

世帯調査または事業所・企業調査における標本の交代について

美添泰人（青山学院大学経営学部）

2019年3月10日

要約 調査を継続的に実施する場合、総務省の世帯調査では標本の部分的な交代 (rotation sampling) が広く行われている。一方、日本の事業所および企業調査では、導入している例は最近まで少なかった。以下は各節の要約である。

- (1) 前期との差を求める場合、標本を交代するより継続する方が標本誤差は小さい。
- (2) 標本の部分的な交代を導入する場合、前期との差の推定に関しては、継続標本のみを利用する方法と、すべての標本を利用する方法があり、世帯調査では両者には大きな差はない。
- (3) 現実には標本の脱落 (attrition) が発生するため、継続標本のみを利用する方法には偏りが発生しやすい。
- (4) 標本調査では母集団の名簿の正確さが前提となる。世帯調査の場合は比較的正確な名簿が月次または年次で更新されているため rotation sampling は有効である。他方、年次決算を導入している企業に関しては正確な名簿整備は1年に一度しかできない。また、事業所に関しては、経済センサスによって新設事業所が補足される2年または3年の間に次第に名簿情報が劣化する。そのため、新たな名簿を用いて抽出された標本と交代する時期に「断層」が発生する。これは母集団名簿の特性であり、現時点では、rotation sampling を利用しても断層の縮小に関しては大きな効果は期待できない。
- (5) 事業所や企業に関する調査の正確性を向上させるためには、法人番号や雇用保険適用事業所の名簿情報を継続的に利用できるような仕組みが必要である。

1 基本的なモデル

目的は所得や消費などについて、母集団における前期の値 X_0 と今期の値 X_1 との差（ないし比）を推定することである。まず、継続標本と固定標本の問題を簡単なモデルに基づいて考える。比較的同質な世帯を想定して、次の仮定をおく。

- (A.1) 標本サイズを n とする。母集団サイズ N は非常に大きく、有限母集団修正は不要とする。
- (A.2) 母集団の名簿は、いつの時点でも、おおむね正確である。
- (A.3) 無回答の存在は無視できる程度に小さい。または継続標本と交代標本の間で違いはない。
- (A.4) 継続標本 (i 番目) における前期と今期の値 (x_{0i}, x_{1i}) について、以下を仮定する。

$$E(x_{0i}) = X_0, \quad E(x_{1i}) = X_1, \quad V(x_{0i}) = V(x_{1i}) = S^2, \quad \text{cor}(x_{0i}, x_{1i}) = R$$

異なる世帯 ($i \neq j$) について (x_{0i}, x_{1i}) と (x_{0j}, x_{1j}) は独立である。

- (A.5) 交代標本の前期と今期の値 (y_{0j}, y_{1j}) については、以下を仮定する。

$$E(y_{0j}) = X_0, \quad E(y_{1j}) = X_1, \quad V(y_{0j}) = V(y_{1j}) = S^2$$

y_{0j} と y_{1j} は異なる世帯の観測値で、独立である。

以上の想定の下では、調査客体の負担や慣れによる記入精度の低下ないし一般の世帯との格差が発生しない限り、継続標本の方が誤差が少ないことが示される。標本をすべて継続する場合と、すべて交代する場合の推定量を、それぞれ

$$\Delta = \frac{1}{n} \sum_{i=1}^n (x_{1i} - x_{0i}), \quad \Delta' = \frac{1}{n} \sum_{j=1}^n (y_{1j} - y_{0j})$$

とする。これらはいずれも不偏、 $E(\Delta) = E(\Delta') = X_1 - X_0$ であり、分散は次のとおりとなる。

$$V(\Delta) = \frac{1}{n} [V(x_{1i}) + V(x_{0i}) - 2\text{cov}(x_{0i}, x_{1i})] = \frac{2}{n} S^2 (1 - R)$$

$$V(\Delta') = \frac{1}{n} [V(y_{1j}) + V(y_{0j})] = \frac{2}{n} S^2$$

多くの問題では $R > 0$ が想定されるから $V(\Delta) < V(\Delta')$ である。継続標本のサイズを $n' = (1 - R)n$ と小さくすれば、 $V(\Delta) = V(\Delta')$ となる。

