った場合と、従業者数だけを使った場合の2通りを評価することにした。補助情報の効果を判断するためである。ここで、検証で使用する変数表記と統計量の定義を改めて整理しておこう。

[母集団]

N : 母集団の大きさ

N_i : 部分母集団 (i 県) の大きさ

$y_{ij}, i=1, \dots, 46, j=1, \dots, N_i$: i 県の j 番目の母集団要素の来期見通し(目標変数)

$x_{1ij}, i=1, \dots, 46, j=1, \dots, N_i$: i 県の j 番目の母集団要素の従業者数(補助変数)

$x_{2ij}, i=1, \dots, 46, j=1, \dots, N_i$: i 県の j 番目の母集団要素の当期実績(補助変数)

$\bar{Y}_i, i=1, \dots, 46$: i 県の母平均(真の県別DI)

$\bar{X}_{1i}, \bar{X}_{2i}$: 従業者数(対数)と当期実績(業況判断)の i 県の母平均(補助情報)

[標本と目標推定量 \hat{Y}_i の特性]

s : 標本(全体), n : 標本サイズ

s_i, s_j : i 県あるいは I ブロックに属する s の部分標本

n_i, n_I : i 県あるいは I ブロックに属する標本サイズ

f : 抽出率(20%)

バイアス: $\text{BIAS} = E(\hat{Y}_i) - \bar{Y}_i$

\hat{Y}_i の分散: $\text{VAR} = V(\hat{Y}_i)$

平均絶対偏差: $\text{AD} = E|\hat{Y}_i - E(\hat{Y}_i)|$

平均平方誤差: $\text{MSE} = E(\hat{Y}_i - \bar{Y}_i)^2$

注) 分散および平均平方誤差の推定量は、小文字で $v(\hat{Y}_i)$, mse と表すことになる。

以下の結果表には、BIAS, AD, MSE を掲載している¹⁸⁾。MSE = BIAS² + VAR の関係より、不偏推定量であれば、平均平方誤差 = 分散であり、そうでない場合にはバイアスの2乗を引けば分散が求められる。そのため、直観的に理解しやすい推定量のばらつきの指標として平均絶対偏差(AD)を示している。なお \hat{Y}_i の添字・の位置には推定量の候補を示す記号が入る。真値 \bar{Y}_i は既知であるから、100回の抽出実験による推定結果の平均でこれらの統計量の理論値を近似計算している。

4. 推定量と検証結果

4.1 直接推定量

県別母集団に属する標本だけを用いた直接推定量としては、以下のa~b2までの3つの推定量を計算している。それぞれ分散推定の式も参考のため示している。

a. 直接推定(補助情報なし)

標本以外に利用できる情報が何もなければ、通常、県別標本平均 \hat{Y}_{ia} を \bar{Y}_i の推定量とするしかない。 \hat{Y}_{ia} とその分散推定式は次の通りである。

$$\hat{Y}_{ia} = \frac{1}{n_i} \sum_s y_i$$

$$v(\hat{Y}_{ia}) = (1-f) \frac{s_i^2}{n_i}$$

ただし、 s_i^2 は i 県の標本分散

b1. 直接回帰推定

(補助情報あり; 従業者数)

回帰推定（正確には差分推定）により県別DIを推定すればよい。手続き的には、従業者数を説明変数とする回帰式を県別標本から推定し、補助情報である県別従業者数平均を代入すればよい。

$$\hat{Y}_{ib1} = \hat{\beta}_{i0} + \hat{\beta}_{i1} \bar{X}_{1i}$$

$$v(\hat{Y}_{ib1}) = (1-f) \frac{s_{ie}^2}{n_i}$$

ただし、 s_{ie}^2 は*i*県の回帰残差の標本分散

b2. 直接回帰推定

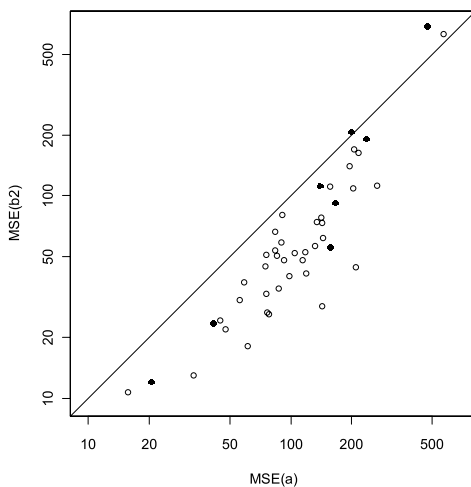
（補助情報あり；従業者数と当期実績）

b1と同様に、2つの補助変数を用いた回帰推定量を求めればよい。

$$\hat{Y}_{ib2} = \hat{\beta}_{i0} + \hat{\beta}_{i1} \bar{X}_{1i} + \hat{\beta}_{i2} \bar{X}_{2i}$$

$$v(\hat{Y}_{ib2}) = (1-f) \frac{s_{ie}^2}{n_i}$$

ただし、 s_{ie}^2 は*i*県の回帰残差の標本分散直接推定量のパフォーマンス（表3）をみると、いずれも不偏推定量であるからバイアスは0近くに分布している。しかし、標本サイズが小さな県では平均偏差と平均平方誤差に関してほとんど実用的でない過大な数字が



