数理解析研究所講究録
第749巻 1991年 101-138

国民生活基礎調査における平均所得の推定精度の改善について（２）

千葉大学理学部　田栗　正章（Masaaki Taguri）
千葉大学工学部　井上　隆勝（Takakatu Inoue）

１．前回の研究結果と今回の研究課題

前回の研究［４］では、厚生省の実施している国民生活基礎調査をとりあげ、補助情報を用いる回帰推定により、推定値の精度を向上させる方法について検討した。具体的には、まず世帯票および健康票の調査データと、所得票および貯蓄票の調査データとのリンケージを行なう。ここで後者の調査は、前者の調査対象をマスターサンプルとする、抽出率約1/5の標本調査と考えられる。そこで所得票の１つの主たる調査項目である世帯の総所得について、いくつかの世帯票の調査項目に対する調査結果を補助情報（説明変数）として用い、重回帰モデルにより、その平均値推定量の精度を向上させる方法について研究を行なった。その研究結果は、［４］にまとめられている。［４］では、４種類の推定方法を提案し、

従来から用いられてきた方法を含めて、シミュレーション実験により比較、検討を行なった。ここで推定目標は、6個に分けた各年齢階級における世帯総所得の平均値であり、推定方法の精度の評価は、推定値のシミュレーション実験に亘る平均2乗誤差を、6つの年齢階級について平均した規準を用いて行なった。その結果、次のような知見が得られた。

(ｲ) 提案した、重回帰モデルを用いる推定方式による推定値の精度は、従来からの推定値（各年齢階級における単純平均）の精度に比べて高い。特に n = 300（抽出率1/20）の場合には、平均2乗誤差は、1/6程度になる。

(ﾛ) 提案した推定量の内では、予測域を考慮した Modified OLS推定量による推定方式を用いたものが、平均2乗誤差の意味では最も良さそうである。したがって予測域を考慮した重回帰分析を行なう必要がある。

このように提案した推定方法によれば、従来から用いられてきた方法と比べて精度の高い推定が行えることが判明したが、この方法が実際の官庁統計の場で常用されるためには、いくつかの検討すべき点がある。そこで今回の研究課題としては、これらの問題点に対する検討を採り上げ、以下に述べるような研究を進めることにした（［4］p.21～22, 問題点参照）。

まず第1に、前回の研究においては、解析の対象としたの

は1地域ブロックのみであったが、提案する手法の安定性を検討するために、ここでは与えられた12個の地域ブロック全てに対して解析を行い、結果の考察を行う。また標本抽出方法に関しては、前回の研究では単純無作為抽出法を用いたが、現実には集落抽出法が使われている。そこで今回の研究では、現実と同じ集落2段抽出法を用いる場合の解析を行う。さらに前回は、推定点としては想定した母集団の各年齢階級の平均値$\overline{X}_k$を用いていたが、現実の場では例えば世帯票の情報等から$\overline{X}_k$を推定しなければならない。そこで本研究では、$\overline{X}_k$の推定値を用いる場合について解析を行う。

第2に、重回帰モデルにおいて用いられる説明変数に関する検討を行う。上で述べた各研究においては、家計支出額およびそれに関連する2種類の合成変数を、重回帰モデルの説明変数として用いている。これらの変数は、適当な仮定の下で導出された合成変数をベキ変換したものと解釈でき、数理統計学的観点からの意味付けは可能である。しかし現実の場で用いるべき変数としては、それが常識的に必ずしも理解され易いとは限らないという意味で、さらに検討すべき余地がある（［3］ p.4, p.159, p.204 参照）。そこでここでは、重回帰モデルの説明変数として家計支出額とそれに関連する合成変数を用いない場合について検討する。

第3に、上に述べた第1の研究においては、ある地域ブロック（地域12）で、提案した方法による推定精度が極めて悪かった。その原因は、シミュレーション実験における標本抽出でかなりの重複があったことに起因していると考えられた。さらに世帯票の調査対象（マスターサンプル）から所得票の調査対象を抽出する第2段の抽出率 f_2 の大きさに応じて、推定精度がどの程度になるかということは、実際の場で f_2 を決定するという観点からも興味のある問題である。そこでここでは、f_2 の変更に対応する推定精度の変化について検討を行う。

本研究における解析手法は、前回の研究［4］と同様であるが、本稿第2節で推定方法とその評価方法について、簡潔に記述しておく。第3節では、今回の第1の研究課題について考察を行なう。第2，第3の研究課題については、それぞれ第4節，第5節でその検討を行なう。最後に第6節で、得られた結果をまとめておく。

2. 推定方法とその評価方法

2. 1. 調査の概要と解析の目的

厚生省は昭和61年以降、国民生活基礎調査を実施している。

この調査では、調査票の種類は、①世帯票、②健康票、③所得票、④貯蓄票に分かれている。そして3年に1度大規模調査が行なわれ、中間年については世帯票，所得票のみについて、約1/5の規模の小規模調査が行なわれている。調査対象は全国の世帯及び世帯員であり、標本抽出は国勢調査区を1次抽出単位とする、層化集落抽出法が用いられている。

ところで所得票，貯蓄票に関する調査は、世帯票，健康票の調査対象の中から標本を抽出して行なわれている。しかるに現行の推計方法では、例えば世帯の平均所得の推定は、所得票の調査対象から得られるデータのみに基づいて行なわれている。しかし、世帯票の調査項目の中には、世帯の平均所得と関連を持つものもあると考えられるので、所得票の調査より大規模な世帯票の調査の結果を利用することにより、推定精度が改善できる可能性がある。これは標本調査論における2重抽出法と類似の考え方である。このような考え方に基づいて、世帯票の属性を活用することにより、地域ブロック別，世帯主の年齢階級別に、世帯の平均所得の推定精度の向上を目的とするいくつかの推定方法を提案し、従来から使用されてきた方法も含めて、比較，検討を行なうことが、本解析の目的である。

上述したように、解析の方針としては、世帯の平均所得の

値y（目的変数）と世帯票の各調査項目の値x（説明変数ベクトル）との間の相関関係を利用して、重回帰モデルにより平均所得推定値の精度の改善を計る（図1参照）。

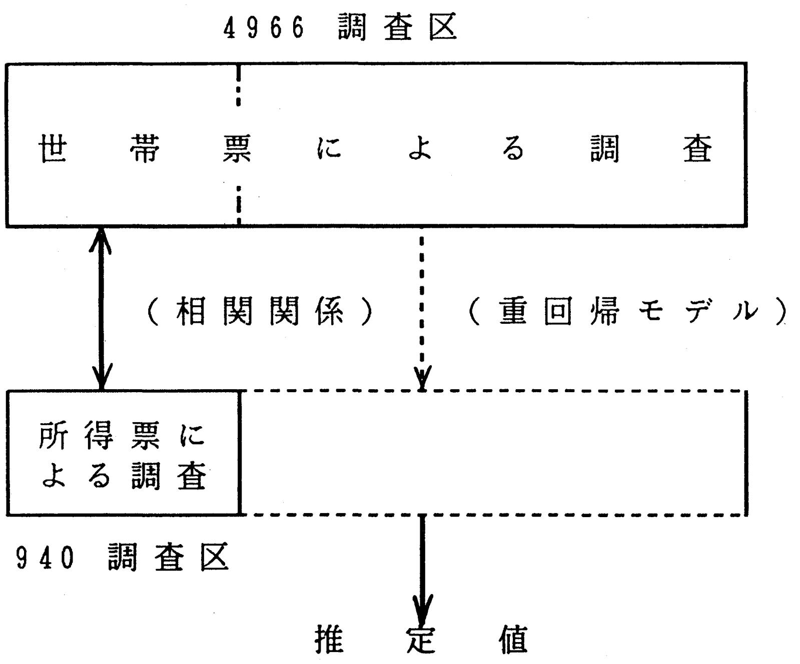

図1. 提案する推定方法の模式図

2. 2. シミュレーション実験と評価の方法

ある地域ブロックにおける、世帯主の年齢階級別の平均世帯所得を推定する、5種類の方法を比較，検討するために、シミュレーション実験を行なう。使用するデータは、昭和61年の国民生活基礎調査における、世帯票および所得票による調査結果のデータである。 本解析では1世帯の所得（総所

