資料２

現時点における計測方法の候補

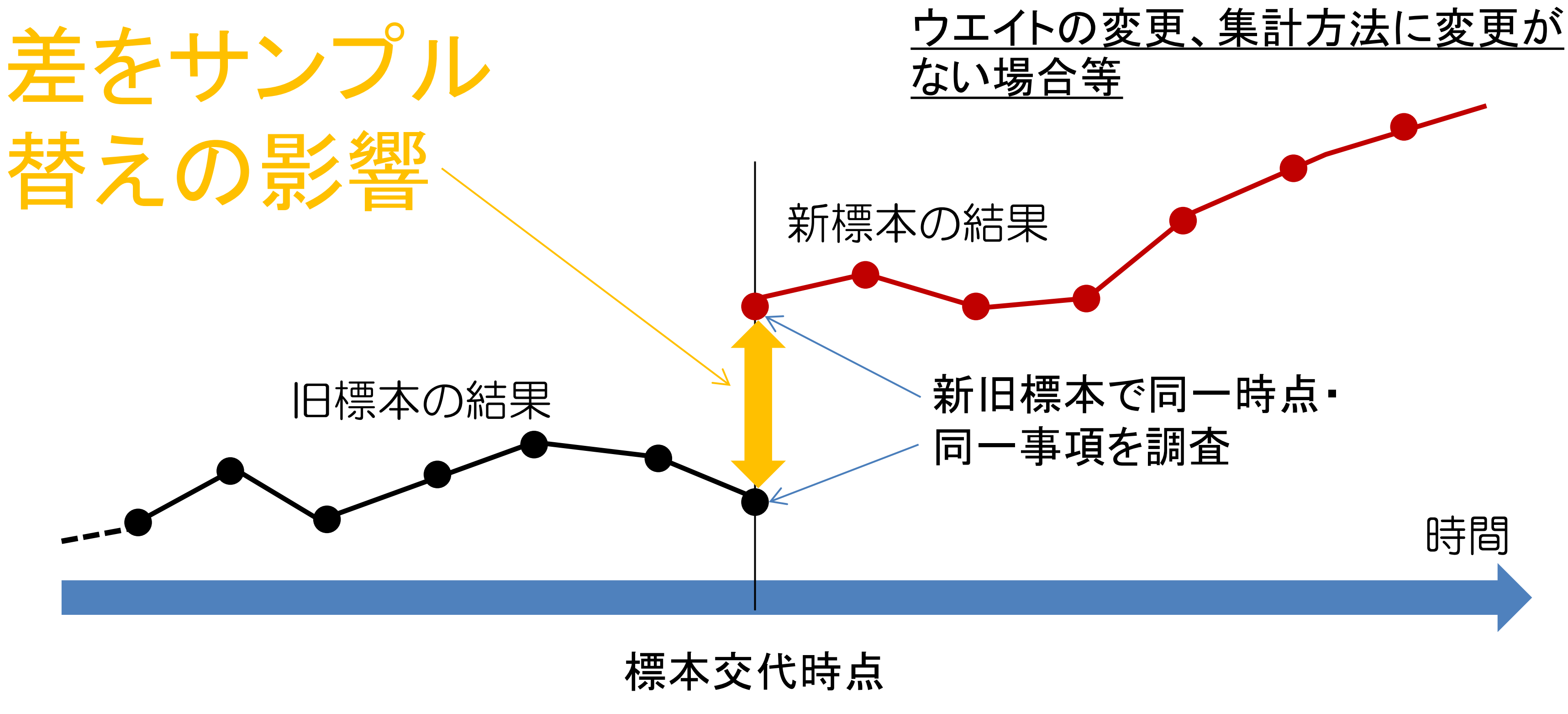
～より分かりやすい説明～

1．新標本と旧標本の双方において、同一時点・同一事項を調査している場合、新旧　双方の標本で同一時点の平均等の統計量を計算し、その差異をサンプル替えの影響とする。

＜例＞

○毎月勤労統計調査では、標本交代の際、新旧標本の双方で１月の　状態を調査

○法人企業統計では、ストックについて、旧標本で期末の状態、新標本で期首の状態を調査



2．新標本と旧標本の差の算式を変形し、要因分解の形式にしてサンプル変更要因と解釈可能な特定の項をサンプリング変更の影響とみなす方法

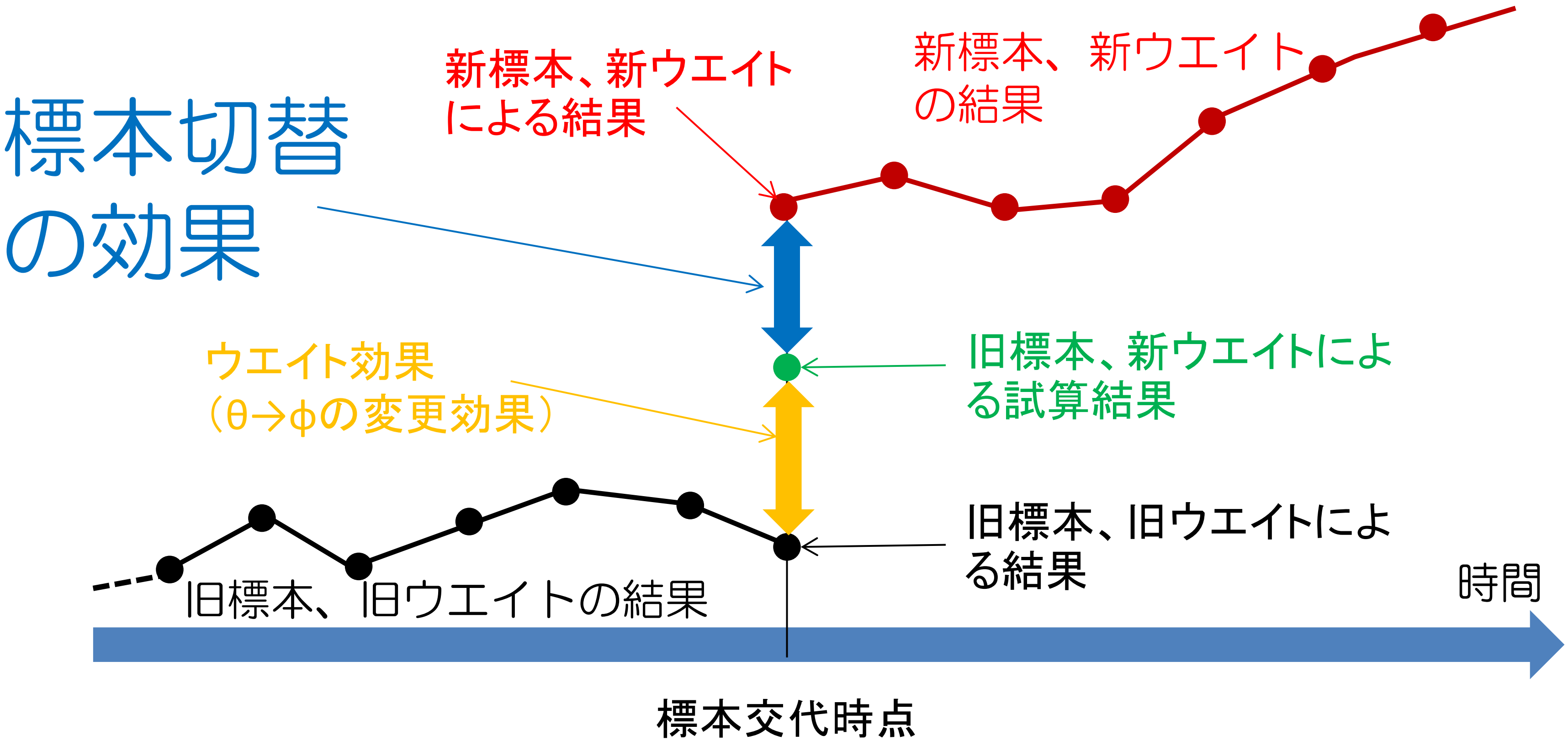
＜例＞

x：新標本　y：旧標本、θ、φ：新旧ウエイト

統計量の変動 $= f(x,\theta) - f(y,\phi) = f(x,\theta) - f(y,\theta) + f(y,\theta) - f(y,\phi)$