注) 結果数字掲載県は黒塗りで示している。後掲図3, 4, 6についても同様である。

図2 MSEの散布図

(a vs. b2; 46都道府県, 対数軸)

みてとれる。

補助情報の役割に着目すれば、相関が低い補助情報（従業者数）を使う場合（b1）では、回帰パラメータの推定誤差の大きさが補助情報による精度向上を相殺するためか、推定誤差を改善しないこと、しかし相関が高い情報（当期実績）も利用可能であれば、かなりの程度パフォーマンスが高まることを確認できる（図2）。

4.2 間接推定量 — 合成推定 —

県別母集団特性値は、それを含むより広いエリアの母集団特性値と同じであると（暗黙裡に）仮定して、当該県が属する地域ブロックの標本を用いて間接推定する。方法的には直接推定量と同じロジックを使えばよい。ただし、仮定が一般に正しい保証はないから、当然推定量には偏りがあると考えるのが自然である。推定量の良さの評価指標として、以下では平均平方誤差の推定量mseの計算式も示しておく¹⁹⁾。

c. 合成推定（補助情報なし）

所属する地域ブロック別標本平均をそのまま単純に県別母平均の推定量とすればよい。

$$\hat{Y}_{ic} = \frac{1}{n_i} \sum_{S_i} y_j$$

$$\text{mse} \approx \left(\hat{Y}_{ic} - \hat{Y}_{ia} \right)^2 - v(\hat{Y}_{ia})$$

d1. 合成回帰推定（補助情報あり；従業者数）

地域ブロック標本から従業者数を説明変数とする回帰式を求め、従業者数の県別母平均に対応するDI値を求める。

$$\hat{Y}_{id1} = \hat{\beta}_{i0} + \hat{\beta}_{i1} \bar{X}_{1i}$$

$$\text{mse} \approx \left(\hat{Y}_{id1} - \hat{Y}_{ib1} \right)^2 - v(\hat{Y}_{ib1})$$

d2. 合成回帰推定

（補助情報あり；従業者数と当期実績）

d1と同様であるが⁸⁾、補助情報として当期実績を追加した合成回帰推定を適用する。

$$\hat{Y}_{id2} = \hat{\beta}_{i0} + \hat{\beta}_{i1} \bar{X}_{1i} + \hat{\beta}_{i2} \bar{X}_{2i}$$

表3 直接推定量の結果特性

都道府県	a			b1			b2		
	BIAS	AD	MSE	BIAS	AD	MSE	BIAS	AD	MSE
北海道	-0.2	5.0	41.3	-0.5	5.0	46.9	0.1	4.0	23.5
福島県	-1.7	10.2	156.4	-2.1	10.1	161.7	-0.9	6.2	56.1
東京都	-0.1	3.7	20.5	-0.2	3.6	19.2	-0.3	2.7	12.1
山梨県	-2.1	17.3	469.1	-3.8	21.3	1119.3	-3.1	19.5	690.9
奈良県	0.8	11.8	234.5	-0.9	13.1	272.7	0.0	10.9	192.0
鳥根県	-0.7	11.3	197.7	0.0	12.1	219.8	0.9	10.2	207.4
香川県	-1.0	10.4	165.5	-2.1	10.5	174.6	-0.6	7.9	92.3
大分県	-1.1	9.1	138.9	-1.4	8.9	135.0	-0.1	8.2	112.0
全県平均	-0.1	2.3	133.8	-0.6	2.6	181.3	0.3	2.2	91.3

表4 合成推定量の結果特性

都道府県	c			d1			d2		
	BIAS	AD	MSE	BIAS	AD	MSE	BIAS	AD	MSE
北海道	-1.6	2.7	14.4	-2.9	3.2	30.3	-0.5	2.1	6.5
福島県	-5.6	2.7	43.7	-6.2	3.1	59.1	-2.0	2.0	9.5
東京都	-1.2	2.3	10.0	-1.6	2.6	15.6	-0.8	1.8	5.7
山梨県	21.5	2.5	472.5	22.0	2.7	496.6	9.3	1.7	90.9
奈良県	15.1	2.4	238.5	12.8	3.1	184.6	1.3	2.4	10.7
鳥根県	-7.3	3.5	71.8	-6.3	3.8	65.3	-1.8	2.2	11.1
香川県	7.9	5.8	115.8	6.3	6.9	115.9	6.6	3.4	62.6
大分県	3.3	3.4	29.0	2.3	4.0	31.5	0.6	2.4	9.8
全県平均	1.2	0.7	53.9	1.1	0.8	62.5	0.5	0.4	18.0