得）の平均値を精度よく推定するために、世帯票のいくつかの項目の調査結果を用いるが、ここで使用する調査項目およびデータの型等は、表1に示す通りである。

表1. 世帯所得推定のために用いる世帯票の調査項目

変数名	項目	データの型
y	総所得	実数型
x_1	世帯人員	実数型
x_2	有業人員	実数型
x_3	家計支出額	実数型
x_4	世帯主年齢	実数型
x_5	夫婦組数	実数型
d_1	世帯構造	ｶﾃｺﾞﾘｰ型
d_2	世帯業態	ｶﾃｺﾞﾘｰ型
d_3	世帯種	ｶﾃｺﾞﾘｰ型

ここで、カテゴリー型説明変数のカテゴリーは、次の通りである。

世帯構造 ： 1.単独 2.夫婦のみ 3.夫婦と子
4.片親と子 5.三世代 6.その他

世帯業態 ： 1.役員 2.小企業 3.大企業 4.日雇
5.自営雇人有 6.自営雇人無 7.内職
8.専業農家 9.兼業農家

世帯種 ： 1.国民保険 2.被雇用者保険 3.両方

これらの変数については、上記の 6 x 9 x 3 個の各カテゴリー（層）における度数，yの平均，標準偏差，yとxとの散布図，相関係数の値等を参考にしてカテゴリー（層）の合併を行い、ダミー変数として重回帰モデルに取り込むことにした。すなわち層の数を、全部で 2 x 3 x 2 個とし、次節以降の解析では、これらの層に対応する 12－1＝11 個（計画行列のランク落ちを防ぐため11とする）のダミー変数を用いる。なお各カテゴリー型変数の層の分割は、［3］，［4］に与えられたものと同じである（［3］ p.193, p.195 参照）。
世帯主年齢の階級の区分は、実際に使われているものと同じで、次の通りである。

階級1： 〜29歳 階級2：30〜39歳 階級3：40歳〜49歳
階級4：50歳〜59歳 階級5：60〜69歳 階級6：70歳〜

さていくつかの推定方法の比較を行うためには、母集団の情報が既知でなければならない。そこで本シミュレーション

実験において想定する母集団は、ある地域ブロックにおいて、世帯票・所得票両方のデータが得られた世帯とする。ここで解析の対象とした母集団の大きさNは、表4等に示すように12個の地域ブロックに対応してN＝731～6158世帯であった。ここで12個の地域ブロックは次の通りであるが、それらに付けた地域ブロック番号は、無作為化してある。

北海道	東北	関東Ⅰ	関東Ⅱ	北陸	東海
近畿Ⅰ	近畿Ⅱ	中国	四国	北九州	南九州

標本抽出方法としては、なるべく実際に使われている方法に近づけるために、上記の母集団からの擬事的な集落（2段）抽出法を用いた。ここで、次のように記号を定める。

f_1: 集落の抽出率,

m : 抽出された集落の数,

n_1: 集落に含まれる世帯数,

f_2: 第2段の世帯の抽出率（$\geqq 1/5$）,

n_2: 調査対象とする世帯数（$n_2 \geqq 600$）。

シミュレーション実験における第1段の抽出単位は、現実の集落（調査区）から1/2の確率で無作為抽出した世帯の集まりとし、これを擬似単位区（大きさは約25世帯）と呼ぶ。調査区の抽出、および抽出された調査区からの擬似単位区の抽出は重複を許して行なった。次節以降で行なう重回帰分析の信

頼性を保証するために、$n_2 \geqq 600$ が全ての場合について満たされるように、$n_1 = n_2 / f_2 \fallingdotseq 3000$と設定し、この数が得られるまで擬似単位区を抽出した（詳しくは［3］p.190～191,［5］を参照）。

シミュレーションのやり方は、上で想定した母集団からの標本抽出実験を、100回行なう。毎回の実験では、得られた標本に基づいて、5種類の推定方法による推定値を計算する。そして100回のシミュレーション実験に亘る、推定値の平均，標準誤差および平均2乗誤差を計算し、推定方法の精度の評価を行なう。

2．3．重回帰モデルと5種類の推定方法

推定目標であるk番目の年齢階級における平均値μ_kを推定するために、次の重回帰モデルを考える。

$$y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon, \qquad \varepsilon \sim (0, \sigma^2)$$

ここでεは平均 0，分散σ^2をもつある分布に従うとする。

μ_kの推定値としては、

$$\hat{\mu}_k = \overline{\mathbf{x}}_k' \mathbf{b}$$

を用いる。ここで$\overline{\mathbf{x}}_k$はk番目の年齢階級における説明変数ベクトル$\mathbf{x}$の平均で、世帯票の情報から推定する。また$\mathbf{b}$は、

以下で述べるいくつかの方法により推定された$\boldsymbol{\beta}$の値を、一般的に記述した記号とする。各年齢階級における平均世帯所得μ_kの推定値の安定性という観点からは、変数選択（モデル選択）を行なうべきである。ここでは実数型説明変数のみを変数選択の対象とし、ダミー変数は変数選択の対象としない強制変数とする。

ここで5種類の推定方法について説明を行なう。上述の重回帰モデルを、目的変数ベクトル$\mathbf{y}$，計画行列X，回帰係数ベクトル$\boldsymbol{\beta}$，誤差ベクトル$\boldsymbol{\varepsilon}$を用いて書き直しておくと、

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2 I_n)$$

となる。ただしI_nはn次の単位行列である。ところで我々の目標は、個々の世帯の所得を精度よく推定することではなく、6つの年齢階級における世帯所得の平均値$\mu_1, ..., \mu_6$を精度よく推定することである。したがって推定目標は

$$\mu_k = \overline{\mathbf{x}}_k' \boldsymbol{\beta}$$

であり、6個の点の上での推定を考える（k = 1, 2, ..., 6）。

ここで

推定域：推定を行うべき6つの点$\mathbf{x}_1, ..., \mathbf{x}_6$の集合

標本域：標本が得られている$\mathbf{x}$空間内のデータ点の集合

（所得票のデータが得られている世帯の、世帯票の項目［$\mathbf{x}$］の値の集合）

とする。我々の問題では、各推定点上での推定量の精度を同等に評価すべきであると考えられた。すなわち、ある特定の年齢階級における平均所得を特に精度よく推定したい訳ではなく、どの年齢階級における平均所得の推定量のバラツキも同等に評価すべきであると考えられた。したがってモデル選択の規準とするリスク関数としては、次の量を用いる。

$$r(\hat{\mu}_k) = K^{-1}\sum_{k=1}^{K} E\{(\mu_k - \overline{\mathbf{x}}_{ik}'\mathbf{b}_i)^2\}$$

