

公的統計の精度向上の取組

統計技術の評価に資する事項

本日の議題

建築着工統計調査（補正調査）の移行期の集計方法

経済産業省生産動態統計調査の欠測値補完方法の検証

（2）評価及び課題解決に向けた取組の方向性

① 標本設計の見直し

本検査は、様々なシミュレーションに基づき、標本誤差が小さくなる
ことが定量的に示されており、意義のある検査であると評価できる。

このため、国土交通省は、今後、本検査で示された改善提案を実現で
きるように検討を進めることが必要である。ただし、検討に当たっては
以下のア）及びイ）を踏まえる必要がある。

ア） 総務省から提案された標本配分法の変更に関し、厳密なネイマン
配分法を採用するためには、その都度、工事費予定額の標準偏差を計
算しながら対応することが必要になるなど、多くの実務的な課題が想
定される。したがって、国土交通省においては、ネイマン配分法に準
じた範囲で実務的な対応が可能なものとなるよう検討を進めること
が必要である。

イ） ネイマン配分に準じた配分率を算定する際、しつ皆層、標本層別
の回収率等を踏まえながら最終的な回収率として現行の標本サイズ
（約 5000）が維持できるよう検討することが必要である。

なお、標本設計の見直し実施後も、しつ皆層のしきい値として示され
ている工事費予定額 20 億円及び層化基準のしきい値として示されてい
る工事費予定額 1 億円については、定期的に検証を行うことが必要であ
る。

② 標本設計の見直しに伴う調査方法の見直し及びスケジュール

標本設計の見直しを行うに伴い、抽出方法を含む調査方法等の変
更が伴うものと見込まれる。この変更に向け、国土交通省は、平成 30 年
度に試験調査を実施し、その後、必要な準備を進め、平成 33 年（2021
年）1 月に完成する建築物から見直し後の標本設計に基づき抽出された
対象に切り替えて調査を開始することが必要である。

なお、この切り替えの際、ア）見直し後の標本設計に基づき抽出され
た調査対象と、イ）現行の抽出方法に基づき抽出され都道府県から国土
交通省に移管された調査対象、の両方を混在させて調査を行うことが必
要となり、その移行期の結果推定には統計技術的な対応が必要である。
そのため、総務省は移行時期の結果推定の方法の開発について国土交通
省に協力し、円滑な調査の移行を支援することが必要である。

③ 都道府県別集計の取扱い

総務省が提案している補正調査の標本設計の改善は、全国の推定精度
を高めるため、各都道府県に配分されていた標本を都市部に配分する結
果となり、建築着工数の少ない県の結果精度は悪化することが予想され
る。

都道府県別集計結果の精度を向上させるためには、標本規模の増加な
どが考えられるが、現状の行政ニーズ及び実査担当機関の大幅な負担増
加を踏まえると、標本規模は現状の 5,000 件を増加させる状況にない。
このため、都道府県別集計の取りやめは、やむを得ないものとする。

ただし、都道府県別集計結果の利活用の現状は、現行の標本設計に基
づく結果精度の低迷が原因である可能性もあるため、今後の標本設計の
見直しに伴い補正調査の全国の推定精度が向上すれば、都道府県等によ
る結果利活用の可能性も生じる。このため、県民経済計算の精度向上な
どに資する観点から、大規模都道府県における特別集計や地域ブロック
別集計を実施する意義はないかなど、今後の結果利活用について引き続
き検討する必要がある。

3 事業所及び企業を対象とする基幹統計調査における欠測値及び外れ値に
関する原則的な対応の確認及び整理

（1）検査結果と現状

プライバシー意識の高まりや企業の間接部門の合理化などにより、統
計調査の回答を得るための環境は厳しい状況になっている。このため、各
府省では、オンライン調査の導入、電話による督促の実施など回答を得る
様々な努力を行っているが、最終的に調査票の未提出や提出されても一
部の調査項目に未記入があるなど、現実問題として欠測値が発生してい
る。

このような状況の中で、正確な統計を作成していくには、過去の回収率
が高い時に使用していた統計作成方法を見直し、欠測値により適切に対
応することが重要な鍵となる。このような問題意識の下、「平成 26 年度統
計法施行状況に関する審議結果報告書（末諮問基幹統計確認関連分）（平
成 28 年 3 月 22 日統計委員会）」の「5 横断的な課題への対応」におい
て、統計調査横断的に欠測値や外れ値等への適切な対応を推進すること
や統計精度向上の取組の創設を統計委員会として提言したところである。
これを受けて実施された、「事業所及び企業を対象とする基幹統計調査