表5 複合推定量の結果特性

都道府県	e			f1			f2		
	BIAS	AD	MSE	BIAS	AD	MSE	BIAS	AD	MSE
北海道	-0.3	3.0	18.8	-2.5	3.1	21.4	-0.6	2.1	6.3
福島県	-1.7	5.0	61.2	-5.5	3.3	47.9	-1.8	2.1	10.6
東京都	-0.7	2.2	8.4	-1.2	2.5	11.0	-0.8	1.9	6.1
山梨県	2.6	12.2	279.3	19.4	4.5	425.7	8.8	2.2	87.5
奈良県	1.3	5.9	102.7	11.5	3.8	160.5	1.1	2.8	13.7
鳥根県	-0.8	5.1	59.5	-5.6	3.9	54.4	-1.8	2.5	13.1
香川県	2.1	7.1	96.0	5.8	6.5	102.6	6.3	3.7	63.3
大分県	0.4	4.9	59.1	2.1	4.0	32.5	0.4	2.7	13.1
全県平均	0.0	1.3	59.1	1.1	0.8	54.4	0.5	0.4	18.4

$$mse \approx (\hat{Y}_{id2} - \hat{Y}_{ib2})^2 - v(\hat{Y}_{ib2})$$

表4から明らかのように、単に周辺地域情報を用いたブロック平均値を県の推定量cとすることで、県平均レベルのMSEは、直接推定量aよりも大幅に低下する。合成推定によってバイアスが生じているが、ADの値が示すように見かけ上の標本サイズの増加により推定量のばらつきは縮小している。もちろん、県によってそれらのバランスは異なるが、県全体の傾向としては補助情報の条件が同じであれば、直接推定量よりも推定誤差を大幅に低下させる結果となった(図3)。ただし、合成推定量の枠内では、単純なブロック平均cに比べ補助情報を使う合成回帰推定が優れているというわけではない。直接推定の場合と同じく、相関が低い補助変数の場合にはむしろMSEは悪化する。相関が高い補助変数(当期実績)の利用可能性がパフォーマンス向上の鍵であることがわかる(d2)。

4.3 間接推定量 — 複合推定量 —

合成推定量はバイアスをもつが、標本サイ

ズの増加により標本誤差は減少する。そのバランス次第で、合成推定量のMSEを直接推定量よりも低下させることができる。しかし、不偏性をもつ直接推定量を利用してバイアスをうまく調整できれば、さらにMSEが減少できるかもしれない。そこで両者の加重平均形を考え、MSEを最小化するようなウェイトを用いて県別母平均の推定量を構成することを考える。直接推定量と合成推定量、両者の加重平均としての複合推定量である。ここでは直接推定量として当該県の標本平均(a)を、合成推定量としては、補助情報を利用する合成回帰推定量(d1, d2)を採用している。

このとき複合推定量において、個々の県別推定量のMSEを最小化するというアプローチ(県別に異なるウェイト)と、共通ウェイトで46県全体の合計MSEを最小化するというそれでは結果が異なる。そのためアプローチが異なる2つの推定量をそれぞれ算出している。なお、共通ウェイトによる複合推定量の特性はJames-Stein推定量のそれと同種のものと考えてよい。

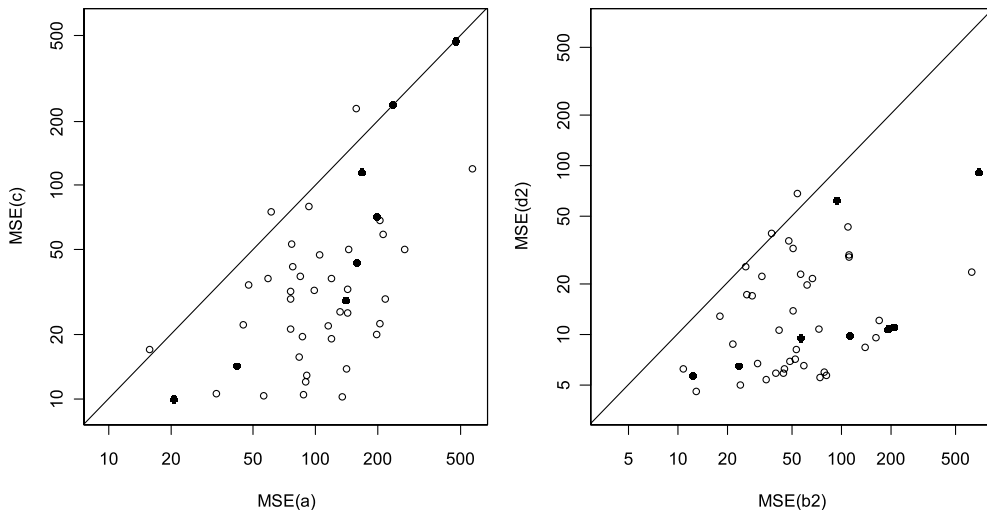


図3 MSEの散布図 (a vs. c, b2 vs. d2 ; 46都道府県, 対数軸)

e. 複合推定 (県毎の個別ウエイト, a+d2)

県別の個々のMSEを最小化するような複合推定を試みる。このとき推定量 \hat{Y}_{ie} はウエイト $\hat{\phi}_i$ により次のように表せる。合成推定量としては従業者数と当期実績を補助情報として用いたd2のケースを利用した。

$$\hat{Y}_{ie} = \hat{\phi}_i \hat{Y}_{ia} + (1 - \hat{\phi}_i) \hat{Y}_{id2}$$

$$\text{ただし } \hat{\phi}_i = 1 - \frac{v(\hat{Y}_{ia})}{(\hat{Y}_{id2} - \hat{Y}_{ia})^2}$$

f1. 複合推定 (共通ウエイト, a+d1)