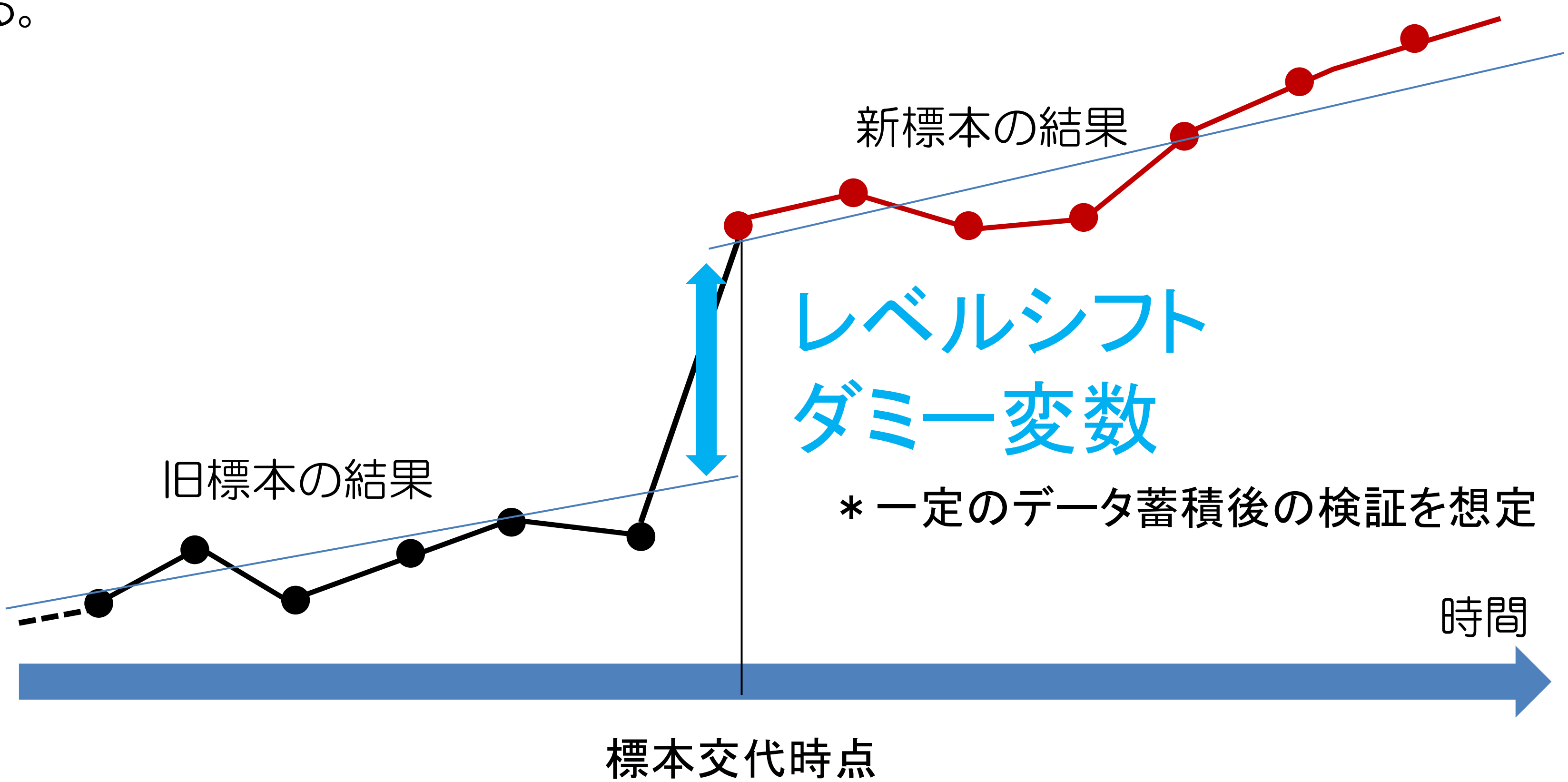
$$= \underbrace{[f(x,\theta) - f(y,\theta)]}_{\text{標本変動要因項}} + \underbrace{[f(y,\theta) - f(y,\phi)]}_{\text{ウエイト変動要因項}}$$

※　前回資料４では「パラメータ等」としていましたが、ここでは一例として指数の基準改定などで見られる、ウエイト効果の分解などの手法の応用を想定しています。



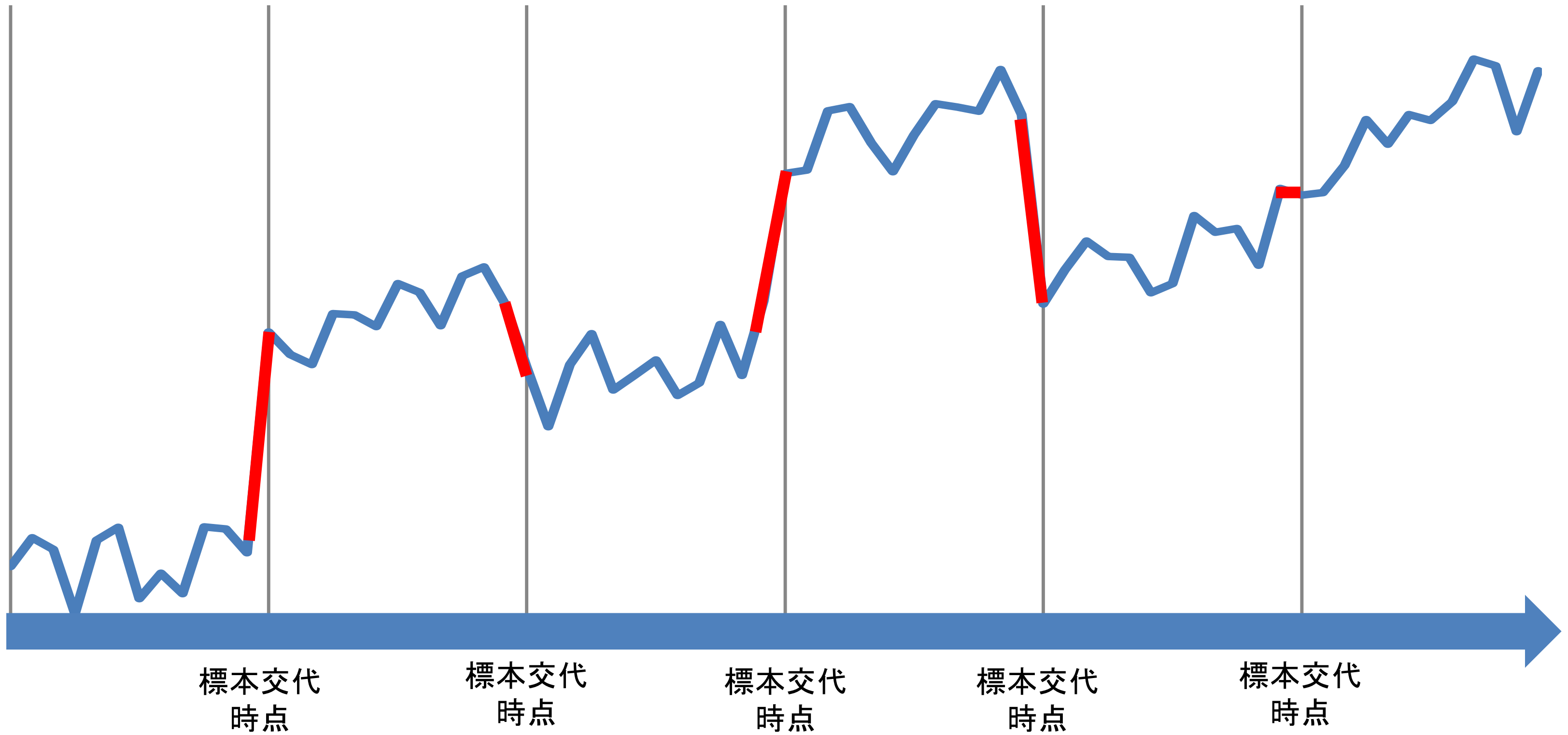
３．標本交代により、売上高等の推定結果には標本交代前と標本交代後でレベルシフトが発生すると想定し、時系列解析ソフト（X−12等）で、標本交替時におけるレベルシフトダミー要素を検出する方法