建築着工統計調査 補正調査の概要

調査の目的

- ・ 建築物の竣工時に実際にかかった費用（工事実施額）及び床面積（実施床面積）を調査し、着工時における工事費予定額及び床面積とのかい離を明らかにする。

抽出方法（二段抽出法）

- ・ 調査実施市区を選定（291市区）
- ・ 調査実施市区の着工建築物について、都道府県別木造・非木造別に定めた抽出率により抽出 ※標本サイズは5,000件程度／年（実績）

調査方法

- ・ 抽出した建築物の工事が完了したときに職員等が実地調査

総務省・統計委員会が示した補正調査の見直し

	現行	見直し案
標本サイズ	約5,000件/年（実績）	同左
抽出方法	層化二段抽出 抽出単位 1 段目：市区（固定） 2 段目：建築物（層化抽出）	層化抽出 抽出単位：建築物 ※工事費予定額20億円以上は しっ皆
層化基準	<ul style="list-style-type: none"> 都道府県（47区分） 建築物の構造（木造・非木造） 	<ul style="list-style-type: none"> 建築物の構造（木造・非木造） 工事費予定額階級 （1億円未満、1～20億円の2区分）
標本配分	層別に抽出率を設定 （1/10～1/100）	工事費予定額による ネイマン配分

木造・非木造計
の予定単価の誤差率

4.8%



0.8%

工事予定額階級	木造	非木造
1億円未満	3,060件	1,711件
1億～20億円	12件	213件
20億円以上	0	4件



工事予定額階級	木造	非木造
1億円未満	1,415件	726件
1億～20億円	50件	1993件
20億円以上	1件	815件

完 成 年

補正調査の調査対象年

t 年

t+1 年

t+2 年

届
出
年

t 年

t+1 年

t+2 年

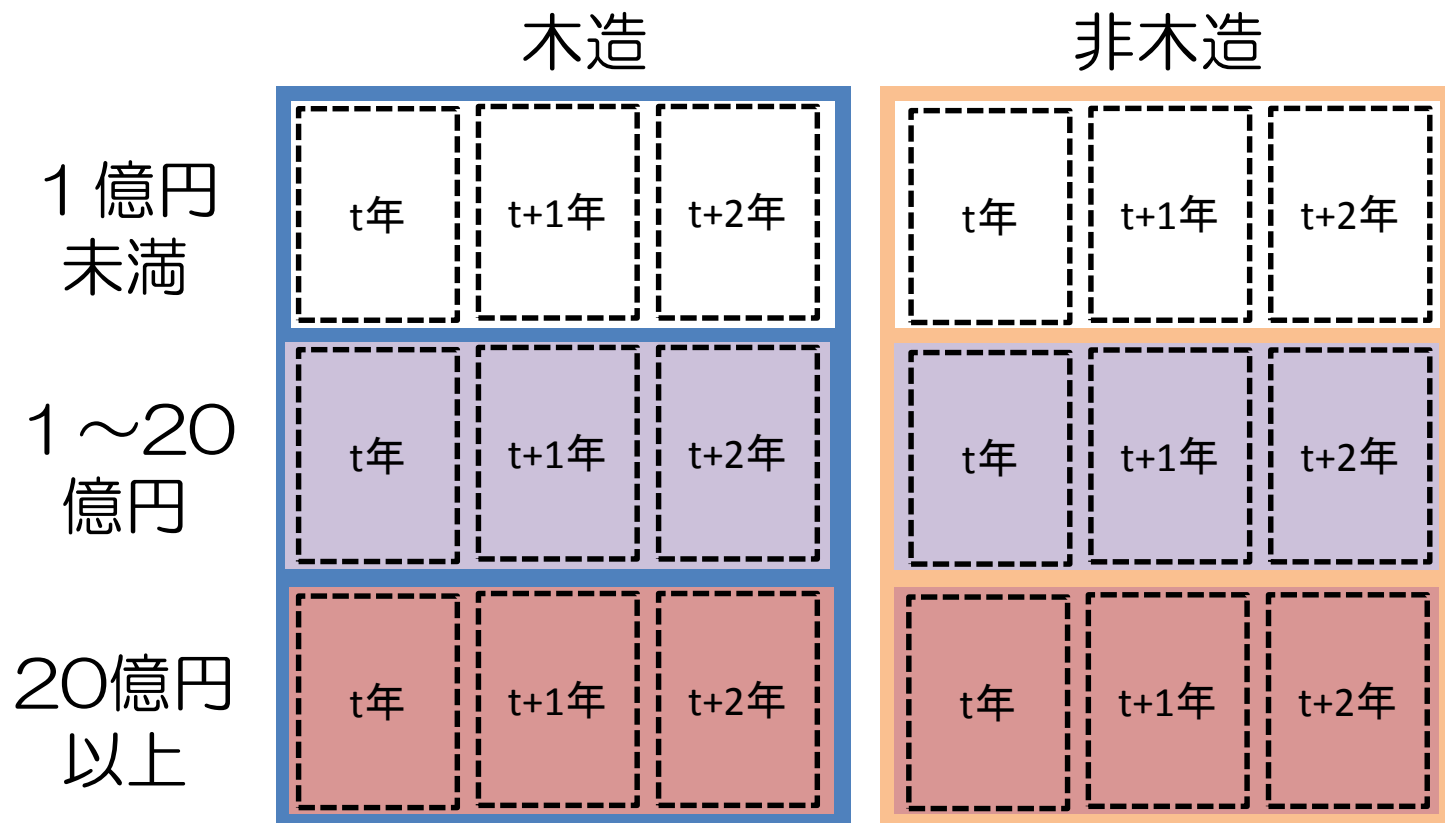
旧抽出法
(比例配分)