$\Delta_{ii} = K^{-1}\sum_{k=1}^{K}(\overline{\mathbf{x}}_{ik}\overline{\mathbf{x}}_{ik}')$ とおき、Δ_{ii}^{+}を Δ_{ii}のMoore-Penrose型一般逆行列と定義すると、モデル M_iの下でリスク関数 $r(\hat{\mu}_k)$ を最小にする回帰係数ベクトル$\boldsymbol{\beta}_{i*}$は、$\boldsymbol{\beta}_{i*} = \Delta_{ii}^{+}\Delta_{iq}\boldsymbol{\beta}$ と表せる。ここでリスク関数を分解して考えると、「推定域」上でのμ_kの推定に関しては、モデルM_iの選び方および$\boldsymbol{\beta}_i$の推定量$\mathbf{b}_i$の選び方により、異なる推定方法が考えられることが判る。［4］では4種類の推定方法を提案したが、本解析でもこれらの方法により推定を行なうことにする。以下ではこれらの推定量に、従来から用いられてきた年齢階級別単純平均値（Y-SMPと略す）を加えた5種類の推定方法について検討を行なうが、それらの記号と簡単な説明を表2にまとめておく。

1. $\mathbf{b}_1$: OLS推定量 $\hat{\beta}_1$, 選択規準：Cp 規準
2. $\mathbf{b}_1$: OLS推定量 $\hat{\beta}_1$, 選択規準：標本域と推定域の違いを考慮した規準
3. $\mathbf{b}_1$: M-OLS推定量 $\tilde{\beta}_1$, 選択規準：標本域と推定域の違いを考慮した規準
4. $\mathbf{b}_1$: $\hat{\beta}_1$, $\tilde{\beta}_1$を毎回選択, 選択規準：標本域と推定域の違いを考慮した規準

(5.)従来から用いられてきた年齢階級別単純平均$\overline{y}$

表2. 用いた推定方法

#	記号	β_1の推定量	選択規準
1.	OLS-X	: $\hat{\beta}_1 = (X_1'X_1)^{-1}(X_1'X)\hat{\beta}$	標本域での評価
2.	OLS	: $\hat{\beta}_1 = (X_1'X_1)^{-1}(X_1'X)\hat{\beta}$	推定域での評価
3.	M-OLS	: $\tilde{\beta}_1 = \Delta_{11}{}^{+}\Delta_{1q}\hat{\beta}$	推定域での評価
4.	SEL.	: $\hat{\beta}_1$, $\tilde{\beta}_1$ の選択	推定域での評価
(5.)	Y-SMP	: $\overline{y}$ (階級別単純平均)	(従来からの方法)

3. 集落2段抽出法を用いた場合の、全地域ブロックに対する推定結果

本節では、実際に用いられている抽出法を模擬した2.2節に示した集落2段抽出法を用いる。その際の推定点$\overline{x}_k$の値は、実際に得られる1次標本からの値を用いることにする。また、以下の解析では、説明変数として世帯主年齢(x_4)を用いた合成変数$(x_4 - 50)^2$を追加する。これは、地域08における残差分析の結果、および参考文献[1]より判断した。

3. 1. 集落2段抽出法の場合の推定結果

2.3節で与えた5種類の推定方式を用いて、2.2節で述べた集落2段抽出法によるシミュレーション実験を、各地域ブロックについてH＝100回ずつ行なった。地域01における結果、および各地域ごとのMSE値はそれぞれ表3、表4にまとめられている。まず、表3の結果を前回の研究結果（［4］ p.27, 表9参照）と比較すると、標本抽出法として単純無作為抽出法のかわりに集落2段抽出法を用いると、推定精度にかなり大きな影響のあることがわかる。これは、標本抽出法が調査コストを下げる目的で提案された方法で、各集落(調査区)が母集団の構造を反映していない状況では、単純無作為抽出法と比べて、その推定精度が低下することによるものと考えら

れる。

次に、表4の結果より、提案した4つの推定方式について比較すると、3:M-OLSが1カ所(地域04では4:SEL.の方が2.5%小さい)を除き最小値を与えている。特に推定域を考慮しない方法1:OLS-Xとの比較を行なうため、表4の右から2列目に、それらの比の値が示されている。この値を見ると、推定域を考慮した方法を用いると、MSE値からみた改善は2～7%程度と小さいが、改悪のケースは存在しなかった。

また、従来からの方法(Y-SMP.)と3:M-OLSとを同様に比較すると、表4の最右列の結果を得る。この値を見ると、地域01で約7割、地域04, 12を除く他の地域で1～4割程度の改善が見られる。

以下では、問題の2つの地域04, 12に対する検討を含めて、推定方式の改良を試みる。

3. 2. 推定方式の改良

地域01を例に、母集団における回帰誤差に着目し、年齢階級別に見た回帰誤差の平均、分散を求めると表5の結果を得る。これを見ると、高年齢階級における誤差分散が他の年齢階級のそれに比べて顕著に大きい。このことは、年齢階級別に重みを与える重み付き最小2乗法の適用を示唆している。

そこで、実際に得られる2次標本に対して通常の回帰分析を行ない、そのときの規準化残差(規準化誤差の推定量)に対して各年齢階級別に分散を求め、この値の逆数を重みとする重み付き最小2乗法を適用した。その結果を表6に示す。表6の下段の数値は重みを付けない場合のMSE値に対する比率を示す。表6を見ると、地域12および地域05の一部を除く全地域で重み付きの方がよく、3～4割の改善が見られる地域もある。また、表6の右から2列目に示されている、3:M-OLSの1:OLS-Xに対する改善度は、重みを付けない場合に比べて大きく、2～20%である。Y-SMP.との比較では、地域04で重みを付けることによる改善が認められるが、地域12では認められなかった。

地域12において、回帰モデルに基づく1:OLS-Xから4:SEL.までの推定方式がY-SMP.に比べて推定精度が悪い原因についてさらに検討を行なった結果、シミュレーションにおける標本抽出の方法に問題があることが判った．この点についての詳しい解析は［5］に記述されている。

4. 重回帰モデルの説明変数に対する考察

第3節までの解析においては、表1に示した連続型の原変数から合成された2つの変数を、各年齢階級の平均所得の推

定を行なうための重回帰モデルの説明変数として用いた。すなわち $x_6=(x_3)^{1/4}$ [x_3は家計支出額] および $x_7=(x_1*x_3)^{1/4}$ [x_1は世帯人員] である。合成変数x_7は、1人当りの所得(総所得/世帯人員)が家計支出額に比例するとして導き出されたもので、x_6, x_7とも試行錯誤によりベキ変換のパラメータを決定した。これらの変数の値は、原変数のデータが与えられれば、それから計算できるものであり、上記のようなある程度までの意味付けが行えるので、これらを用いて重回帰モデルにより推定を行うことは、数理統計学的にはあまり問題がないと考える。しかし経済学的観点からすると、世帯の総所得の推定に家計支出額に関する変数を用いることには、かなりの抵抗があるようである。もちろん変数の解釈のし易さを考えれば、これらの変数を用いない方が望ましいが、この場合には推定精度が問題となる。そこで本節では、表1の5個の原変数から家計支出額を除いた4個の変数、すなわちx_1: 世帯人員, x_2: 有業人員, x_3: 世帯主年齢, x_4: 夫婦組数 を説明変数として用いる重回帰モデルについて検討する。カテゴリー型変数である世帯構造, 世帯業態, 世帯種については、前回の解析方法でとくに問題はないと考えたので、第2節で述べたようにこれらは11個のダミー変数(層別変数)として用いる。