全県共通のウエイトを $\hat{\phi}$ とおくと、合計MSEを最小化する推定量は次式となる。当然推定量eに比べ、県によってはMSEが悪化する。まず従業者数だけを用いた合成回帰推定量d1を使用した複合推定量を求めた。

$$\hat{Y}_{if1} = \hat{\phi} \hat{Y}_{ia} + (1 - \hat{\phi}) \hat{Y}_{id1}$$

$$\text{ただし } \hat{\phi} = 1 - \frac{\sum_i v(\hat{Y}_{ia})}{\sum_i (\hat{Y}_{id1} - \hat{Y}_{ia})^2}$$

f2. 複合推定 (共通ウエイト, a+d2)

f1と同様であるが、従業者数と当期実績を補助情報とする合成回帰推定量d2を使用して、複合推定量を計算した。

$$\hat{Y}_{if2} = \hat{\phi} \hat{Y}_{ia} + (1 - \hat{\phi}) \hat{Y}_{id2}$$

$$\text{ただし } \hat{\phi} = 1 - \frac{\sum_i v(\hat{Y}_{ia})}{\sum_i (\hat{Y}_{id2} - \hat{Y}_{ia})^2}$$

結果(表5)をみると、まず推定量eについては、個別県レベルのMSE最小化を目指しながら、推定量のパフォーマンスが単純な合成推定量cに比べても大して改善されていないこと、むしろ合成推定量d2よりも大幅に悪化していることがわかる。複合推定の考え方は悪くないが、本研究の標本サイズレベルでは最適ウエイト $\hat{\phi}_i$ 計算の前提となる県別mseの推定のぶれが大きく、むしろMSEの増大を招いている。

これに対して合計MSEの最小化を共通ウエイトで実現しようとするケースに関して

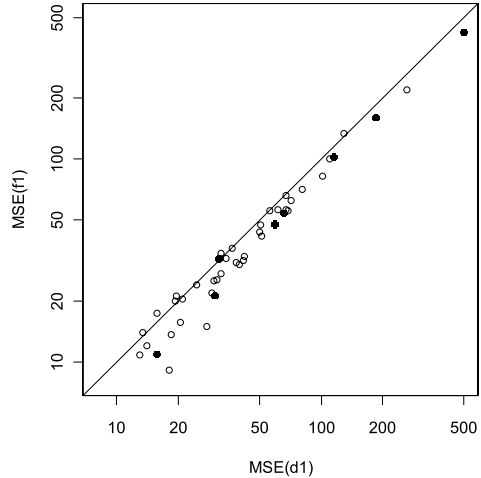


図4 MSEの散布図
(d1 vs. f1; 46都道府県, 対数軸)

は、とくに合成推定量d1に対する推定量f1の分布が示すように、ある程度の改善傾向が認められる(図4:d1とf1の散布図)。顕著ではないものの、全体としては共通ウエイトによる複合推定のアプローチがかなり有効に作用している。共通ウエイトにより標本変動固有のぶれがかなり抑え込まれたせいでもある。

いずれにしても、確率標本の下での母数推定という標準的な推定図式の枠内では、複合推定は限界点に位置する工夫といえる。とくに直接推定量と合成推定量の両者をMSE最小化をめざして結合するというアイデアは秀逸である。しかし実際には、小標本の下での確率変動が推定図式の複雑化により増幅されるマイナス効果と、他方で理論上MSEを小さくする方向で作用するプラス効果とのトレードオフ関係は曖昧である。そのため、適用に当たっては、ケースバイケースの実質的な判断が不可欠であり、実用的な推定方式と推奨するには大きな難点が残されている。

4.4 モデルによる推定

標本はある確率モデルからの実現値と考え、関心対象である変数 y_{ij} に対するモデルを標

本から推定し、それを利用して地域母数を推定する。ここでは補助変数の利用を前提に、まず回帰型のモデルから検討している。

g. 回帰モデル（補助情報あり；従業者数）

従業者数を説明変数とする次の回帰モデルを想定する。パラメータは全県共通とし、誤差項は単純に正規分布に従うものと仮定しており、実質的には合成回帰推定的一种である。

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

推定された回帰モデルを用いて、県別母平均 \hat{Y}_{ig} の推定量は次式で得られる。

$$\hat{Y}_{ig} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_{1i}$$

h1. 混合効果モデル

（補助情報あり；従業者数）

上記モデル g を拡張し、県固有の変動を量効果 τ_i として導入する。補助情報は同じく従業者だけ利用可能とする。次の混合効果モデル

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \tau_i + \varepsilon_{ij}$$

$$\tau_i \sim N(0, \sigma_\tau^2), \varepsilon_{ij} \sim N(0, \sigma^2)$$

を推定すれば、県別母平均が次式のように求められる。

$$\hat{Y}_{ih1} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_{1i} + \hat{\tau}_i$$

h2. 混合効果モデル

（補助情報あり；従業者数と当期実績）

さらに目標変数との相関が高い当期実績を補助情報として利用できる場合には、

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \tau_i + \varepsilon_{ij}$$

$$\tau_i \sim N(0, \sigma_\tau^2), \varepsilon_{ij} \sim N(0, \sigma^2)$$

を利用して、下記を県別母平均の推定量とすればよい。

$$\hat{Y}_{ih2} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_{1i} + \hat{\beta}_2 \bar{X}_{2i} + \hat{\tau}_i$$