例えば、一定の統計モデルを仮定し、標本交代時点でレベルシフトの発生があったとする仮説をAIC基準等で判断。当該基準で仮説（レベルシフトダミー変数の存在）が妥当と判断される場合、当該ダミー変数の係数を標本交代の要因として検出する。



４．①標本交替時に生じる変動（前期比、前年同月比）と、②標本交代を行なわない時の変動（同）の絶対値を比較すると、①の変動の方が大きいことが想定される。そのため、①の標本交代時の前期比・前年同期比と②の標本交代を行なわない時の前期比・前年同期比の大きさを長期間のデータよって比較し、標本交替時の平均的な影響の大きさ（①から②を引いた値等）で検出する方法

※　標本交代時（赤「━」）の変動の大きさと、それ以外（青「━」）の変動の大きさを長期にわたって情報収集し、双方の変動の大きさを平均化することで標本交代時の変動の大きさの特徴を検出する。



5．シミュレーションを行いラフに影響を計算する方法

ア）N個の2時点（t,t+1）の継続標本データ
　を用意する。

x(1,t)　x(1,t+1)
x(2,t)　x(2,t+1)
x(3,t)　x(3,t+1)
x(4,t)　x(4,t+1)
　　：　　　　：
x(n,t)　x(n,t+1)

イ）「ア）」のデータからｎ個の継続標本データ
　をランダムに複数回取り出し、その都度平均を
　計算する

x(i1,t)　x(i1,t+1)
x(i2,t)　x(i2,t+1)
x(i3,t)　x(i3,t+1)
x(i4,t)　x(i4,t+1)
　　：　　　　：
x(in,t)　x(in,t+1)

\-----------------------------

u(t)　　u(t+1)・・・・抽出の都度、平均を計算

ウ）「ア）」のデータからｔ期、t+1期それぞれ
　からｎ個のデータをランダムに複数回取り出し、
　その都度平均を計算する。

x(j1,t)　x(k1,t+1)
x(j2,t)　x(k2,t+1)
x(j3,t)　x(k3,t+1)
x(j4,t)　x(k4,t+1)
　　：　　　　：
x(jn,t)　x(kn,t+1)

\-----------------------------

s(t)　　s(t+1)・・・・抽出の都度、平均を計算

エ）「イ）」で複数回計算される「u(t+1)－u(t)」
　と「ウ）」で複数回計算される「s(t+1)－s(t)」
　の標準偏差の差を調べる。

※　継続サンプルから擬似母集団を作成し、そこからリサンプリングを繰り返して行い統計量を複数回計算することで、抽出の方法の違い（継続標本を重視するサンプリング、継続標本を考えないサンプリングの違い）が結果の変動にどの程度影響するか実験的に計測する。

以下の研究などが参考となる。

統計数理（2009）第57 巻第2 号413-424　2009 統計数理研究所

法人企業統計調査における推計方法の比較
ー疑似母集団に基づく実験ー

土屋隆裕・吉岡完治・松田芳郎

＜要旨＞
法人企業統計調査において，計数値の総計とその成長率を推定するいくつかの方法を，疑似母集団を使ったシミュレーションにより比較した．従来の方法は，原則として標本全体を毎年交替し，各年の計数値の総計を求めた上で成長率を推定する方法である．これに対し標本の半分を順次交替していく標本ローテーションを行うと，成長率の推定量は，従来の方法に比べ標準誤差が2/3 程度となることが示された．さらに，標本は全て交替するとしても，標本からまず成長率を推定した上で総計を推定する方が，総計の標準誤差については従来の1/10 から1/3，成長率の標準誤差については1/10 程度になることが示された．

http://www.ism.ac.jp/editsec/toukei/pdf/57-2-413.pdf

（資料２の参考）

統計数理（2009）
第 57 巻 第 2 号 413–424

特集「ミクロ経済データによる統計解析」
［研究ノート］

法人企業統計調査における推計方法の比較

—疑似母集団に基づく実験—

土屋 隆裕[1] ・吉岡 完治[2] ・松田 芳郎[3]

（受付 2009 年 3 月 23 日；改訂 5 月 13 日；採択 5 月 14 日）

要　旨

法人企業統計調査において，計数値の総計とその成長率を推定するいくつかの方法を，疑似母集団を使ったシミュレーションにより比較した．従来の方法は，原則として標本全体を毎年交替し，各年の計数値の総計を求めた上で成長率を推定する方法である．これに対し標本の半分を順次交替していく標本ローテーションを行うと，成長率の推定量は，従来の方法に比べ標準誤差が 2/3 程度となることが示された．さらに，標本は全て交替するとしても，標本からまず成長率を推定した上で総計を推定する方が，総計の標準誤差については従来の 1/10 から 1/3，成長率の標準誤差については 1/10 程度になることが示された．

キーワード： 標本ローテーション，成長率．

1.　本稿の目的

1.1　はじめに

法人企業統計は，指定統計として我が国の法人企業に関する調査統計の中心を占めてきた．平成 20 年度からは金融・保険業を追加することによって産業の網羅性がさらに高まったこと，企業規模を大から小までとるカバレッジの高さもこの統計の重要性を裏づけている．そのため景気の状況，産業構造の変化などをこの統計によって見る向きが多く，内閣府の発表する GDP 統計の基礎統計の一つとしても重要な役割を果たしている．しかしまたこの統計が重要なるが故に，標本取り替えにまつわるいわゆる時系列値の変化の上ブレ，下ブレ，時系列の断層が長年指摘されてきたことも事実である(社会工学研究所，1976)．サンプリングバイアスに絡むこの課題は，おそらく一朝一夕に解決できるものではなく，調査統計全般にかかわる根深いものと言えるかもしれない．

我々は，この法人企業統計の個票に基づいて，いろいろな角度からこの断層問題に取り組んでみたい．本論文はその第一段階として，とりあえずひな形の疑似的母集団を作成して，いくつかの推定方法を比較してみることにする．特に平成 21 年度からは，標本抽出の方法としていわゆるローテーション抽出法が採用されようとしている．そこで吟味する方法の中にこのローテーション抽出法も含めることとしたい．

[1] 統計数理研究所：〒190–8562 東京都立川市緑町 10–3
[2] 慶應義塾大学 産業研究所：〒108–8345 東京都港区三田 2–15–45
[3] 青森公立大学 経営経済学部：〒030–0196 青森県青森市大字合子沢字山崎 153–4

1.2 本稿の目的

法人企業統計調査には，年度の決算計数を調べる年次別調査と，四半期ごとの仮決算計数を調べる四半期別調査とがある．本稿では季節変動などノイズが少ない年次別調査に焦点を絞ることにする．

母集団 U に含まれる法人企業 i の t 期の計数値を $x_{i,t}$ とする．またその総計を $X_t = \sum_{i \in U} x_{i,t}$ とする．このとき調査の目的は大きく二つある．一つは t 期の総計 X_t を推定することであり，もう一つはその成長率

$$G_{t-1,t} = \log X_t - \log X_{t-1} \tag{1.1}$$

を推定することである．もちろん X_t と $G_{t-1,t}$ の両者は密接に関連しており，最終的には両者を推定することになるとしても，調査の重心をどちらに置くのかに応じて調査の設計や推定方法は自ずと異なってくる (Duncan and Kalton, 1987)．本稿では調査の重心を X_t の推定に置く方法を“従来の方法”と呼び，$G_{t-1,t}$ の推定に置く方法を“指数型の方法”と呼ぶ．