ところで［3］ p.201 で述べたように、 $x_5 \equiv (x_3 - 50)^2$［x_3は世帯主年齢］という変数を追加して解析を行なうと、かなりの推定精度の改善のあることが判っている。これは世帯主の年齢と世帯の所得との関係を解析した結果得られたもので、その現実的意味付けは容易であろう。すなわち種々の職業を平均して考えれば、世帯の所得は世帯主の年齢が50歳位の時に最大となることを表わす変数と解釈できよう。そこで本節の解析では、上記4個の連続型変数と11個のダミー変数に、上で定義した変数 x_5 を追加して、重回帰分析を行なうことにする。

表7は、地域01のデータに対して、これらの説明変数を用いて重回帰分析を行なった結果である。これと比較するために、3.1節で行なった合成変数を用いた場合の計算結果を、表8に与えてある。これらの表を見ると、重相関係数，系列相関係数の値等は、2つの場合であまり差がない。また $x_5 \equiv (x_3 - 50)^2$ の係数は負であり、この変数に関しては上で述べたような意味付けが行える。分散分析の結果は、いずれの場合も極めて有意である。さらに回帰平面上の値は、いずれの年齢階級においても、y（世帯所得）の真の平均値に近い。また真の平均値，回帰平面上の値とも、5番目の年齢階級において、一度減少した後増加に転じていることを考えると、

この重回帰モデルは、yの平均値の構造をある程度まで捉えていると解釈できよう。したがってyの平均値の推定という観点からは、家計支出額およびそれに関連する2つの合成変数を用いなくても、それらを用いた場合の推定と同程度の精度が得られる可能性がある。

そこでこれら16個の説明変数を用いて、2.3節で述べた5つの推定方法（従来からの方法を含む）により、全地域ブロックに対して、平均世帯所得の推定のシミュレーション実験を行なった。まず重みなしの場合の結果が、表9にまとめられている。各セルの1段目，2段目には、合成変数を用いない今回の解析の場合および合成変数を用いる前回の解析の場合の、シミュレーション実験に亘る、yの平均値推定量の平均2乗誤差の値が与えられている。またこれらの値の比が、各セルの3段目の（　）内に示されている。これらの数値を見ると、3.1節で述べた合成変数を用いる場合に成立していた性質が、今回の合成変数を用いない場合にも成立していることが判る。すなわち表9の右から2列目の数値を見ると、ほとんどの場合、3の推定法（3:M-OLS）がよいことが判る。また表9の1番右の列の数値から、地域12を除けば、3の推定方法は従来の方法による推定精度を、かなり改善していることが判る。地域12における推定精度が悪い原因については、第3

節の解析の場合と同様、標本抽出にかなりの重複があったためと考えられる。したがって実際の場では、全く問題がないと考える（詳しくは、［３］ p.203 および［５］参照）。これらの事実から、合成変数を用いない今回の場合においても、推定方法としては、3:M-OLSがよいと考えられるので、以下では主として3の方法について議論を行なう。

さて表9の各セルの3段目の数値を、3の推定方法について見ると、合成変数を用いない場合と用いる場合の差は、かなり小さいことが判る。この数値が1より小さいことは、合成変数を用いない場合の方が、平均2乗誤差の値が小さいことを意味しており、このような場合が、12地域中5地域ある。1より大きな場合でも最大1.186であり、合成変数を用いなくても、それらを用いる場合と同程度の精度の推定が行えることが判明した。これは、変数 x_5 の追加とダミー変数（層別変数）の構成が妥当なものであったことによると考えられる。

次に重み付けを行なった場合の、同様な計算結果が、表10に与えられている。表中の各セルの1段目，3段目には、合成変数を用いない場合と用いる場合の、シミュレーション実験に亘る、yの平均値推定量の平均2乗誤差の値が与えられている。これらの値の比は、各セルの5段目の（　）内に示されている。各セルの2段目，4段目の数値は、合成変数を用

いない場合と用いる場合についての、重みなしの場合に対する比の値を表わしている。これらの数値を見ると、重み付きの場合も重みなしの場合と同様の考察ができることが判る。すなわち表10の右側の2つの列の数値より、地域12を除き、3の推定方法（3:M-OLS）の精度が最も高く、それは従来からの方法による精度をかなり改善している。また合成変数を用いない場合と用いる場合について、3/1, 3/Yの値は近い値となっていることも判る。

次に表中の2段目，4段目の数値を比較すると、重み付きの場合の重みなしの場合に対する比は、合成変数を用いない場合と用いる場合で、かなり近い値となっている。そして地域12を除けば、合成変数を用いない今回の場合では、この比の値は全て1以下となっており、0.6～0.8程度の値もかなり存在している。したがってこの場合には、重み付き最小2乗法を行なうことにより、平均2乗誤差が多少改善される傾向にある。

表10の5段目に示されている合成変数を用いない場合と用いる場合の比較に関しては、3の推定方法ではそれらの差は小さく、地域12を除けば、最大でも1.189である。また比の値が1より小さい地域が5個あり、これらの地域では、合成変数を用いない場合の方が推定精度が高いことを示している。

以上の考察より、家計支出額およびそれに関連する合成変数を用いなくても、かなり精度の高い推定が行えることが判った。提案した4種類の推定方法の内では、第3節の場合と同様3:M-OLSがよく、この方法によれば、従来の方法による推定精度をかなり改善できる。すなわち地域12を除けば、重み付き最小2乗法を用いた場合、平均値推定量の平均2乗誤差は、23～85％程度にまで減少できる。また重み付き最小2乗法を行なうことにより、多少精度が改善されると考えられる。

5. 第2段抽出率と推定精度の関係

本節では、世帯票の調査対象から所得票の調査対象を抽出する際の、第2段の抽出率f_2を変化させ、世帯所得の平均値推定量の平均2乗誤差にどのような影響があるかを検討する。実際の国民生活基礎調査において用いられている数値は、$f_2 \fallingdotseq 1/5$ であるが、現実的観点からは f_2 を1/5より大きくした場合に興味があると考えられたので、本解析においては、$f_2 = 1/5, 1/4, 1/3, 1/2$ の場合について計算を行なう。これらの場合の結果を比較するために、ここでは1次抽出単位の数、すなわち1次標本として得られる単位区内に含まれる世帯の数n_1が、$n_1 \fallingdotseq 3000$ （$n_1 \geq 3000$）となるように、シ

ミュレーション実験を設定した。これにより、全ての場合について、2次抽出単位の数n_2は、$n_2 \geq 600$となり、16個の変数を用いる重回帰分析を行なっても、安定性については問題がないと考える。また第4節で考察したように、家計支出額およびそれに関連する合成変数を用いなくても、比較的精度の高い推定が可能であると考えられるので、ここでは第4節で与えた16個の説明変数を用いた重回帰モデルにより推定を行なう。また回帰係数の推定には、重み付き最小2乗法を用いる。

第2段抽出率 f_2 を変化させた場合の、全地域ブロックに対する平均2乗誤差の値は、表11にまとめられている。この表より、当然のことながら、f_2 を大きくしていくと、平均2乗誤差は小さくなることが判る。従来からの方法 Y-SMP の場合は、この推定量は不偏であるから、表中の値は分散の値と同じになる。したがって f_2 と表中の値とは反比例するはずである。例えば、$f_2 = 1/4$ と $1/2$ の場合を比較すると、平均2乗誤差（分散）の値は、いずれの地域ブロックにおいても、約1/2（＝1/4÷1/2）となっている。しかし提案した4つの推定方法の場合には、このような関係が成立していない。例えば $f_2 = 1/2$ の場合の値は、$f_2 = 1/4$ の場合の値の半分にはなっていない。この事実から、これらの推定量には偏り