2 一部の標本を継続する場合の比較

標本はいずれは交代する必要があること、およびその他の理由から、多くの継続的な世帯調査では rotation sampling が導入されている。継続標本の比率を p とし、標本 n のうち継続する標本のサイズを $n' = pn$ 、新規（交代）標本のサイズを $n'' = (1-p)n = n - n'$ とする。この状況で、改めて継続標本のみを利用する推定量 Δ と、交代標本を含めた推定量 Δ' を次のように定義する。

$$\Delta = \frac{1}{n'} \sum_{i=1}^{n'} (x_{1i} - x_{0i}), \quad \Delta' = \frac{n'}{n} \Delta + \frac{1}{n} \sum_{j=1}^{n''} (y_{1j} - y_{0j})$$

これらはいずれも不偏で、分散は次のようになる。

$$V(\Delta) = \frac{2S^2}{n'}(1-R) = \frac{2S^2}{n} \frac{1-R}{p} \quad (1)$$

$$\begin{aligned} V(\Delta') &= \left(\frac{n'}{n}\right)^2 \frac{2S^2}{n'}(1-R) + \frac{n''}{n^2} 2S^2 = \frac{2S^2}{n} \left[\frac{n'}{n}(1-R) + \frac{n''}{n}\right] \\ &= \frac{2S^2}{n} [p(1-R) + (1-p)] = \frac{2S^2}{n} (1-pR) \end{aligned} \quad (2)$$

(1) 式と (2) 式の大小関係は p と R のみに依存する。総務省の家計調査の場合は $p = 5/6$ が継続だから、分散の比 $V(\Delta)/V(\Delta') = (1-R)/p(1-pR)$ は表 1 のとおりである。

表 1 $p = 5/6$ と R に対する分散の比

R	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$V(\Delta)/V(\Delta')$	1.20	1.18	1.15	1.12	1.08	1.03	0.96	0.86	0.72	0.48	0.00

筆者の経験では家計調査における前月と当月の消費支出などの相関係数は $R = 0.2$ から 0.5 程度であり、実際に利用されている推定量 Δ' と、継続標本のみを用いた Δ では分散はほとんど変わらない。家計調査の分析例は [5] にある。また [4] には全国消費実態調査の分析が収録されている。

3 脱落標本の特性

世帯調査の場合には、脱落は、子育て、共働き、不在が多い若年単身者世帯、高齢者世帯などで調査協力が次第に困難になる、などの要因で発生する。これらは、最初の調査開始時点においても非協力の原因となものであるが、標本に偏りを生じさせるため、rotation sampling は有効な解決策である。

事業所や企業の調査では、調査期間中に業績が悪化すると廃業にならないまでも調査に協力する余裕がなくなる。財務省の法人企業統計季報における 1 社あたり固定資産の変動を見ると、図 1-(1) の全規模では毎年第 1 四半期 (4-6 月期) に落ち込みがあり、いわゆる「断層」が見える。1981 年から 1990 にかけては断層は明確に見えないが、このバブルの時期は例外的で、その後は断層がある。新しい名簿から法人が抽出されるまでの 1 年間には比較的業績の好調な企業が回答を継続するために、翌年 4 月からの新規法人企業との間で断層が生じるのである。大規模法人のみを対象に描いた図 1-(2) ではそのような断層が小さいことから、断層の原因の一部は説明できる。法人企業統計に関しては、参考文献 [2] において舟岡から一つの解決策が提示されていたが、財務情報に関しては名簿以外の原因もあるため、最終的な結論は得られていない。断層に関する他の原因および対策については参考文献 [6, 7, 8] に記述があり、5 節で簡単に触れる。

また毎月勤労統計や工業統計でも断層が発生することは、すでに [3] で指摘しているとおり、周知の現象である。

同様な例に、日本銀行の全国企業短期経済観測における中小企業の動きがある。対象企業に関する個別情報は公開されていないが、筆者が統計調査の負担に関して聞き取りを行った企業の多くは「日本銀行の調査には優先的に回答しています」と協力的である。したがって回答率は非常に高いが、継続的に回答できている中小企業の業績は平均的な企業より好調であり、エコノミストもこの傾向を指摘している。

標本の脱落に対応するために、継続して回答した標本のみを用いて前期との差を評価することがあるが、一般に、この方法には偏りがあり、しかも、偏りの大きさを客観的に評価することは困難である。ひとつの例として、