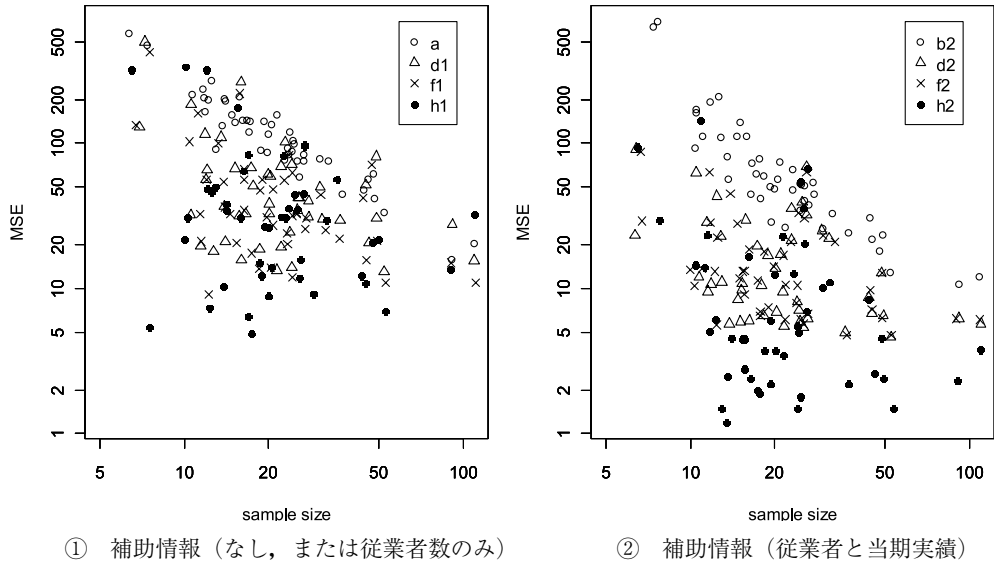
通常の回帰モデルによる推定量 g と県別変量効果を想定したモデル h1 のパフォーマンスを比較すると、県別変量効果の導入によって MSE が全体（全県平均）としてはある程度改善していることがわかる。この場合、県別要因を量効果として導入することの有用性を示唆している。また、目標変数との相関が高い補助変数を利用した場合（ケース h2）、これまでと同様に大幅に MSE は低下している。MSE の全県平均をみると、数値上は推定量の候補のうちもっともよいパフォーマンスを示している。なお、対応する複合推定量と比較した場合、格段のパフォーマンスの向上が確認できるわけではないが、f1 に対する h1、f2 に対する h2、いずれも MSE の全県平均値は若干低下している。

5. おわりに

本稿では、小地域推定の方法評価の試みとして、全国小企業動向調査の標本データを仮想的な母集団に措定して、そこからの抽出サンプルを用いた地域母数の推定実験を行った。

表6 モデルによる推定結果の特性

都道府県	g			h1			h2		
	BIAS	AD	MSE	BIAS	AD	MSE	BIAS	AD	MSE
北海道	0.0	1.5	7.1	3.2	2.5	22.0	0.5	1.1	2.4
福島県	-3.2	1.4	17.3	-2.1	2.2	14.0	-1.4	1.0	3.7
東京都	-6.0	1.5	43.1	-4.1	2.8	32.2	-1.0	1.2	3.8
山梨県	20.5	1.3	428.3	17.7	1.8	318.6	9.7	0.9	94.7
奈良県	13.8	1.6	197.6	18.1	2.0	336.5	3.4	1.1	14.0
島根県	-4.4	1.3	25.8	-6.5	1.8	48.1	-2.0	1.0	6.1
香川県	18.2	1.4	336.1	17.7	1.9	322.5	12.0	1.0	144.4
大分県	4.1	1.4	23.3	4.7	2.2	34.6	-1.7	1.0	4.5
全県平均	2.2	0.1	61.4	1.5	0.2	53.9	0.6	0.1	15.0



① 補助情報（なし，または従業者数のみ） ② 補助情報（従業者と当期実績）

図5 推定量別MSEと標本サイズ
(46都道府県，対数軸)

小地域推定の方法論理は，母集団にできるだけ仮定をおかず，標本設計に忠実な，いわばデザインベースの推定法から，標本データをモデルからの実現値とみなし，空間特性を含む関係を柔軟にモデル化し地域母数を求めるモデルベースの推定へとシフトしつつある。方法評価の焦点もそこにあり，該当地域に属する標本だけを利用する直接推定，その周辺データも利用する間接推定（合成推定，複合推定），およびこれらと同等の条件下でのモデルベースの推定量を比較した。補助情報はすべて回帰推定を基本として，できるだけ比較条件をコントロールしている。単純ではあるが，基本的な推定量を検討候補として選んでいる。