があり、重回帰モデルに取り入れるべき変数には、まだ改良の余地があると考えられる。したがってこの実験のように f_2 をかなり大きくしていくと、提案した方法の偏りが問題となり、従来から用いられてきた単純平均推定量の分散と比較して、改善の度合が小さくなっていく。これは、表11の1番右の列に与えた数値より理解できる。すなわち f_2 を大きくすると、3/Y の値は増加する。とくに $f_2=1/2$ の場合は、12地域中4地域において、3/Y の値が1を越えており、提案した方法の精度が単純平均推定量の精度より悪くなっている。f_2 を大きくすると提案する方法の精度が相対的に悪くなる傾向は、地域 04, 06, 09, 12 のように、標本数Nが比較的小さな所で観察される。したがってシミュレーション実験における標本抽出の仕方に問題があるのではないかと考えられる。しかしNが十分大きく、第1段の標本抽出で重複の少ない地域については、f_2 が 1/2 になっても、提案した方法は従来の方法よりかなり精度が高い。

次に表11の右から2列目の数値を見ると、提案した方法1と3による推定精度の比の値は抽出率 f_2 にあまり依存しないが、f_2 を大きくすると多少比の値が大きくなる。すなわち抽出率を大きくすると、提案した4つの方法による推定精度は、あまり差がなくなることが判る。

6. 結果のまとめ

本節では、今回の3つの研究課題について、第3節～第5節で得た結果をまとめておく。

まず第1の研究課題である集落2段抽出法を用いた場合の、全地域ブロックに対する推定については、次のような結果が得られた。

(1-1) 提案した重回帰モデルを用いる方法による推定値の精度は、従来からの推定値（各年齢階級における単純平均）の精度に比べて高い。特に重み付けをした場合には、平均2乗誤差は、20～80％程度になる。

(1-2) 提案した推定量の内では、推定域を考慮した修正最小2乗(Modified OLS)推定量による推定方法を用いたものが、平均2乗誤差の意味では最も良さそうである。特に重み付けをした場合には、推定域を標本域と同じとする従来の推定と比べて1～2割程度の改善が見られる。従って推定域を考慮した重回帰分析を行なう必要がある。

(1-3) 提案した推定方法は、全ての地域ブロックについて従来の方法より平均2乗誤差の意味で、精度がかなり高いと考えられる。この意味で、提案した方法は安定であり、実際の場での使用に耐え得るものと考えられる。

次に第2の研究課題である重回帰モデルの説明変数に対する考察については、次のような結果が得られた。

(2-1) 家計支出額およびそれに関連する合成変数を用いない場合でも、比較的精度の高い推定を行なうことが可能である。すなわち現実的意味付けの行い易い変数のみを用いて、各年齢階級における平均世帯所得を、精度よく推定することができる。

(2-2) 提案した推定方法の内では、3:M-OLSの精度がよく、重み付き最小2乗法を用いると、多少推定精度が向上する。この方法によれば、従来からの方法に比べ、推定量の平均2乗誤差を、23～85%程度にまで減少させることができる。

また第3の研究課題である第2段抽出率と推定精度の変化の関係については、以下のような結果が得られた。

(3-1) 第2段抽出率 f_2 を大きくすると、提案した推定方法と従来からの方法との差は減少する。特に第1段の抽出数が小さな場合には、注意が必要かもしれない。しかし第1段の抽出数が大きな場合には、$f_2 = 1/2$ の場合でも、提案した方法は、従来からの方法に比べて平均2乗誤差をかなり小さくする場合も多い。

(3-2) 提案した推定方法の内では、3:M-OLSの精度がよいが、f_2を大きくしていくと、4つの方法による推定精度にはあまり差がなくなる。

結論として、標本抽出方法として集落2段抽出法を用いる場合、家計支出額を除く4つの原変数に $x_5 \equiv (x_3 - 50)^2$ [x_3は世帯主年齢] および2.2節で示した11個のダミー変数を説明変数として追加した重回帰モデルに基づいて、3の方法(3:M-OLS)により推定を行なえば、かなり精度の高い安定した推定が、全地域ブロックに対して行えることが判明した。前節までの解析では、推定値の安定性という観点から、毎回の重回帰分析に際してモデル選択を行なった。しかし現実の解析における計算の簡便性およびモデルの解釈の容易さという観点からは、モデル選択を行なう必要のない方が望ましいと考えられる。そこで第4節で行なった計算の過程を検討したところ、上記全ての変数を取り入れたモデル(最大モデル)を用いても、安定した推定が行えると考えられた。したがってここで提案した推定方法を実際の場に適用する場合には、最大モデルを用いてもよいと考える。

第2段の抽出率に関しては、現実のほとんどの場では、$f_2 \leq 1/2$ ならば、提案した3の推定方法(3:M-OLS)により、精度を向上させることができると考えられる。

謝　辞

本研究における問題の所在について示唆をいただいた、東洋大学・浅井 晃教授に深く感謝致します。またデータの提供等についてお世話になった厚生省大臣官房・統計情報部管理企画課・稲垣誠一氏に厚く御礼申し上げます。また一般逆行列の計算に際しては、名古屋大学大型計算センターのライブラリープログラム SVDD（倍精度型特異値分解, 1985年版）を使用させていただきました。

参考文献

[1] 厚生省大臣官房統計情報部（1986）：「昭和61年国民生活基礎調査の概況」.

[2] 名古屋大学大型計算センター（1982）：「ライブラリー・プログラム利用の手引（数値計算編）」.

[3] 総務庁統計局統計基準部（1989）：「昭和63年度統計調査における規模区分等に関する調査研究報告書」.

[4] 田栗 正章, 井上 隆勝（1988）：国民生活基礎調査における平均所得の推定精度の改善について, 京都大学数理解析研究所講究録, No. 682.

[5] 田栗 正章, 井上 隆勝：国民生活基礎調査における平均所得の推定精度の改善について, 日本統計学会誌 (to appear).

表3．　地域ブロック０１に対する検討［n_2=600，f_2=1/5，重みなしの場合］

(1) 各方式の下での推定値のシミュレーションに亘る平均

年齢階級k	1:OLS-X	2:OLS	3:M-OLS	4:SEL.	Y-SMP.	μ_k	REG_k
1	294.	293.	291.	292.	308.	303.	294.
2	532.	532.	532.	532.	536.	532.	531.
3	641.	641.	640.	641.	628.	625.	639.
4	715.	715.	716.	715.	722.	718.	710.
5	687.	688.	687.	688.	690.	690.	680.
6	765.	764.	764.	765.	783.	751.	751.

(2) 各方式の下での推定値のシミュレーションに亘る平均２乗誤差

年齢階級k	1:OLS-X	2:OLS	3:M-OLS	4:SEL.	Y-SMP.
平均	4020.	4040.	3960.	3970.	13500.
1	1570.	1490.	1540.	1530.	1100.
2	463.	472.	446.	458.	743.
3	864.	896.	828.	849.	1020.
4	1180.	1170.	1150.	1180.	2340.
5	2610.	2690.	2490.	2530.	9190.
6	17400.	17500.	17300.	17300.	66900.