新抽出法
(悉皆・ネイマン配分)

移行期に問題

- 届出年（着工年）に抽出
- 完成年（竣工年）に調査

届出年の条件付確率による復元が必要



集計算式 1

$$\sum_i \sum_j \sum_l \sum_m \overset{\text{計測不能なデータ}}{N(i, j, k_{=t+2}, l)} \frac{X(i, j, k_{=t+2}, l, m)}{n(i, j, k_{=t+2}, l)}$$

$$\doteq \sum_i \sum_j \sum_l \sum_m N'(i, j, k'_{=t+2}, l) \frac{X(i, j, k_{=t+2}, l, m)}{n(i, j, k_{=t+2}, l)}$$

i: 構造(木造・非木造)

j: 工事予定額階級 (20億円以上、1~20億円、1億円未満)

k: 完成年 \longleftrightarrow *k'*: 完成予定年 *t*: 旧方式による抽出最終年, *t*+2: 調査年

l: 届出年 (着工年)

m: 標本インデックス

N: 構造*i*, 工事予定額階級*j*, 完成年*k*, 届出年*l* の建築物 総数 \longleftrightarrow *N'*: 構造*i*, 工事予定額階級*j*, 完成予定年*k'*, 届出年*l* の建築物 総数

n: 構造*i*, 工事予定額階級*j*, 完成年*k*, 届出年*l* の抽出標本数

x: 構造*i*, 工事予定額階級*j*, 完成年*k*, 届出年*l*, 標本インデックス*m*の標本の観測値

集計算式2

$$\sum_i \sum_j \sum_l \sum_m \boxed{N'(i, j, l) \frac{X(i, j, k_{=t+2}, l, m)}{n(i, j, l)}}$$