従来の方法による現行の法人企業統計調査では，標本を毎年抽出し直し，t 期の標本 S_t によって X_t を推定する．成長率は，この $\hat{X}_t$ と前年の標本 S_{t-1} による推定値 $\hat{X}_{t-1}$ を用いて

$$\hat{G}_{t-1,t} = \log \hat{X}_t - \log \hat{X}_{t-1} = \log\left(\sum_{i \in S_t} w_i x_{i,t}\right) - \log\left(\sum_{i \in S_{t-1}} w_i x_{i,t-1}\right) \tag{1.2}$$

と推定する．ただし w_i は標本抽出デザイン等を反映した推定用のウェイトである．あくまでも総計 X_t の推定が第一の目的であり，成長率 $G_{t-1,t}$ の推定は二義的である．さらに現行の標本設計では，企業の負担を考慮し，t 期の標本 S_t は原則として $t-1$ 期で標本とならなかった企業から選んでいる．二期間で標本企業が重複しないため，$\hat{G}_{t-1,t}$ は大きな標本誤差を持つとともに，誤差は負の系列相関を持つと考えられる．

一方，指数型の方法であれば，t 期の標本 S_t に $t-1$ 期の計数値 $x_{i,t-1}$ と t 期の計数値 $x_{i,t}$ の両方を回答してもらい，そこからまず後述の方法で成長率 $G_{t-1,t}$ を推定する．次に総計 X_t を推定するには，$t=0$ 期に全数調査などを行い基準値を得ておく．これをベンチマークと呼ぶ．t 期の総計 X_t は，$t-1$ 期の総計の推定値 $\hat{X}_{t-1}$ に，成長率の推定値 $\exp(\hat{G}_{t-1,t})$ を乗じることで推定する．同一の標本を用いて $\hat{G}_{t-1,t}$ を求めるので，その標本誤差は従来の方法より小さいと期待できる (Cochran, 1977)．反面，前期の総計 $\hat{X}_{t-1}$ が大きな誤差を含むと，$\hat{X}_t$ もその誤差を継承してしまう．またベンチマークとなる値が必要であるし，各法人企業にとっては二期分の計数値を回答してもらうため負担が増える．なお，この指数型の方法を採用している我が国の代表的な統計調査としては，経済産業省の商業動態統計調査がある．

従来の方法と指数型の方法の折衷案としては，従来の方法において標本を毎期全て入れ替えるのではなく，その一部だけを順に入れ替えていく標本ローテーションがある (Patterson, 1950; Eckler, 1955; Cochran, 1977; Sunter, 1977; Kish, 1987)．一部分は重複した標本によって X_{t-1} と X_t を推定するため，毎回標本を抽出し直すよりも $\hat{G}_{t-1,t}$ の標本誤差は縮小すると期待できる．ただし各企業にとっては，数期にわたって調査が続くこととなり，現行の設計よりも負担は重くなる．

調査を設計する際には，企業の負担とそれによって得られる精度とのバランスを見極める必要がある．理論的には，成長率 $G_{t-1,t}$ の推定が目的であれば，従来の方法よりも指数型の方法の方が精度の面で優れていることは古くから知られている．また後述する $\hat{G}_{t-1,t}$ の標本ローテーションの下での分散の理論式は，土屋・吉岡 (2008) が既に導いている．しかし $\hat{X}_t$ に関してはどちらの方法がより優れているのか直ちに明らかではない．また現実の調査で採用する手法を決定するには，シミュレーションなどを通じて各手法の性質に関して経験を積み重ねてお

く必要がある．そこで本稿ではシミュレーションを通じて，両型の得失と標本ローテーションの効果を明らかにする．特に指数型の方法と，標本ローテーションを取り入れた従来の方法は，いずれも現行の設計よりも企業の負担が増すことになる．それによって得られる精度の向上がどの程度なのか比較検討する．シミュレーションとはいえ現実の調査データを基にするので，仮に調査の設計を変更したとき，推定値がどの程度改善するのか有益な示唆が得られよう．

2. シミュレーションの方法

2.1 シミュレーションに用いるデータ

1996 年度から 2004 年度までの年次別法人企業統計調査を用いる．この 9 回全てで回収できた法人を疑似母集団 U とする．ただし 9 回全ての調査で資本金が 10 億円以上であった法人に限定し，法人のパネル化は法人番号により行う．また法人番号が同じでも，企業の分割により大きな断層が生じている 1 社を除く．疑似母集団法人数は $N=3,456$ となる．

資本金を 10 億円以上に限定したのは，法人のパネル化を行うためである．現実の法人企業統計調査では資本金 10 億円未満を標本抽出としているが，標本抽出を行っているがゆえにパネル化は不可能である．また実際には業種による層化を行っているが，今回のシミュレーションでは層化抽出は行わない．企業の業種変更に伴う推定方法の問題とも絡み，問題が複雑となってしまうからである．

対象項目は，当期資産合計，当期資本金，売上高，売上原価，販売費，営業外収益，営業外費用，従業員数，従業員給与とする．これらの項目は法人企業の実態ある活動を金額などで評価した数値で，負の値は生じないからである．他方，差で求められる営業利益や当期純利益などは企業活動の結果を評価する数値で，負の値をとりうる．これらの項目の成長率を直接推定することは不自然である．そこで売上高や費用等の総計を推定し，それらを基に推定値を求めることとする．

2.2 標本抽出と推定の方法

本稿では，従来の方法では三種類の標本抽出法を試み，指数型の方法では二つの推定の方法を試みる．つまり全部で五つの方法を比較する．この五つは本節の最後に簡潔にまとめてある．

まず従来の方法である．t 期の標本を S_t とすると，S_t の抽出法として以下の三種類を考える．第一の抽出法では，母集団 U から $t-1$ 期の標本 S_{t-1} を除いた企業 $U-S_{t-1}$ の中から n 社を非復元単純無作為抽出する．これは現行の方法であり，本稿では“従来型”と呼ぶ．第二の抽出法では，S_{t-1} の半分の $n/2$ 社を t 期にも継続し，$n/2$ 社を $U-S_{t-1}$ の中から非復元単純無作為抽出する．どの企業も 2 期継続して標本となる．最初に $t=0$ 期から $t=1$ 期へ継続する $n/2$ 社は S_0 の中から非復元単純無作為抽出により選ぶ．Wolter (1979) はこれを one-level ローテーション抽出法と呼んでいるが，本稿では単に“ローテーション型”と呼ぶ．第三の抽出法は，$t=0$ 期に非復元単純無作為抽出した標本 S_0 を全期にわたり標本とし続けるものであり，“固定型”と呼ぶ．

従来の方法では，どの抽出法を用いるにせよ，t 期の総計 X_t は t 期の標本 S_t を用いて次式で推定する．

$$\hat{X}_t^{[m]}=\frac{N}{n}\sum_{i\in S_t}x_{i,t},\quad t=0,\dots,T \tag{2.1}$$

ただし $m=1,2,3$ は上記の標本抽出法の種類を順に表す．また成長率 $G_{t-1,t}$ は次式で推定する．

$$\hat{G}^{[m]}_{t-1,t}=\log\hat{X}^{[m]}_{t}-\log\hat{X}^{[m]}_{t-1}=\log\left(\frac{\sum_{i\in S_t}x_{i,t}}{\sum_{i\in S_{t-1}}x_{i,t-1}}\right),\quad t=1,\dots,T \tag{2.2}$$

標本サイズ n が十分大きいとき，$\hat{G}^{[m]}_{t-1,t}$ の分散は次式で近似できる（土屋・吉岡，2008）．

$$V(\hat{G}^{[m]}_{t-1,t})\approx(1-f)\frac{1}{n}\left\{\left(\frac{\sigma_{t-1}}{\mu_{t-1}}\right)^2+\left(\frac{\sigma_t}{\mu_t}\right)^2\right\}-2\left(\frac{n_c}{n}-f\right)\frac{1}{n}\rho_{t-1,t}\left(\frac{\sigma_{t-1}}{\mu_{t-1}}\right)\left(\frac{\sigma_t}{\mu_t}\right) \tag{2.3}$$