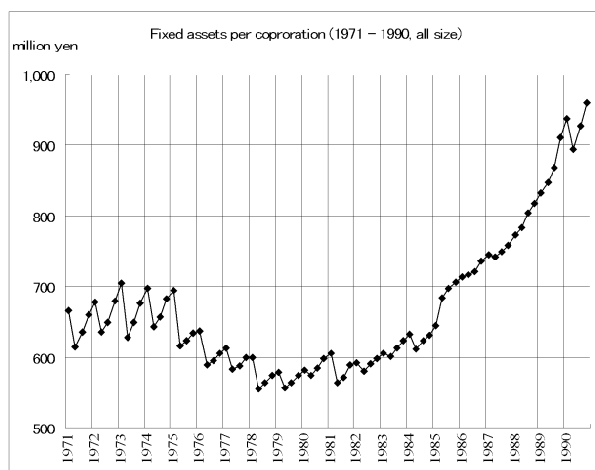


図 1-(1) 全規模

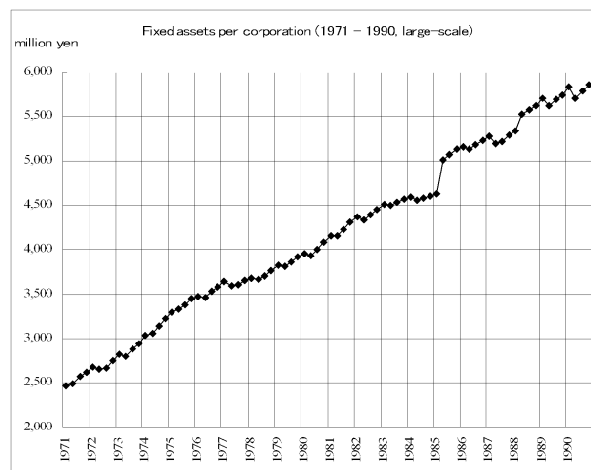


図 1-(2) 大規模

図 1: 1 社あたり固定資産（法人企業統計季報，1971-1990）

経済産業省の「商業動態統計」では、毎月の商品販売額などを調査しているが、百貨店、スーパーについては店舗調整（現在の名称は「既存店」として調査月と前年同月でも存在した事業所の数値を比較している。調整前は前月と今月の回答事業所の販売額をそれぞれ推定して比較するものであり、エコノミストの一部からは調整後の方が信頼できるという主張があったが、問題もある。たとえば大きな百貨店が開業して、その周囲にある既存店の販売額が減少したときに、連続して回答した事業所の販売額は減少するが、小売業全体では増加していたと判断できる根拠もあった。この例のように、複数の結果を提供することが多様な利用のためには望ましい。

4 事業所または企業調査の場合

世帯調査との決定的な違いは、新設事業所に関する最新の名簿情報が入手できないことにある。従来から、旧通商産業省の実施する工業統計調査における事業所数は、3年周期（現在は5年の間に2年、3年という変動）で断層が生じることが知られている（[3] 参照）。これは名簿情報を整備する事業所統計調査（後に事業所企業統計調査、現在は経済センサス）が3年周期で実施されていて、その間に名簿情報が古くなるからである。事業所全体では3年間の開業・廃業は約15%であったが、規模、業種によってははるかに大きな変動がある。筆者が関わった通商産業省の商業調査では、平成6年から平成9年にかけて3年間で、神奈川県の小売業の事業所は約6割が開業・廃業し、名簿情報が大きく変動した例もある（この分析結果は一般には公開されていない）。

事業所が調査対象として抽出され、3年間継続的に調査するうちに業績の悪い事業所は回答を停止するため、調査を継続するのは比較的業績のよい事業所となる。この状況は新しい名簿に基づく抽出が実行できる3年後まで続き、その時点で事業所あたりの支払賃金総額などに断層が発生する。

法人企業あたりの計数についても、同様な理由で1年周期の断層が発生する。財務省の法人企業統計は四半期調査と年次調査からなるが、毎年、新しい名簿が利用可能となる4月から翌年の3月にかけて業績不振の企業が調査から脱落する傾向があり、四半期調査では1年周期の断層が生じることが知られている。したがって世帯調査の場合に成立した(A-2)の想定は、日本の企業・事業所調査に関しては正しくない。