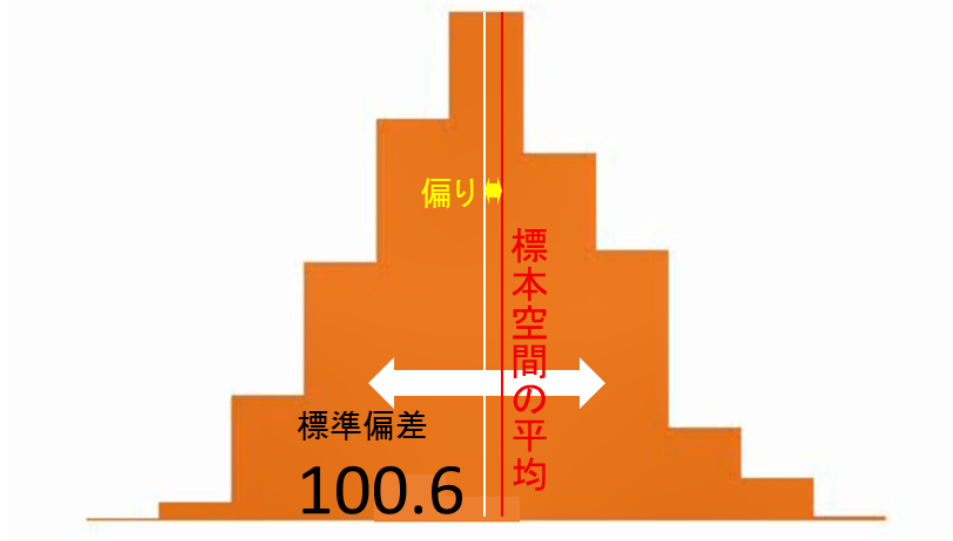
※ 母数Nと標本数nの「完成年」条件を無視
 つまり観測値 x 抽出時の抽出率の逆数
 ⇒ N/nが抽出作業時に予め計算されている

i: 構造 (木造・非木造)
j: 工事予定額階級 (20億円以上、1~20億円、1億円未満)
k: 完成年 ←→ *k'*: 完成予定年 *t*: 旧方式による抽出最終年, *t+2*: 調査年
l: 届出年 (着工年)
m: 標本インデックス

N: 構造*i*, 工事予定額階級*j*, 完成年*k*, 届出年*l* の建築物 総数 ←→ *N'*: 構造*i*, 工事予定額階級*j*, 完成予定年*k'*, 届出年*l* の建築物 総数
n: 構造*i*, 工事予定額階級*j*, 完成年*k*, 届出年*l* の抽出標本数
x: 構造*i*, 工事予定額階級*j*, 完成年*k*, 届出年*l*, 標本インデックス*m*の標本の観測値

リサンプリング実験による確認

- ① 2010年届出を旧抽出法、2011年と2012年を新抽出法で抽出を行い
- ② その時点の抽出率の逆数をウエイトとした集計を実施（集計算式2で集計）
- ③ 1000回繰り返し、集計算式2で歪みが生じないか確認
- ④ 「2010～2012年」「2011～2013年」・・・「2014～2016年」のデータセットでそれぞれ1000回のリサンプリング実験



改定率のリサンプリングシミュレーション結果

木造			非木造		
リサンプリング データ	標本空間の 平均	試算結果 1000件の平均 (標準偏差)	リサンプリング データ	標本空間の 平均	試算結果 1000件の平均 (標準偏差)
2010-2012年	1.01	1.014(0.004)	2010-2012年	1.07	1.069(0.009)
2011-2013年	1.02	1.022(0.002)	2011-2013年	1.02	1.014(0.007)
2012-2014年	1.03	1.029(0.003)	2012-2014年	1.04	1.049(0.015)
2013-2015年	1.03	1.031(0.004)	2013-2015年	1.04	1.043(0.005)
2014-2016年	1.03	1.023(0.003)	2014-2016年	1.05	1.037(0.003)

改定率のリサンプリングシミュレーション結果

- いずれの年においても概ね偏り無く推計できていると判断？
- 特殊処理は短期であることも踏まえれば、簡易な方法による集計においても、公的統計としての品質を維持し単価及び補正率が推計されるものと判断？

まとめ 集計方法1 (2つの提示式のうちの1つ目)

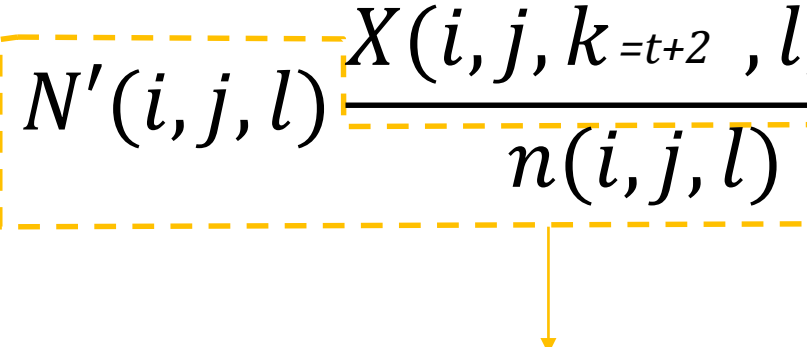
集計算式1

$$\sum_i \sum_j \sum_l \sum_m N'(i, j, k'_{=t+2}, l) \frac{X(i, j, k_{=t+2}, l, m)}{n(i, j, k_{=t+2}, l)}$$