図5は，シミュレーション結果による推定量別のMSE(46県)の分布を，標本サイズ(横軸)との対応で示している。図5-①には，現実的なケースとして，利用可能な補助情報がないか，あるとしても相関が低い補助情報(従業者数)しか利用できない場合の推定量を整理した²⁰⁾。それに対して図5-②には，

かなり相関が高い補助情報(当期実績)も利用できる，稀ではあるが，幸運なケースだけを取り上げている。推定量のアプローチの違いによる特性とその良し悪しの大まかな傾向を捉えるにはこれで十分である²¹⁾。小地域推定の名が示すように，明らかに，標本サイズが小さいエリアでの推定誤差の改良が間接推定やモデル推定のアプローチの効果であることが確認できるであろう。グラフが示すように，MSEの観点では，標本サイズが相対的に小さいエリアでの低下傾向が顕著である。

また図5-①では次のような特徴を看取できるはずである。直接推定量 a に対して間接推定量(合成推定量 d1，複合推定量 f1)とモデル推定量 h1 が MSE の観点からは優位であること，なおかつ合成推定量よりも複合推定量で全体的には若干の改善がみられ，さらにモデル推定量ではかなりの県で比較的大きな MSE の低下がみられる。つまり，補助情報がないか，あるとしても相関が低い補助情報しか利用できないような場合には，直接推定量よりは合成推定量，さらには複合推定量や

モデル推定量といった推定アプローチの順に、程度の差はあれ推定精度が向上している。

他方で、相関が高い補助情報が利用できる場合（図5-②）は、そうでない場合（図5-①）に比較して下方向にシフトしたMSEのばらつきを示しており、補助情報の相関特性の高低が推定法によっては決定的であることがわかる。そのせいもあり、推定方式の違いがグラフ上で明確に浮かび上がっている。まず、直接推定量としての回帰推定量 b_2 （もっとも高いMSEを示す）に対して、合成推定量と複合推定量は明らかな改善を示しているが、補助情報の効果が高いため両者の優劣は判別し難い。しかし、モデル推定はそれら2つの推定量以上に大きくMSEを低下させていることが明らかである。いずれにしても、目標地域に属する十分な標本がなくとも推定精度を高める可能性とそのための推定量の候補は明らかとあってよい。図6は両極に位置する直接推定量 a とモデル推定量 h_2 を単純に比較したMSEの散布図であり、モデルベースへの推定方式の転換と有効な補助情報の存在が、どのように地域母数の推定量を改善するのかを端的に示している。

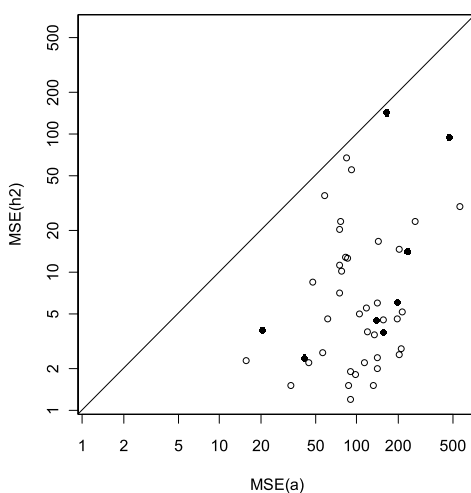


図6 直接推定 a と混合効果モデル h_2 : MSEの分布（46都道府県，対数軸）

ところで検証結果のグラフ（とくに図5-①）では、モデルベースの推定方式への転換が、間接推定（合成推定量や複合推定量）に比べ際立った優位性を示しているようにはみえないかもしれない。単純な標本平均に比べても大した改善がみられない県もいくつか存在する。当然、標本調査の論理に反してまで採用すべきアプローチなのか疑念を生じる向きもあろう。しかしそうではなく読み取るべきは、このような単純なモデル推定量でも、従来の推定図式の枠組みではもっとも複雑な複合推定量と同等（もしくはそれ以上）のパフォーマンスを傾向的に示している点である。実際ここで採用したモデル h_2 は複合推定量 $\hat{\theta}$ のいわばモデルバージョンと解され²²⁾、純然たるモデルベースのアプローチの効果をみるために導入したにすぎず、大幅な改善をそもそも期待すべきものではない。

むしろ、モデルベースのアプローチの優位性は、合成推定量にみたような限定的な仮定（例えば県母平均＝ブロック母平均）に止まらず、地域母数の変動や分布についてより複雑な多様な空間モデルを取り込むことができ、そのことが地域母数の推定精度をさらに高める可能性がある。しかも、経常的調査のように対象時点の前後の標本データが存在する場合には、そのような時系列標本も推定に利用できる柔軟性（時空間モデルへの拡張）をモデルベースのアプローチは有している²³⁾。従来の標本調査本来の推定図式の中で暗黙の仮定をおきながら、無理な工夫を凝らしてのデザインベースの推定方式に対して、いわゆる小地域推定モデルと呼ばれるモデルベースの推定へと軸足を移すには十分な理由といえる。

最後に、小地域推定モデルの有効性を実質化する条件（制約）をめぐって、補助情報の利用可能性とモデル評価の方法の2点について触れておきたい。まず、一方の補助情報の効果についてはすでに示した通りである。相関が高い補助情報が利用できれば、地域母数