表4．　５種類の方法による推定結果とその比較
［重みなしの場合の全地域ブロックに対する平均２乗誤差］

地域	N（M）	1:OLS-X	2:OLS	3:M-OLS	4:SEL.	Y-SMP.	3/1*	3/Y**
01	6158(214)	4020.	4040.	3960.	3970.	13500.	.985	.293
02	1487(50)	1620.	1610.	1500.	1580.	2150.	.926	.698
03	2494(69)	1730.	1770.	1650.	1710.	2410.	.954	.685
04	1102(34)	2870.	2790.	2810.	2740.	2660.	.979	1.056
05	1460(42)	951.	945.	914.	934.	1240.	.961	.737
06	2040(64)	1410.	1410.	1390.	1410.	1510.	.986	.921
07	2596(74)	1130.	1100.	1060.	1070.	1580.	.938	.671
08	3633(108)	2510.	2510.	2400.	2480.	2910.	.956	.825
09	2018(68)	1390.	1390.	1360.	1390.	1570.	.978	.866
10	1474(48)	636.	628.	599.	607.	908.	.942	.660
11	3691(138)	5370.	5420.	5160.	5200.	8820.	.961	.585
12	731(27)	3750.	3870.	3650.	3790.	2060.	.973	1.772

*　：3:M-OLS の 1:OLS-X に対する相対比率．
**：3:M-OLS の Y-SMP. に対する相対比率．

表5．母集団における回帰分析(回帰誤差の評価)

k	階級	相対度数	μ_k	REG_k	規準化誤差 平均	規準化誤差 分散
1:	- 29	0.092	303.	294.	0.0225	0.156
2:	30 - 39	0.271	532.	531.	0.0023	0.343
3:	40 - 49	0.279	625.	639.	-0.0334	0.757
4:	50 - 59	0.227	718.	710.	0.0197	0.986
5:	60 - 69	0.089	690.	680.	0.0247	2.27
6:	70 -	0.042	751.	751.	-0.0008	6.14

表6．　5種類の方法による推定結果とその比較
「重み付きの場合の全地域ブロックに対する平均2乗誤差」

地域	N（M）	1:OLS-X	2:OLS	3:M-OLS	4:SEL.	Y-SMP.	3/1*	3/Y**
01	6158(214)	3380. .841	3040. .752	2740. .692	2950. .743	13500.	.811	.203
02	1487(50)	1090. .673	1030. .640	955. .637	979. .620	2150.	.876	.444
03	2494(69)	1720. .994	1440. .814	1370. .830	1400. .819	2410.	.797	.568
04	1102(34)	2150. .749	2060. .738	2020. .719	2030. .741	2660.	.940	.759
05	1460(42)	959. 1.008	985. 1.042	895. .979	928. .994	1240.	.933	.722
06	2040(64)	1080. .766	1080. .766	1060. .763	1080. .766	1510.	.981	.702
07	2596(74)	942. .834	925. .841	922. .870	936. .875	1580.	.979	.584
08	3633(108)	2450. .976	2270. .904	2220. .925	2250. .907	2910.	.906	.763
09	2018(68)	1360. .978	1330. .957	1260. .926	1270. .914	1570.	.926	.803
10	1474(48)	609. .958	574. .914	545. .910	550. .906	908.	.895	.600
11	3691(138)	3770. .702	3440. .635	3090. .599	3390. .652	8820.	.820	.350
12	731(27)	4210. 1.123	3940. 1.018	3700. 1.014	3850. 1.016	2060.	.879	1.796

* : 3:M-OLS の 1:OLS-X に対する相対比率．
** : 3:M-OLS の Y-SMP. に対する相対比率．
各セルの下段の数値は，重みなしの結果に対する相対比率を示す．

表7．母集団における重回帰分析［合成変数を用いない場合］

標本数	:	6158			
目的変数	:	総所得（y）			
重相関係数	: R	0.403			
寄与率	: R**2	0.163			
自由度調整済R	: RA	0.160			
残差の標準誤差	: ROOT-VE	439.0			
系列相関係数	: S-COR.	1:0.076	2:0.065	3:0.041	

++ 回帰係数 ++

変数名	STD. B	B	D(B)	T=B/D(B)
x_1（世帯人員）	4.308D-02	1.452D+01	6.923D+00	2.097D+00
x_2（有業人員）	1.400D-01	7.526D+01	7.437D+00	1.012D+01
x_3（世帯主年齢）	1.364D-01	5.171D+00	4.984D-01	1.038D+01
x_4（夫婦の組数）	5.047D-02	5.308D+01	1.799D+01	2.950D+00
x_5（$(x_3-50)^2$）	-4.912D-02	-1.050D-01	2.710D-02	-3.874D+00
x_6 (Dummy Var.)	2.291D-02	1.082D+02	7.077D+01	1.529D+00
x_7 (Dummy Var.)	5.526D-02	1.114D+02	5.015D+01	2.221D+00
x_8 (Dummy Var.)	-6.893D-02	-1.768D+02	5.349D+01	-3.305D+00
x_9 (Dummy Var.)	-1.717D-02	-3.180D+01	4.963D+01	-6.408D-01
x_{10}(Dummy Var.)	-5.277D-02	-1.619D+02	5.722D+01	-2.830D+00
x_{11}(Dummy Var.)	-2.406D-02	-1.024D+02	6.762D+01	-1.515D+00
x_{12}(Dummy Var.)	5.209D-02	1.049D+02	4.898D+01	2.141D+00
x_{13}(Dummy Var.)	1.409D-01	1.421D+02	4.443D+01	3.199D+00
x_{14}(Dummy Var.)	-7.297D-02	-1.194D+02	4.701D+01	-2.540D+00
x_{15}(Dummy Var.)	1.281D-02	1.431D+01	4.481D+01	3.194D-01
x_{16}(Dummy Var.)	-7.029D-02	-2.150D+02	5.583D+01	-3.850D+00
定 数 項	0.0	1.267D+02	5.446D+01	2.326D+00

++ 分散分析表 ++

要 因	自由度	平方和	平均平方	F
回 帰	16	2.29864D+08	1.43665D+07	7.45525D+01
残 差	6141	1.18339D+09	1.92703D+05	
合 計	6157	1.41326D+09		

== 各年齢階級における y の平均 ==

k	階級	度数	yの平均	回帰平面上の値
1:	-29	0.092	303.	296.
2:	30-39	0.271	532.	533.
3:	40-49	0.279	625.	631.
4:	50-59	0.227	718.	713.
5:	60-69	0.089	690.	693.
6:	70-	0.042	751.	740.