ただし $f=n/N$ は抽出率であり，n_c は継続標本のサイズである．従来型では $n_c=0$，ローテーション型では $n_c=n/2$，固定型では $n_c=n$ となる．σ_t/μ_t は t 期の変動係数であり，$\rho_{t-1,t}$ は $t-1$ 期と t 期の相関係数である．

次は指数型の方法である．標本の抽出方法は上記の“従来型”のみとする．すなわち標本は毎期全て交替する．標本企業に $t-1$ 期と t 期の計数値を同時に回答してもらうため，標本のローテーションを行うと情報が重複し，企業に負担をかけることとなるからである．なお Wolter (1979) は，二期分の計数値を同時に回答してもらう方法を two-level デザインと呼んでいる．本稿では，まず成長率 $G_{t-1,t}$ を次式で推定する方法を“金額指数型”と呼ぶ．

$$\hat{G}^{[4]}_{t-1,t}=\log\left(\frac{\sum_{i\in S_t}x_{i,t}}{\sum_{i\in S_t}x_{i,t-1}}\right),\quad t=1,\dots,T \tag{2.4}$$

(2.2) 式とは異なり，比の分子・分母ともに同一の標本 S_t を用いる．$\hat{G}^{[4]}_{t-1,t}$ の分散 $V(\hat{G}^{[4]}_{t-1,t})$ は，$n_c=n$ とした (2.3) 式である．

また次式で推定する方法を“Trans-log 指数型”と呼ぶ．

$$\hat{G}^{[5]}_{t-1,t}=\sum_{i\in S_t}\phi_{i,t}\log\frac{x_{i,t}}{x_{i,t-1}} \tag{2.5}$$

ただし

$$\phi_{i,t}=\frac{1}{2}\left(\frac{x_{i,t}}{\sum_{i\in S_t}x_{i,t}}+\frac{x_{i,t-1}}{\sum_{i\in S_t}x_{i,t-1}}\right) \tag{2.6}$$

である．これはディヴィジア指数の離散近似として Trans-log 指数を用いたものである (Diewert, 1976)．

$$\sum_{i\in S_t}\phi_{i,t}\frac{d\log x_{i,t}}{dt}\approx\sum_{i\in S_t}\phi_{i,t}\log\frac{x_{i,t}}{x_{i,t-1}} \tag{2.7}$$

なお $x_{i,t}=0$ あるいは $x_{i,t-1}=0$ となる法人 i は除外して計算する．$\hat{G}^{[5]}_{t-1,t}$ の分散 $V(\hat{G}^{[5]}_{t-1,t})$ は次式で近似できる．

$$V(\hat{G}^{[5]}_{t-1,t})\approx N^2(1-f)\frac{1}{n(N-1)}\sum_U\left(z_i-\frac{1}{N}\sum_U z_i\right)^2 \tag{2.8}$$

ただし

$$z_i=\frac{1}{2X_t}\left(x_{i,t}\log\frac{x_{i,t}}{x_{i,t-1}}-\frac{\check{X}_t}{X_t}x_{i,t}\right)+\frac{1}{2X_{t-1}}\left(x_{i,t-1}\log\frac{x_{i,t}}{x_{i,t-1}}-\frac{\check{X}_{t-1}}{X_{t-1}}x_{i,t-1}\right) \tag{2.9}$$

であり，

$$\check{X}_t = \sum\nolimits_U x_{i,t} \log \frac{x_{i,t}}{x_{i,t-1}}, \quad \check{X}_{t-1} = \sum\nolimits_U x_{i,t-1} \log \frac{x_{i,t}}{x_{i,t-1}} \tag{2.10}$$

である．

金額指数型と Trans-log 指数型のいずれについても，総計 X_t は $t=0$ 期の総計 X_0 をベンチマークとし，これに成長率の推定値を乗じることで推定する．

$$\hat{X}_t^{[4]} = X_0 \exp(\hat{G}_{0,1}^{[4]}) \cdots \exp(\hat{G}_{t-1,t}^{[4]}) = X_0 \exp\left(\sum_{t'=1}^{t} \hat{G}_{t'-1,t'}^{[4]}\right), \quad t=1,\ldots,T \tag{2.11}$$

$$\hat{X}_t^{[5]} = X_0 \exp(\hat{G}_{0,1}^{[5]}) \cdots \exp(\hat{G}_{t-1,t}^{[5]}) = X_0 \exp\left(\sum_{t'=1}^{t} \hat{G}_{t'-1,t'}^{[5]}\right), \quad t=1,\ldots,T \tag{2.12}$$

以上の五つの型を以下に整理しておく．

(1) 従来型：$t-1$ 期と t 期で標本を完全に交替する従来の方法
(2) ローテーション型：$t-1$ 期と t 期で標本の半分を重複させる従来の方法
(3) 固定型：$t-1$ 期と t 期で同一の標本を用いる従来の方法
(4) 金額指数型：$t-1$ 期と t 期で標本を完全に交替する指数型の方法
(5) Trans-log 指数型：$t-1$ 期と t 期で標本を完全に交替する指数型の方法

2.3 シミュレーションの方法

まず 1996 年度を $t=0$ 期のベンチマークとし，9 項目について上記五つの型をそれぞれ用いて各年度の X_t と $G_{t-1,t}$ を推定した．また営業利益と経常利益の X_t を以下により推定した．

$$\text{営業利益} = \text{売上高} - \text{売上原価} - \text{販売費} \tag{2.13}$$

$$\text{経常利益} = \text{営業利益} + \text{営業外収益} - \text{営業外費用} \tag{2.14}$$

さらに総計の推定値の誤差 $\varepsilon_t = \hat{X}_t - X_t$ および成長率の推定値の誤差 $u_t = \hat{G}_{t-1,t} - G_{t-1,t}$ も求めた．この作業を 2004 年度まで行うことで 1 回のシミュレーションとする．シミュレーションを 1,000 回独立に繰り返し，1,000 個の $\hat{X}_t$ と $\hat{G}_{t-1,t}$ の平均と標準偏差を求めた．なおシミュレーションでは，抽出率 $f=n/N$ として $f=.50$，.30，.10，.01 の四種類を試みた．結果はいずれも同様の傾向を示したので，以下では基本的に $f=.30$ だけを取り上げる．すなわち $n=fN=.30\times 3,456=1,036$ である．

3. 結果

3.1 成長率の比較

まず図 1 は，9 項目のうち売上高と当期資産合計，従業員数の成長率の推定値 $\hat{G}_{t-1,t}$ の分布を箱ヒゲ図で示したものである．各図を横切る折れ線は，疑似母集団 $N=3,456$ を用いて求めた真の成長率である．図に示していない 6 項目についても，結果は同様であったので詳細は割愛する．なお成長率については，固定型と金額指数型の推定量は等しい．

いずれの型も，1,000 回の繰り返しの平均，すなわち $\hat{G}_{t-1,t}$ の期待値の推定値は真値に近く，成長率の推定値に偏りはないと考えてよい．このことは，売上高に関する結果をまとめた表 1 からも分かる．

1,000 回の繰り返しの標準偏差すなわち標準誤差の推定値は，売上高・当期資産合計・従業員数のいずれについても，従来型よりローテーション型の方が若干小さい．しかし固定型や二つの指数型はさらに小さくなっている．表 1 に示す売上高を例とすると，従来型の標準誤差は