	木造	非木造						
1億円未満	<table border="1"><tr><td>2019年</td><td>2020年</td><td>2021年</td></tr></table>	2019年	2020年	2021年	<table border="1"><tr><td>2019年</td><td>2020年</td><td>2021年</td></tr></table>	2019年	2020年	2021年
2019年	2020年	2021年						
2019年	2020年	2021年						
1~20億円	<table border="1"><tr><td>2019年</td><td>2020年</td><td>2021年</td></tr></table>	2019年	2020年	2021年	<table border="1"><tr><td>2019年</td><td>2020年</td><td>2021年</td></tr></table>	2019年	2020年	2021年
2019年	2020年	2021年						
2019年	2020年	2021年						
20億円以上	<table border="1"><tr><td>2019年</td><td>2020年</td><td>2021年</td></tr></table>	2019年	2020年	2021年	<table border="1"><tr><td>2019年</td><td>2020年</td><td>2021年</td></tr></table>	2019年	2020年	2021年
2019年	2020年	2021年						
2019年	2020年	2021年						

- 本体調査の3年分を全数集計し、
 - ① 2021年完成予定建築物だけ抽出
 - ② 構造別、工事予定額階級別、着工年別に集計し各セルのNを計算
- 各セルの回収標本数nを計算
- N/n を集計ウエイトに用いる

まとめ 集計方法2 (2つの提示式のうちの2つ目)

$$\sum_i \sum_j \sum_l \sum_m N'(i, j, l) \frac{X(i, j, k_{=t+2}, l, m)}{n(i, j, l)}$$


(完成時期・完成予定時期に依存しない) 抽出時の標本 $X(i, j, k, l, m)$ の抽出率の逆数をウエイトとする集計を実施

ご審議をお願いしたい論点

- 簡易な方法を使うことの是非について
- 移行期の集計の設計上、他に留意すべき点はないか
- 完全に移行した後の集計について、アドバイスはないか

本日の議題

建築着工統計調査（補正調査）の移行期の集計方法

経済産業省生産動態統計調査の欠測値補完方法の検証

一部非回答に対しては様々な単一補完が行われている。

このうち、全体合計に対する影響度が1%未満の企業における一部非回答（取引状況（輸出・入）、外部委託の状況、技術所有の内訳などに多く発生）には0値補完が行われているが、当該対応は、過小推計につながる懸念もあることから、0値補完対応の検証について、第Ⅲ期基本計画（表5参照）において指摘されている企業系統計調査の見直しの検討状況を踏まえ、必要に応じて総務省の支援も得ながら、適宜シミュレーション等の検証を行うことが必要である。

表5 「第Ⅲ期基本計画」 抜粋

具体的な措置、方策等	担当府省	実施時期
○ 経済構造実態調査と、経済産業省企業活動基本調査、法人企業統計調査、建設工事施工統計調査、情報通信業基本調査及び中小企業実態基本調査等との役割分担、重複是正等を検討する。	総務省、関係府省	平成34年度(2022年度)までに一定の結論を得る。

⑥ 過去の回答結果を使用した単一補完を実施する場合の使用データの期限のないもの

現在、商業動態統計調査（丙調査（百貨店・スーパー）及び丁3調査（ドラッグストア））、経済産業省生産動態統計調査、経済産業省特定業種石油等消費統計調査及び自動車輸送統計調査（バス関連調査）では、明確な使用期限を定めることなく過去の回答結果を使用した単一補完が行われており、長期にわたってデータを使用した場合、統計精度の悪化につながる懸念がある。このため、これらの統計調査は、以下のとおり対応していくことが必要である。

ア) 経済産業省所管調査

商業動態統計調査（丙調査（百貨店・スーパー）及び丁3調査（ドラッグストア））、経済産業省生産動態統計調査及び経済産業省特定業種石油等消費統計調査では、回収率90%以上を保持していることから、致命的な対応とは判断されないが、平成30年度から総務省の支援を得て、計画的にシミュレーションを行いデータの使用期限も含めた補完方法の検証を行うことが必要である。