事業所統計調査は昭和23年から昭和56年までは3年ごとに調査され、その期間の工業統計における工場数には明確な「3年周期」が見えた。しかし断層の原因の大部分は零細事業所であり、出荷額の推計上は大きな問題とはならない。このことは[3]で指摘している。昭和56年以降の工業統計では周期が明確に見えないが、その理由は、統計審議会の『統計行政に関する中長期構想』[1]を受けて、昭和56年からは5年ごとに本調査を実施し、その2年後（中間年）に簡易調査を実施することになったためである。ところで、この変更にも関わらず、毎月勤労統計の標本の交代はしばらくの間は3年周期のままとされていたため、必ずしも最新時点の名簿ではなかった。その時期の断層は見えにくくなっているが、精度が高くなったわけではない。

事業所統計調査は、平成8年の調査から企業の実態把握を充実させて「事業所・企業統計調査」と名称を変更

し、さらに、現在では「経済センサス」に発展しているが、調査の周期は5年（正確には基礎調査と活動調査があるので2年と3年）であることに変わりはない。

5 断層問題の解決策

一部の論者は、世帯調査で行われている標本の入れ替え（rotation sampling）を導入すれば、断層が解消されると主張するが、正しいとは言えない。業績の悪い事業所や企業が退出する傾向は、最初の抽出に用いた母集団名簿にある「すべての」事業所でも生じている。したがって、同じ名簿から「存続している」事業所を抽出して交代しても、基本的な傾向は標本における継続事業所と同じである。標本の交代に伴う誤差の増大に隠れて名簿更新時の断層が小さく見える可能性はあるが、この方法によっては、新しい事業所名簿が整備された時点で抽出される事業所との断層は解消できない。

断層問題の本質的な解決策は、年次または月次で名簿情報を更新することである。法人企業については法人番号の利用が考えられ、さらに法人に関する税務情報が利用可能であれば、企業統計の断層問題は「部分的に」解決できる可能性がある。しかし財務省の法人企業統計に関しては、これだけでは完全な解決ではない。断層のもう一つの原因は、多くの法人が3月を決算期としているため四半期ごとの仮決算と期末の決算で性格が異なることにある。結論として1年周期でしか正確な名簿を作成できない（[6, 8] 参照）。以上から、1年以内の標本の入れ替えは「法人企業統計の断層問題に関しては」有効な解決策とはならない。

一般の事業所については法人番号は利用できないが、代替的で有効な方法として、雇用保険適用事業所の名簿情報の利用が考えられる。月次でこの情報が利用可能であれば、常に最新の母集団名簿を整備することになり、rotation sampling の効果は大きいと予想される。実際、アメリカの賃金統計である Current Employment Statistics (CES) では失業保険加入事業所の名簿から約15企業、60万事業所が対象として抽出されているため、母集団名簿の不備を原因とする断層は大きくないと考えられる。なお、CESでは標本の脱落に対応する修正方法として“Weighted Difference-Link and Taper”という手法が提案されているが、母集団情報が劣化しないことを想定しているため、この方法は日本の場合には効果が期待できない。

参考文献

- [1] 統計審議会『統計行政の中・長期構想について』昭和60年(1985年)10月答申
- [2] 中村隆英・腰原久雄・舟岡史雄・美添泰人他（共著）『法人企業統計の高度利用に関する調査研究』、社会工学研究所、1976.
- [3] 中村隆英・新家健精・美添泰人・豊田敬『経済統計入門』東京大学出版会、1983、(第2版1992).
- [4] 美添泰人・松田芳郎・舟岡史雄・大瀧雅之・清水誠（共編著）『社会科学研究』経済統計特集号（第53巻第5号）、東京大学社会科学研究所、2002.
- [5] 美添泰人・荒木万寿夫「1980年代以降における家計の消費と資産に関する実証分析」、総務省統計研修所リサーチペーパー、2006.
- [6] Yasuto Yoshizoe, Seisho Sato, Itsuko Takemura, Yoshiaki Hosoya, and Yasumasa Baba, “Correcting Non-sampling Errors in *Financial Statement Statistics* of Japanese Ministry of Finance,” 56th meeting of the International Statistical Institute, Lisbon, *Proceedings of the 56th Meeting of the International Statistical Institute*, 2007.
- [7] Yasuto Yoshizoe, “Economic Statistics,” in *International Encyclopedia of Statistical Science*, (Part 5, pp. 417–421), Springer, Miodrag Lovric (Editor-in-chief), 2010
- [8] Yasuto Yoshizoe, Masuo araki and Hitoshi Motoyama, “Extensive Use of Official Statistics,” *Aoyama Business Review*, Aoyama Institute of Global Business, Aoyama Gakuin University, Vol.40, pp.1–60, 2018.