表8．母集団における重回帰分析［合成変数を用いる場合］

標本数	:	6158			
目的変数	:	総所得（y）			
重相関係数	: R	0.442			
寄与率	: R**2	0.196			
自由度調整済R	: RA	0.193			
残差の標準誤差	: ROOT-VE	430.3			
系列相関係数	: S-COR.	1:0.057	2:0.045	3:0.022	

＋＋ 回帰係数 ＋＋

変数名	STD.B	B	D(B)	T=B/D(B)
x_1 (世帯人員)	-1.731D-01	-5.834D+01	1.977D+01	-2.950D+00
x_2 (有業人員)	1.516D-01	8.151D+01	7.280D+00	1.119D+01
x_3 (家計支出額)	-1.009D-01	-3.027D+00	7.775D-01	-3.894D+00
x_4 (世帯主年齢)	1.179D-01	4.472D+00	4.894D-01	9.138D+00
x_5 (夫婦の組数)	2.788D-02	2.932D+01	1.770D+01	1.655D+00
x_6 ($\sqrt[4]{x_3}$)	4.497D-02	7.985D+01	1.360D+02	5.869D-01
x_7 ($\sqrt[4]{x_1 * x_3}$)	3.855D-01	3.242D+02	9.614D+01	3.372D+00
x_8 (Dummy Var.)	3.033D-02	1.432D+02	7.032D+01	2.037D+00
x_9 (Dummy Var.)	7.415D-02	1.494D+02	5.020D+01	2.977D+00
x_{10} (Dummy Var.)	-3.880D-02	-9.951D+01	5.356D+01	-1.857D+00
x_{11} (Dummy Var.)	1.097D-02	2.032D+01	5.000D+01	4.063D-01
x_{12} (Dummy Var.)	-2.017D-02	-6.189D+01	5.708D+01	-1.084D+00
x_{13} (Dummy Var.)	5.899D-04	2.511D+00	6.697D+01	3.749D-02
x_{14} (Dummy Var.)	3.850D-02	7.753D+01	4.800D+01	1.615D+00
x_{15} (Dummy Var.)	1.160D-01	1.170D+02	4.352D+01	2.689D+00
x_{16} (Dummy Var.)	-7.887D-02	-1.290D+02	4.603D+01	-2.803D+00
x_{17} (Dummy Var.)	-3.316D-03	-3.705D+00	4.389D+01	-8.442D-02
x_{18} (Dummy Var.)	-6.227D-02	-1.904D+02	5.477D+01	-3.477D+00
定 数 項	0.0	-6.555D+02	1.172D+02	-5.589D+00

＋＋ 分散分析表 ＋＋

要 因	自由度	平方和	平均平方	F
回 帰	18	2.76569D+08	1.53650D+07	8.29829D+01
残 差	6139	1.13669D+09	1.85158D+05	
合 計	6157	1.41326D+09		

＝＝ 各年齢階級におけるyの平均 ＝＝

k	階級	度数	yの平均	回帰平面上の値
1:	-29	0.092	303.	299.
2:	30-39	0.271	532.	531.
3:	40-49	0.279	625.	637.
4:	50-59	0.227	718.	708.
5:	60-69	0.089	690.	680.
6:	70-	0.042	751.	761.

表９．合成変数を用いる場合と用いない場合の推定結果とその比較
［重みなしの場合の全地域ブロックに対する平均２乗誤差］

地域 ﾌﾞﾛｯｸ	N	1:OLS-X	2:OLS	3:M-OLS	4:SEL.	Y-SMP.	3/1*	3/Y**
01	6158	3700.	3540.	4080.	3990.	13500.	1.103	.302
		4020.	4040.	3960.	3970.	13500.	.985	.293
		(0.920)	(0.876)	(1.030)	(1.005)			
02	1487	2000.	1960.	1700.	1890.	2150.	.850	.791
		1620.	1610.	1500.	1580.	2150.	.926	.698
		(1.234)	(1.217)	(1.133)	(1.196)			
03	2494	1990.	2000.	1740.	1860.	2410.	.874	.722
		1730.	1770.	1650.	1710.	2410.	.954	.685
		(1.150)	(1.130)	(1.055)	(1.088)			
04	1102	2140.	2130.	1930.	1960.	2660.	.902	.726
		2870.	2790.	2810.	2740.	2660.	.979	1.056
		(0.746)	(0.763)	(0.687)	(0.715)			
05	1460	955.	968.	774.	855.	1240.	.810	.624
		951.	945.	914.	934.	1240.	.961	.737
		(1.004)	(1.024)	(0.847)	(0.915)			
06	2040	1610.	1600.	1510.	1540.	1510.	.938	1.000
		1410.	1410.	1390.	1410.	1510.	.986	.921
		(1.142)	(1.135)	(1.086)	(1.092)			
07	2596	1300.	1300.	998.	1090.	1580.	.768	.632
		1130.	1100.	1060.	1070.	1580.	.938	.671
		(1.150)	(1.182)	(0.942)	(1.019)			
08	3633	2820.	2770.	2490.	2730.	2910.	.883	.856
		2510.	2510.	2400.	2480.	2910.	.956	.825
		(1.124)	(1.104)	(1.038)	(1.101)			
09	2018	1470.	1480.	1330.	1410.	1570.	.905	.847
		1390.	1390.	1360.	1390.	1570.	.978	.866
		(1.058)	(1.065)	(0.978)	(1.014)			
10	1474	681.	665.	572.	582.	908.	.840	.630
		636.	628.	599.	607.	908.	.942	.660
		(1.071)	(1.059)	(0.955)	(0.959)			
11	3691	5190.	5220.	5210.	5200.	8820.	1.004	.591
		5370.	5420.	5160.	5200.	8820.	.961	.585
		(0.966)	(0.963)	(1.010)	(1.000)			
12	731	4830.	4830.	4330.	4480.	2060.	.896	2.102
		3750.	3870.	3650.	3790.	2060.	.973	1.772
		(1.288)	(1.248)	(1.186)	(1.182)			

☆ 各セルの１段目，２段目の数値は、合成変数を用いない場合と用いる場合の結果を示す。
３段目の（ ）内の数値は、合成変数を用いない場合と用いる場合の相対比率を示す。
* ：3:M-OLS の 1:OLS-X に対する相対精度。
** ：3:M-OLS の Y-SMP に対する相対精度。

表10．合成変数を用いる場合と用いない場合の推定結果とその比較
[重み付きの場合の全地域ブロックに対する平均２乗誤差]

地域 ﾌﾞﾛｯｸ	標本数	1:OLS-X	2:OLS	3:M-OLS	4:SEL.	Y-SMP.	3/1*	3/Y**
01	6158	3470.	3510.	3100.	3210.	13500.	.893	.230
		.938	.992	.760	.805			
		3380.	3040.	2740.	2950.	13500.	.811	.203
		.841	.752	.692	.743			
		(1.027)	(1.155)	(1.131)	(1.088)			
02	1487	1270.	1250.	1090.	1140.	2150.	.858	.507
		.635	.638	.641	.603			
		1090.	1030.	955.	979.	2150.	.876	.444
		.673	.640	.637	.620			
		(1.165)	(1.214)	(1.141)	(1.164)			
03	2494	1660.	1640.	1440.	1490.	2410.	.867	.598
		.834	.820	.828	.801			
		1720.	1440.	1370.	1400.	2410.	.797	.568
		.994	.814	.830	.819			
		(0.965)	(1.139)	(1.051)	(1.064)			
04	1102	1980.	1890.	1700.	1730.	2660.	.859	.639
		.925	.887	.881	.883			
		2150.	2060.	2020.	2030.	2660.	.940	.759
		.749	.738	.719	.741			
		(0.921)	(0.917)	(0.842)	(0.852)			
05	1460	846.	851.	706.	748.	1240.	.835	.569
		.886	.879	.912	.875			
		959.	985.	895.	928.	1240.	.933	.722
		1.008	1.042	.979	.994			
		(0.882)	(0.864)	(0.789)	(0.943)			
06	2040	1370.	1330.	1260.	1260.	1510.	.920	.834
		.851	.831	.834	.818			
		1080.	1080.	1060.	1080.	1510.	.981	.702
		.766	.766	.763	.766			
		(1.269)	(1.231)	(1.189)	(1.167)			