の推定精度は向上する。そのためには、ターゲットとなる地域に属するそのような補助情報（地域母数、もしくはその近似としての推定値） $\bar{X}_{1i}, \bar{X}_{2i}$ の入手可能性とともに、目標変数 y_{ij} にリンク可能な標本レベルでの補助変数 x_{1ij}, x_{2ij} の利用可能性が条件となる。その実現には、データ空間拡張に向けたデータアーカイブ論に標本調査データの組込みとそのリンケージ手法を絡めて議論していかねばならない。

他方のモデル評価の方法については、本稿のように推定実験で求めたMSEではなく、実際には推定量としてのmseを頼りに最終的なモデルを選択せざるを得ない。しかし、複合推定量の最適ウェイト $\hat{\phi}$ に関して指摘したように、必ずしも安定的で信頼できる推定量mseが得られるわけではない。mseをはじめとするモデル評価の規準統計量の問題につ

いては、モデルベースのアプローチを中心に小地域推定モデル論として稿を改めて論じることになりたい。

（付記）本稿は、「政府統計データのアーカイビングシステムの構造と機能に関する国際比較研究」日本学術振興会科学研究費補助金基盤研究(B)（課題番号：22330070，研究代表者：法政大学 森博美，平成22年度～25年度）の成果の一部である。また、本研究は個票データの二次分析に基づいている。二次分析に当たっては、東京大学社会科学研究所附属日本社会研究情報センターSSJ データアーカイブから「〔全国小企業動向調査 2004年7-9月調査〕日本政策金融公庫総合研究所（旧国民生活金融公庫）」の個票データの提供を受けたことを付記して、関係諸機関への謝辞としたい。

注

- 1) 完全失業率の例は労働力調査結果（総務省）の参考数値として時系列回帰モデルによる四半期別推計値が公表されている（URL：<http://www.stat.go.jp/data/roudou/pref/index.htm>）。平均所得の事例についてはFay and Herriott(1979)，貧困世帯についてはNational Research Council(2000)を参照されたい。
- 2) 統計体系と調査形態については、森(1984, 2011)などの一連の研究を参照されたい。
- 3) 小地域推定の議論についてはRao(2003)を参照されたい。坂田(2010)はその推定論理を整理している。また労働力調査への適用をめぐって推定モデルを整理した元山・山口(2007)や高部(2004)、小泉(2004)などがある。
- 4) 本学会において、部分母集団の推定に関して明確に問題を意識した論考には、統計調査論の立場からの大屋(1959)の先駆的研究がみられる程度である（大屋(1995), pp.201-222参照）。関連して付言すれば、近年の政府統計マイクロデータの提供は、層化変数などの標本設計情報が一部秘匿された下での推定量とその誤差計算への解法を切実なものとしているが、本学会での研究蓄積は社会生活基本調査（総務省）を取り上げた栗原(2010)など、こちらもまだ数える程にすぎない。部分母集団の推定問題とともに学会としての取り組みが必要な領域である。
- 5) 本節の詳細については、Rao(2003)の2-7章、あるいは坂田(2010)を参照されたい。
- 6) 厳密には抽出法の違い（復元、非復元）によって抽出確率や包含確率による定義が必要だが、このような直観的表現でもいまの議論には影響しない。なお、直接推定における標本調査法の数理については土屋(2009)を参照されたい。
- 7) 単純無作為抽出のケースについてであるが、Cochran(1977, pp.34-38)の記述を参照されたい。
- 8) 単純無作為抽出の場合は、 $\hat{Y}_i = \hat{Y} = \frac{1}{n} \sum_s y_j$ と推定していることになる。
- 9) 全地域の合計MSEを最小化するような共通ウェイト ϕ を用いた複合推定量を考えればよい。
- 10) 複合推定量とJames-Stein推定量、およびこれらのモデルとの関係についてはRao(2003), p.63以降を参照されたい。