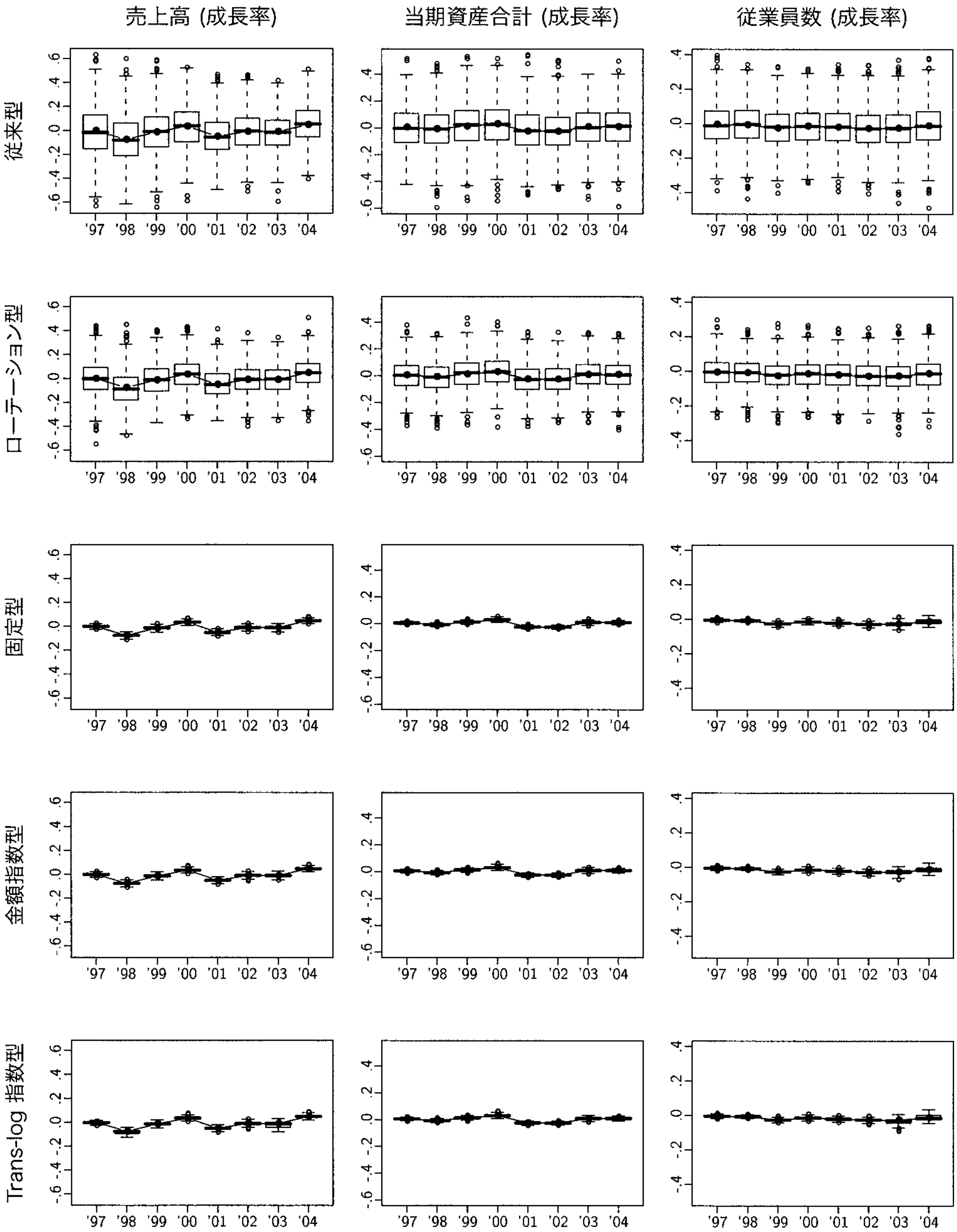


図 1. 成長率の推定値の分布.

表 1. 売上高の成長率と総計の推定値 (期待値および標準誤差).

		1997年	1998年	1999年	2000年	2001年	2002年	2003年	2004年
売上高	成長率G 真値	-0.001363	-0.076716	-0.014045	0.034294	-0.051016	-0.009793	-0.010508	0.047735
期待値	従来型	-0.011452	-0.072460	-0.011531	0.034365	-0.049258	-0.005691	-0.019912	0.052180
	ローテーション型	-0.002294	-0.083208	-0.011909	0.035062	-0.045533	-0.006495	-0.011109	0.044410
	固定型	-0.000827	-0.075815	-0.013464	0.034464	-0.051087	-0.010070	-0.010584	0.047685
	金額指数型	-0.001099	-0.075848	-0.013469	0.034746	-0.050357	-0.009591	-0.010387	0.047971
	Trans-log 指数型	-0.003786	-0.081142	-0.013470	0.035233	-0.050473	-0.010314	-0.019703	0.048920
標準誤差	従来型	0.211373	0.201452	0.190857	0.177944	0.170388	0.164044	0.158124	0.159419
	ローテーション型	0.141526	0.141891	0.133751	0.130823	0.120580	0.121103	0.116375	0.116646
	固定型	0.008987	0.011522	0.012232	0.010042	0.010930	0.011498	0.013781	0.009744
	金額指数型	0.008995	0.011319	0.012680	0.010471	0.010906	0.012132	0.014072	0.010037
	Trans-log 指数型	0.007871	0.016028	0.012717	0.010632	0.010927	0.012288	0.024366	0.011784
売上高	総計X 真値	463,019,536	428,826,710	422,845,988	437,598,641	415,834,145	411,781,810	407,477,327	427,400,092
期待値	従来型	460,098,291	427,563,110	422,339,494	436,896,645	415,739,588	413,184,181	405,009,867	426,698,389
	ローテーション型	464,120,861	426,956,501	421,589,007	436,334,731	416,847,546	413,978,391	409,346,018	427,951,604
	固定型	464,955,664	430,714,934	424,727,982	439,433,617	417,361,034	413,052,719	408,635,005	428,588,145
	金額指数型	463,160,392	429,361,899	423,637,863	438,625,288	417,099,677	413,138,008	408,901,535	429,015,864
	Trans-log 指数型	461,913,349	425,970,853	420,291,145	435,372,365	413,959,210	409,731,471	401,845,441	422,021,859
標準誤差	従来型	56,130,970	49,175,649	45,637,274	45,067,919	41,546,552	39,403,737	38,203,808	40,158,785
	ローテーション型	54,745,773	49,804,765	46,424,309	45,030,052	42,722,223	40,526,500	39,308,664	41,237,480
	固定型	56,276,220	49,693,234	47,000,751	46,943,133	42,845,477	41,154,655	40,008,305	41,890,910
	金額指数型	4,166,619	6,503,555	7,644,208	8,407,882	8,760,603	9,579,249	10,791,983	12,084,985
	Trans-log 指数型	3,633,769	7,763,997	8,666,073	9,415,231	9,651,176	10,385,311	13,780,066	15,305,747

.158～.211 であるのに対し，ローテーション型は .116～.142 と 2/3 程度に減っている．さらに固定型や指数型では .008～.024 であり，従来型の 1/10 以下にまで縮小している．

3.2 総計の比較

次は総計の推定値 $\hat{X}_t$ の比較である．図 2 に売上高・当期資産合計・従業員数の総計の推定値を箱ヒゲ図で示す．従来型・ローテーション型・固定型の三つの従来の方法では，総計の推定値は等しくなる．

五つの型のいずれも偏りは見られない．表 1 に示す売上高の結果でもこのことは確認できる．標準誤差は，三つの従来の方法に比べ，二つの指数型の方法の方が明らかに小さい．指数型の方法では，ベンチマークとした 1996 年度から年を経るにしたがい推定量の標準誤差は若干拡大するが，8 年を経た 2004 年度であっても，従来の方法より標準誤差はかなり小さい．具体的にどの程度小さいのかを見るため，売上高について従来型の標準誤差に対する金額指数型の標準誤差の比を求めたものが表 2 である．表 2 には $f=.30$ の結果だけでなく，他の抽出率の結果も示した．例えば 1997 年における $f=.30$ の 7.4%とは，表 1 に示される従来型の標準誤差 56,130,970 に対する金額指数型の標準誤差 4,166,619 の比である．抽出率によらず，1997 年時点では指数型の方法の標準誤差は従来の方法の 1/10 より小さい．2004 年であっても，指数型の方法は従来の方法の三割程度の標準誤差となっている．なお標準誤差率にすると，表 1 によれば従来の方法は 9%から 12%程度であるのに対し，指数型の方法は 1%から 4%程度である．