☆ 各セルの１段目，３段目の数値は、合成変数を用いない場合と用いる場合の結果を示す。
５段目の（ ）内の数値は、合成変数を用いない場合と用いる場合の相対比率を示す。
また各セルの２段目，４段目の数値は、重みなしの結果に対する相対比率を示す。

* ：3:M-OLS の 1:OLS-X に対する相対精度。

** ：3:M-OLS の Y-SMP に対する相対精度。

表10．合成変数を用いる場合と用いない場合の推定結果とその比較（つづき）
［重み付きの場合の全地域ブロックに対する平均２乗誤差］

地域 ﾌﾞﾛｯｸ	標本数	1:OLS-X	2:OLS	3:M-OLS	4:SEL.	Y-SMP.	3/1*	3/Y**
07	2596	1040.	954.	813.	862.	1580.	.782	.515
		.800	.734	.815	.791			
		942.	925.	922.	936.	1580.	.979	.584
		.834	.841	.870	.875			
		(1.104)	(1.031)	(0.882)	(0.921)			
08	3633	2570.	2600.	2180.	2260.	2910.	.848	.749
		.911	.939	.876	.828			
		2450.	2270.	2220.	2250.	2910.	.906	.763
		.976	.904	.925	.907			
		(1.049)	(1.145)	(0.982)	(1.004)			
09	2018	1340.	1360.	1330.	1350.	1570.	.993	.847
		.912	.919	1.000	.957			
		1360.	1330.	1260.	1270.	1570.	.926	.803
		.978	.957	.926	.914			
		(0.985)	(1.023)	(1.056)	(1.063)			
10	1474	590.	565.	527.	539.	908.	.893	.580
		.866	.850	.921	.926			
		609.	574.	545.	550.	908.	.895	.600
		.958	.914	.910	.906			
		(0.969)	(0.984)	(0.967)	(0.980)			
11	3691	3880.	3890.	3350.	3500.	8820.	.863	.380
		.748	.745	.643	.673			
		3770.	3440.	3090.	3390.	8820.	.820	.350
		.702	.635	.599	.652			
		(1.029)	(1.131)	(1.084)	(1.032)			
12	731	5120.	5160.	5040.	5020.	2060.	.984	2.447
		1.060	1.068	1.164	1.121			
		4210.	3940.	3700.	3850.	2060.	.879	1.796
		1.123	1.018	1.014	1.016			
		(1.216)	(1.310)	(1.362)	(1.304)			

☆ 各セルの１段目，３段目の数値は、合成変数を用いない場合と用いる場合の結果を示す。
５段目の（ ）内の数値は、合成変数を用いない場合と用いる場合の相対比率を示す。
また各セルの２段目，４段目の数値は、重みなしの結果に対する相対比率を示す。

* : 3:M-OLS の 1:OLS-X に対する相対精度。

** : 3:M-OLS の Y-SMP に対する相対精度。

表11. 第2段の抽出率 f_2 を変化させた場合の推定結果 [n_1 = 3000]
[重み付きの場合の全地域ブロックに対する平均2乗誤差]

地域 ﾌﾞﾛｯｸ	標本数	抽出率 f_2	1:OLS-X	2:OLS	3:M-OLS	4:SEL.	Y-SMP.	3/1*	3/Y**
01	6158	1/5	3470.	3510.	3100.	3210.	13500.	.89	.23
		1/4	2350.	2440.	2140.	2150.	8620.	.91	.25
		1/3	2590.	2610.	2470.	2570.	6620.	.95	.37
		1/2	1720.	1760.	1540.	1540.	3750.	.90	.41
02	1487	1/5	1270.	1250.	1090.	1140.	2150.	.86	.51
		1/4	979.	982.	894.	891.	1430.	.91	.63
		1/3	735.	702.	688.	685.	1300.	.93	.53
		1/2	683.	683.	690.	685.	721.	1.01	.96
03	2494	1/5	1660.	1640.	1440.	1490.	2410.	.87	.60
		1/4	1350.	1410.	1300.	1330.	1930.	.96	.67
		1/3	1320.	1320.	1200.	1220.	1750.	.91	.69
		1/2	943.	931.	889.	897.	1070.	.94	.83
04	1102	1/5	1980.	1890.	1700.	1730.	2660.	.86	.64
		1/4	1870.	1850.	1800.	1800.	2240.	.96	.80
		1/3	1440.	1430.	1330.	1330.	1450.	.92	.92
		1/2	1170.	1170.	1160.	1150.	1070.	.99	1.08
05	1460	1/5	846.	851.	706.	748.	1240.	.83	.57
		1/4	528.	591.	497.	514.	959.	.94	.52
		1/3	385.	400.	381.	393.	627.	.99	.61
		1/2	308.	308.	307.	308.	533.	1.00	.58
06	2040	1/5	1370.	1330.	1260.	1260.	1510.	.92	.83
		1/4	1110.	1090.	1090.	1090.	1020.	.98	1.07
		1/3	1040.	1030.	1020.	1040.	913.	.98	1.11
		1/2	953.	945.	954.	951.	535.	1.00	1.78

* : 3:M-OLS の 1:OLS-X に対する相対精度。
** : 3:M-OLS の Y-SMP に対する相対精度。

表11. 第2段の抽出率 f_2 を変化させた場合の推定結果 [n_1＝3000] (つづき)
[重み付きの場合の全地域ブロックに対する平均2乗誤差]

地域ブロック	標本数	抽出率 f_2	1:OLS-X	2:OLS	3:M-OLS	4:SEL.	Y-SMP.	3/1*	3/Y**
07	2596	1/5	1040.	954.	813.	862.	1580.	.78	.51
		1/4	822.	765.	643.	680.	1140.	.78	.56
		1/3	568.	558.	500.	504.	836.	.88	.60
		1/2	524.	493.	469.	481.	568.	.90	.83
08	3633	1/5	2570.	2600.	2180.	2260.	2910.	.85	.75
		1/4	2170.	2280.	1920.	2030.	2900.	.88	.66
		1/3	1830.	1800.	1630.	1690.	1870.	.89	.87
		1/2	1300.	1290.	1250.	1260.	1330.	.96	.94
09	2018	1/5	1340.	1360.	1330.	1350.	1570.	.99	.85
		1/4	1170.	1160.	1150.	1160.	1210.	.98	.95
		1/3	1110.	1110.	1100.	1110.	940.	.99	1.17
		1/2	982.	981.	977.	977.	583.	.99	1.68
10	1474	1/5	590.	565.	527.	539.	908.	.89	.58
		1/4	443.	424.	390.	397.	695.	.88	.56
		1/3	416.	402.	374.	378.	583.	.90	.64
		1/2	244.	244.	239.	239.	328.	.98	.73
11	3691	1/5	3880.	3890.	3350.	3500.	8820.	.86	.38
		1/4	3770.	3860.	3360.	3440.	6770.	.89	.50
		1/3	2930.	2900.	2580.	2750.	4190.	.88	.62
		1/2	2130.	2130.	1930.	2080.	2600.	.91	.74
12	731	1/5	5120.	5160.	5040.	5020.	2060.	.98	2.45
		1/4	5040.	5030.	4940.	4960.	1720.	.98	2.87
		1/3	4710.	4710.	4670.	4680.	1580.	.99	2.96
		1/2	4760.	4760.	4740.	4750.	963.	1.00	4.92

* : 3:M-OLS の 1:OLS-X に対する相対精度。
** : 3:M-OLS の Y-SMP に対する相対精度。