- 11) 東京大学社会科学研究所附属社会調査・データアーカイブ研究センターは、日本における社会科学の実証研究を支援することを目的として、データアーカイブ (SSJDA) を構築し、個票データの提供を1998年4月から行っている (URL: <http://ssjda.iss.u-tokyo.ac.jp>)。
- 12) SSJDAによる提供情報に基づく。また、日本政策金融公庫総合研究所のサイト (URL: <http://www.jfc.go.jp/findings/gri/>) も参照されたい。
- 13) DIの算出については、例えば坂田(2009)参照。なお同調査の公表DI値 (日本政策金融公庫、旧国民金融公庫)の計算は、3つのカテゴリーに再統合して計算されており、本稿のDI値とは異なることに注意されたい。
- 14) 本来DIは動向把握統計としてDIの差の推定を問題とすべきだが、今回は推定法の比較に関心があるので単純にDIの水準を取り上げている。
- 15) 全国小企業動向調査結果では10地域ブロック別の数字が公表されていたが、本稿では抽出率の関係もあるので7ブロックにグループ化している。
- 16) ブロック別母平均が既知の場合の回帰推定量についても検証を行ったが、推定精度は必ずしも向上せず、むしろ悪化するケースも見られる。そのためここでは取り上げていない。
- 17) 実際には、地域区分の表章の問題もあり、利用可能な補助情報の範囲は極めて狭い。
- 18) 小地域推定の性格としては、真値に対する誤差 (あるいはその2乗) の比率を問題にすべきかもしれないが、DIの定義域を $-100 \sim +100$ の区間に設定したこと、またDI作成のための原データは多項分布特有の制約をもつことなどを考慮して、真値に対する相対誤差指標はここでは取り上げていない。推定法の良し悪しの相対比較が可能であれば本稿の目的には十分である。
- 19) 平均平方誤差の推定量mseについては結果表に掲載していないが、実際には1回の標本抽出により評価せざるを得ないので、その重要性を考慮してmseの推定式を示している。なお、3.3節の複合推定量の計算はmseに基づいている。推定式についてはRao(2003)、坂田(2010)などを参照。
- 20) 直接推定量の b_1 については、単純な標本平均を推定量とした a よりMSEの悪化を示しているので、グラフからは除外した。また、地域ブロック平均を推定量とする c については、本稿の趣旨が明確に読み取れるように図示を控えている。
- 21) 推定量の特性を示す様々な統計量を計算できるが、紙面の制約もあり、本稿ではMSEの分布に限定している。
- 22) 混合効果モデルと複合推定量、JS推定量の関係についてはRao(2003)、pp.116-118参照。
- 23) 本稿で取り上げたDIは景況の時間的変動を捉えるための指標であり、そもそも経常調査として実施されている。そのため本格的な都道府県別DIの推定に当たっては、このような枠組みにおいて推定モデルを検討しなければならない。

参考文献

- [1] 大屋祐雪 (1959) 「標本統計資料の吟味 — 統計利用者のための標本統計論 I —」, 『熊本商大論集』, 9号, pp.85-115.
- [2] 大屋祐雪 (1995) 『統計情報論』, 九州大学出版会.
- [3] 栗原由紀子 (2010) 「社会生活基本調査マイクロデータにおける平日平均統計量と標本誤差の計測」, 『統計学』, 99号, pp.20-35, 経済統計学会.
- [4] 小泉英希 (2004) 「いくつかのSteinタイプの推定量の導入と評価方法」, 『統計研究彙報』, 第61号, pp.139-179, 総務省統計研修所.
- [5] 坂田幸繁 (2009) 「景気動向調査 — “Business Tendency Surveys”(OECD) — の方法と性格について」, 『熊本学園大学経済論集』, 第15巻第3・4合併号, pp.127-153, 2009.03.
- [6] 坂田幸繁 (2010) 「標本調査データからの地域母数の推定 — 直接推定と間接推定 —」, 『中央大学経済研究所年報』, 第41号, pp.191-210, 中央大学経済研究所.
- [7] 高部 勲 (2004) 「小地域推定各手法の労働力調査への適用 — 都道府県別完全失業率の推定 —」, 『統計研究彙報』, 第61号, pp.1-138, 総務省統計研修所.
- [8] 土屋隆裕 (2009) 『概説 標本調査法』, 朝倉書店.
- [9] 元山 齊, 山口幸三 (2007) 「小地域推計と労働力調査への適用」, pp.66-73, 『統計』, 2007年

2月号.

- [10] 森 博美 (1984) 「統計調査の諸形態」, 広田, 大屋, 是永, 野村編 『統計学』, 産業統計研究社.
- [11] 森 博美 (2011) 「調査形態論再論」, オケージョナル・ペーパー, No. 23, 法政大学日本統計研究所.
- [12] Cochran, W.G. (1977), *Sampling Techniques*, 3rd ed., New York : Wiley.
- [13] Fay, R.E., and Herriott, R.A. (1979), Estimation of Income from Small Places : An Application of James-Stein Procedures to Census Data, *Journal of American Statistical Association*, **74**, pp.269-277.
- [14] National Research Council (2000), *Small-Area Estimates of School-Age Children in Poverty : Evaluation of Current Methodology*, C.F. Citro and G. Kalton (Eds.), Committee of National Statistics, Washington, DC : National Academy Press.
- [15] Rao, J.N.K. (2003), *Small Area Estimation*, John Wiley & Sons.

Extraction of Small Area Information based on Sampling Survey Data :

Experimental Estimation of Prefectures' Diffusion Indexes

Yukishige SAKATA

(Faculty of Economics, Chuo University)

Summary

Methods of small area estimation based on sampling survey data have improved with the shift from the traditional direct design-based approach to the indirect model-based approach. To evaluate the efficiency of these different methods with statistical simulation, this study uses the micro data for a business tendency survey for small business as the virtual population of small businesses in Japan. The estimates for each method are then calculated from the resampled data and are iterated 100 times. The differences in the characteristics of these methods are determined by the distribution of the estimators from this simulation. Thus, the study concludes that estimation through the model-based approach is strikingly predominant, as compared with the other approach, in terms of the mean squared error, and that the degree of this predominance depends on the statistical characteristics of the auxiliary information.

Key Words

small area estimation, sampling survey, resampling method, diffusion index of business conditions, random effect model