図 3 は，営業利益の総計の推定値 $\hat{X}_t$ である．前述のとおり，この値はまず売上高，売上原価，販売費の総計を推定し，そこから求めたものである．三つの従来の方法と二つの指数型の方法との間の差は，図 2 に示す三つの項目ほどは大きくない．ベンチマークの年である 1996 年から 5 年後の 2001 年頃までは指数型の方法の方が標準誤差は小さい．しかし 2002 年頃以降は，従来の方法と指数型の方法の差はほとんどないと言ってよい．なお経常利益についても同様の傾向が認められた．すなわちベンチマークの年に近いほど指数型の方法の方が標準誤差は小さいが，年を経るに従って指数型の方法の標準誤差は拡大し，2002 年頃以降は従来の方法と

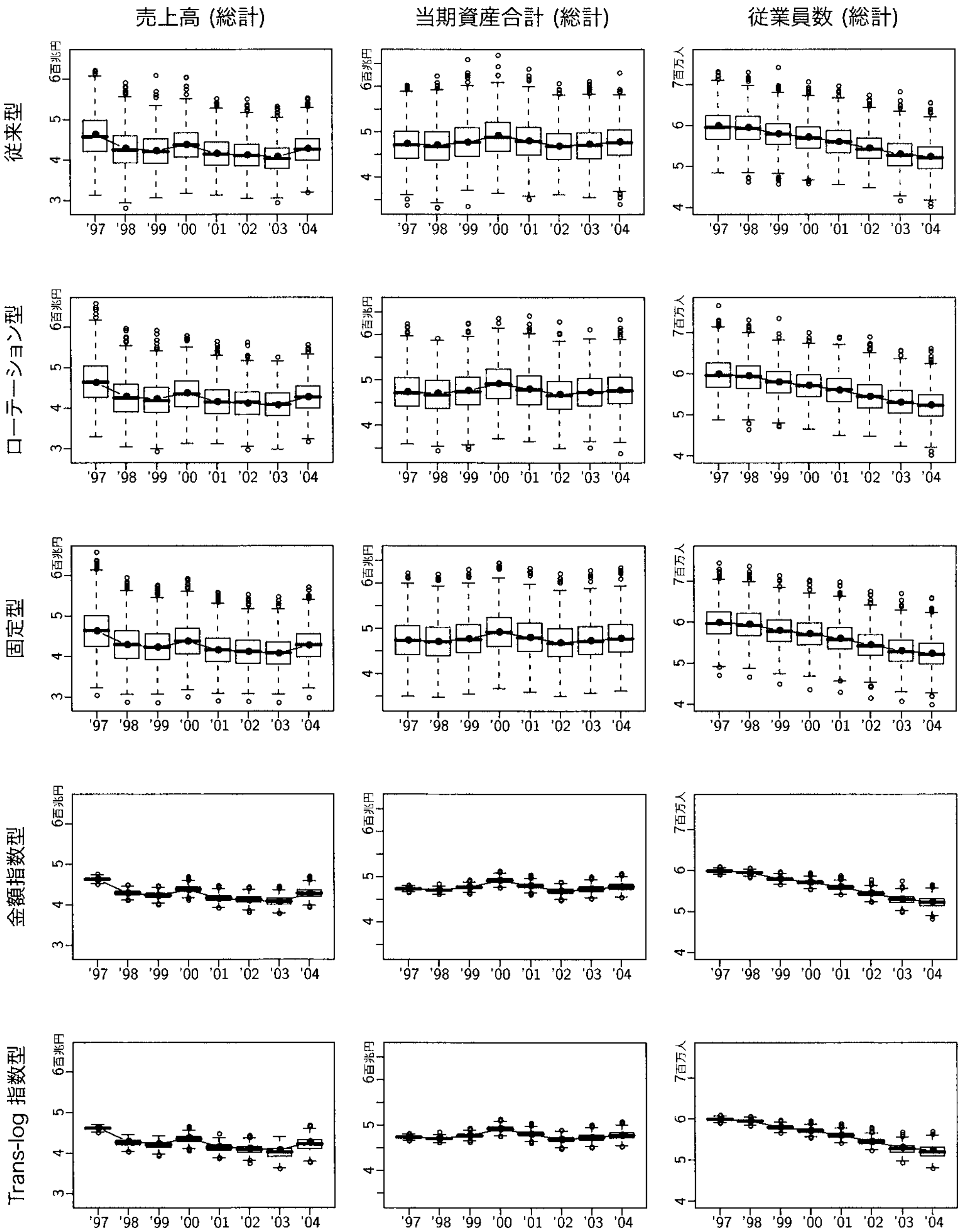


図 2. 総計の推定値の分布.

表 2.　売上高総計の推定量の標準誤差比較（従来型の標準誤差に対する金額指数型の標準誤差の比）.

総計X　標準誤差 金額指数型／従来型	1997年	1998年	1999年	2000年	2001年	2002年	2003年	2004年
f=.50	7.3%	14.7%	15.0%	16.4%	16.5%	18.1%	23.3%	24.6%
f=.30	7.4%	13.2%	16.7%	18.7%	21.1%	24.3%	28.2%	30.1%
f=.10	7.3%	12.0%	16.6%	20.0%	22.6%	25.8%	30.0%	31.9%
f=.01	7.9%	11.8%	14.7%	18.0%	21.5%	24.9%	27.6%	32.3%

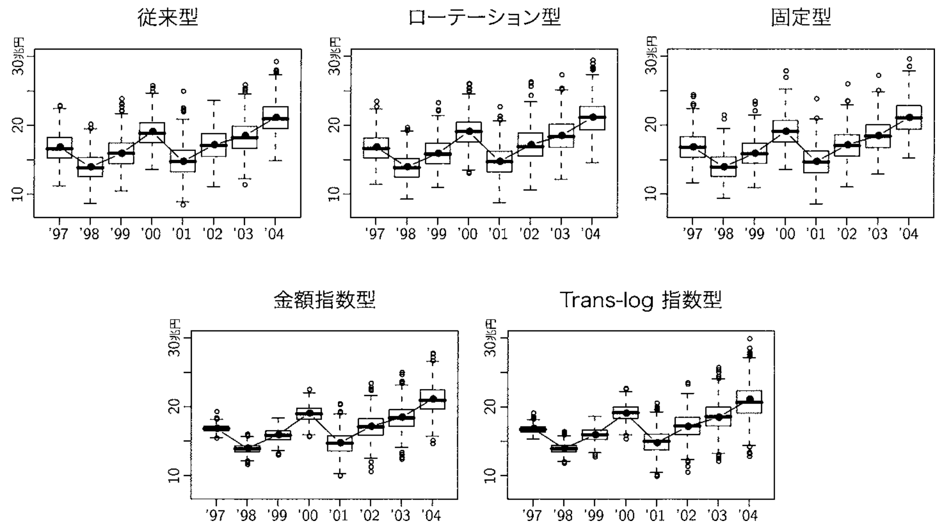


図 3.　推定値の分布（営業利益の総計）.

の差は見られなくなる．したがって指数型の方法によって合計額を推定するときには，例えば 5 年間など適度な短い期間でベンチマークをとる必要がある．

3.3　誤差の比較

最後は誤差の比較である．図 4 の上段は，売上高の成長率の推定値 $\hat{G}_{t-1,t}$ について，一期前の誤差 $u_{t-1}=\hat{G}_{t-2,t-1}-G_{t-2,t-1}$ を横軸，当該期の誤差 $u_t=\hat{G}_{t-1,t}-G_{t-1,t}$ を縦軸として，$t=1998$ 年度から 2004 年度までの 7 年分を重ねてプロットし，さらにそれを 1,000 回のシミュレーションのうち 200 回分について重ねたものである．

従来型では明らかに負の系列相関が認められる．すなわち，ある年の成長率が過大に推定されれば，翌年の成長率は過小に推定される．ローテーション型や金額指数型，Trans-log 指数型では系列相関が 0 に近く，この傾向は緩和される．

図 4 の下段は，売上高総計の推定値 $\hat{X}_t$ について，成長率のときと同様に誤差 $\varepsilon_t=\hat{X}_t-X_t$ をプロットしたものである．従来型では負の系列相関があり，ある年で推定値が過大となれば翌年は過小となる．固定型では正の高い系列相関があり，推定値がいったん過大となればその後も過大であり続けることになる．規模の大きな企業が選ばれれば，その後も継続してしまうのである．つまり固定型は，成長率の推定に関しては指数型と同様の性質を持つ．しかし総計の推定に関しては，最初に選ばれた標本に依存しすぎてしまう点が限界と言える．指数型の方法でも系列相関は正であるが，そもそも誤差自体が固定型よりもかなり小さい．

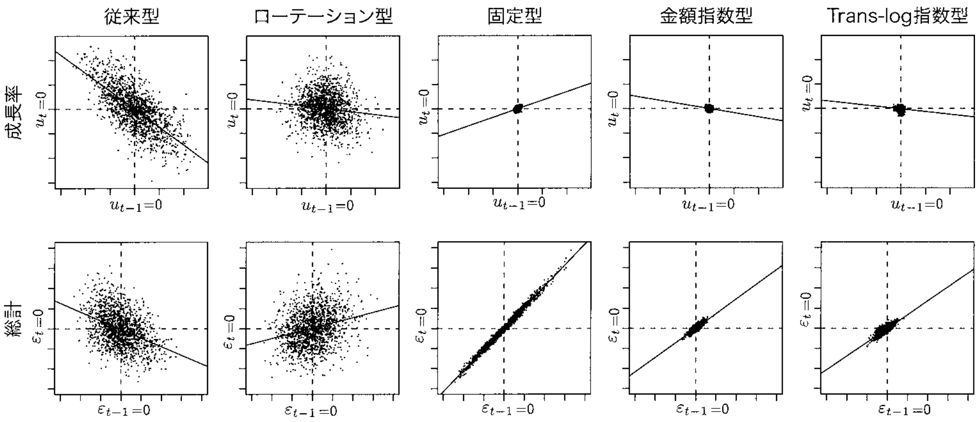


図 4. 成長率と総計の推定値の誤差 (売上高).

4. まとめと今後の課題

得られた結果を以下に箇条書きの形でまとめる.

(1) 五つの型のいずれについても，標本サイズが十分に大きかった今回のシミュレーションでは，成長率と総計の推定値に偏りはないと考えてよい.

(2) 成長率 G については，ローテーション型とすることで標準誤差を従来型の 2/3 程度に小さくすることができる．さらに指数型の方法では，1/10 以下とすることができる.

(3) 総計 X については，指数型の方法を採用することで標準誤差を従来の方法よりも小さくすることができる．ただし指数型の方法では，ベンチマーク年から時間を経るにつれ標準誤差は大きくなるので，従来の方法に比べた縮小の程度は年によって異なる．ベンチマーク年の翌年は 1/10 未満であるが，8 年を経ると 1/3 程度となる.

(4) 営業利益など他の項目の推定値から総計を求める項目については，ベンチマーク年から 5 年程度の間は，従来の方法よりも指数型の方法の方が標準誤差は小さい．しかしその後は従来の方法と指数型の方法は同程度となる.

(5) 従来型では成長率 G の誤差に負の相関が見られる.

本稿の実験は，成長率の新たな推定方法を提案するための第一歩に過ぎない．本稿で試みた推定方法を実用化するまでには少なくともまだ以下の諸課題が残されている．まず現実の母集団は固定されていない．法人企業の参入・退出がある．同一企業の資本金規模や業種が移動することも少なくない．また実験に当たってはパネル化が可能な資本金 10 億円以上の企業を疑似母集団としたが，現実に標本調査の対象とするのは資本金 10 億円未満の企業である．10 億円未満の企業であっても $t-1$ 期と t 期の相関係数 $\rho_{t-1,t}$ は非常に高いと推測される．したがって (2.3) 式に示す推定量の分散式に鑑みれば，成長率の推定において指数型の方法の方が従来の方法よりも誤差が小さいという本稿の結論は，10 億円未満の企業に対しても当てはまると考えられる．しかし誤差がどの程度小さくなるのかは不明である．さらに今回は年次別調査のデータを用いたが，四半期別調査のデータを用いた実験も必要である.

特別損益など 0 が多い項目の扱い方についても検討が求められる．本稿では Trans-log 指数型において $x_{i,t}=0$ あるいは $x_{i,t-1}=0$ となる法人は除外したが，除外する法人が多いときの影

響は明らかではない．また仮に調査の重心を，本稿で試みた指数型の方法に移すとすれば，ベンチマークが不可欠である．当然，全数調査であっても例えば無回答などの非標本誤差は生じる．ベンチマークが誤差を含めば，指数型の方法では総計の推定に影響を与えることになる．いつ，どのような方法で全数調査を行うのかが課題となる．なおベンチマークの作成に当たっては，2009 年 7 月の基礎調査を端緒とする経済センサスが一定の役割を果たすものと期待される．

さらに指数型の方法では，各法人企業には現在の一期分に加え，もう一期分の計数値の記入も求めることとなる．プレプリントが一つの有効な対処法であるが，それでも各法人企業の負担増は否めない．

これらの点に関して今後の方向性を判断する一つの材料は，内閣府と財務省が共同で実施している法人企業景気予測調査から得られるかもしれない．数期前からの計数値を併記してもらう調査票を既に採用しているが，推定自体は“従来の方法”を用いているからである．指数型の方法を試みることで有用な示唆が得られるものと考えられる．

謝　辞

投稿原稿を精読し，有益なコメントをいただきました査読者の先生に感謝いたします．本研究は科学研究費補助金(課題番号：17203017，基盤研究(A)法人企業統計調査と事業所・企業統計調査のミクロデータの統合新統計の編成と解析研究，研究代表者: 松田芳郎)の助成を受けたものである．また法人企業統計調査の調査票は平成 19 年 4 月 12 日官報第 4562 号総務省告示第 240 号に基づき使用した．

参 考 文 献

Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed., John Wiley & Sons, New York.

Diewert, W. E. (1976). Exact and superlative index numbers, *Journal of Econometrics*, **4**, 115–145.

Duncan, G. J. and Kalton, G. (1987). Issues of design and analysis of surveys across time, *International Statistical Review*, **55**, 97–117.

Eckler, A. R. (1955). Rotation sampling, *Annals of Mathematical Statistics*, **26**, 664–685.

Kish, L. (1987). *Statistical Design for Research*, John Wiley & Sons, New York.

Patterson, H. D. (1950). Sampling on successive occasions with partial replacement of units, *Journal of the Royal Statistical Society, Series B*, **12**, 241–255.

社会工学研究所(1976).『法人企業統計の高度利用に関する調査研究(統計編-1,2,3，総合解説編)』，社会工学研究所.

Sunter, A. B. (1977). Response burden, sample rotation, and classification renewal in economic surveys, *International Statistical Review*, **45**, 209–222.

土屋隆裕，吉岡完治(2008). 標本ローテーションの下での成長率の推定量，Research Memorandum, No.1082, The Institute of Statistical Mathematics, Tokyo.

Wolter, K. M. (1979). Composite estimation in finite populations, *Journal of the American Statistical Association*, **74**, 604–613.

A Comparison of Estimators for the Survey of Corporation Statistics

Takahiro Tsuchiya[1], Kanji Yoshioka[2] and Yoshiro Matsuda[3]

[1]The Institute of Statistical Mathematics
[2]Keio Economic Observatory, Keio University
[3]Department of Management and Economics, Aomori Public College

Several methods for estimating both population totals and their growth rates were compared by simulation with microdata from a survey of corporation statistics as a pseudo population. The conventional survey design basically replaces the whole samples with a new one, and the growth rates are estimated from the estimates of population totals of each year. The simulation study revealed that the rotating design, in which half the sample is kept in the panel until the next year, reduces standard errors of growth rates by almost two thirds that of the conventional design. The study also indicated that the estimating strategy, in which the growth rates are first estimated and then secondly used for estimating the population totals, decreases standard errors of population totals by one tenth to one third and growth rates by one tenth of the conventional design even if all the sample is replaced annually.

Key words: The survey of corporation statistics, sample rotation, growth rates.