イ) 国土交通省所管調査

自動車輸送統計調査（バス関連調査）では、以下⑦でも示してい

るとおり第Ⅲ期基本計画において自動車輸送統計調査そのものの見直しが指摘されていることから、当該見直しと整合するよう並行して、データの使用期限も含めて検証を行うことが必要である。

⑦ 調査そのものの見直しが必要であるもの

○ 自動車輸送統計調査

自動車輸送統計調査では、調査票に応じて、単一補完、ウエイト調整などの欠測値への対応の措置がとられているものの、一部の調査票様式について回収率が50%を下回るなど、欠測値に対する統計技術的な対応では補正しきれない状態に置かれているものと判断される。

このため、第Ⅲ期基本計画（表6参照）において調査そのものの見直しが指摘されていることに併せて推計方法や欠測値への対応についても検討を行い、結論を得ることが必要である。

表6 「第Ⅲ期基本計画」 抜粋

具体的な措置、方策等	担当府省	実施時期
◎ 自動車輸送統計調査について、精度向上を図る観点から、平成29年度(2017年度)に実施する予備的調査、他の輸送統計及び行政記録情報の活用も含めて分析・検討を促進し、新たな調査手法による調査を開始する。また、公表の早期化やニーズに応じた公表事項の充実、品目別輸送量の数値の安定化方策等について検討し、早期に結論を得る。	国土交通省	平成32年度(2020年度)から実施する。

⑧ その他、見直しの余地があると判断されるもの

○ 民間給与実態調査

民間給与実態調査では、源泉徴収義務者に対し、所属する給与所得者から抽出された給与所得者に関する情報の報告を求めている。

この給与所得者の情報において一部非回答が発生した場合、当該源泉徴収義務者から得られた情報全体を集計対象外とした上でウエイト調整を行っている（調査対象となる源泉徴収義務者の0.6%）。当該対応は、一部非回答に対する1つの対処方法ではあるものの、単一補完を行うことで集計対象外とした情報を利用できる余地もあると考えられる。このため、平成30年度に有識者の意見を聞く等により、改善の余地の有無について判断を行うことが必要である（判断の結果、改

善の余地がある場合、行政記録情報による欠測値への対応も含めその対応について引き続き検討を進めることが必要である。）。

⑨ 外れ値及び異常値の検出

外れ値及び異常値については、集計業務の中で、検出を行い、審査する対応が採られていたところ、疑義として検出し照会等の対応を行ったエラー数については、管理されていないものも見られた。

疑義への対応は、調査対象に照会が行われることから調査対象者への負担増大にもつながることを考慮すると、今後、業務効率化や統計調査の負担軽減の観点から、まずは①疑義の数と、②効果（訂正に繋がったもの）があった疑義の数を管理し、①と②との関係で効果のない疑義が生じている場合に、結果精度とのバランスも考慮しつつ、検出するしきい値の範囲を見直すなどの取組も必要である。

⑩ 欠測値や外れ値への対応の公表

欠測値や外れ値への対応は、統計調査の結果を解釈する際にも重要な情報となる場合があることから、各府省は欠測値や外れ値への対応について、ホームページに開示することが必要である。

4 平成30年度における精度向上の取組

1) 標準検査

平成30年度における標準検査（見える化状況検査）については、前述Ⅱ-1-(2)に記載したとおり。

2) オプション検査

平成30年度に総務省が行うオプション検査は、経済産業省が所管する

- ・ 商業動態統計調査
- ・ 経済産業省生産動態統計調査
- ・ 経済産業省特定業種石油等消費統計調査

のうちいずれかの調査を総務省及び経済産業省が選択して、過去の回答結果を使用した単一補完を実施する場合の使用データの期限の検討に資するシミュレーションを実施し、その結果について統計委員会に報告するとともに、経済産業省に対し欠測値への対応のより適切な対応を提案する。

なお、オプション検査については、季節調整法のチェックなど、調査統計に限らない検査も考えられることから、今後、オプション検査のあり方について検討することが必要である。

3) フォローアップ

前述Ⅱ-3-(2)-③から⑧までで対応を求めた事項のうち、平成30年度に検討期限を迎えるもの及び平成30年度から着手することとされている取組について、平成30年度後半にその進捗状況について統計委員会に報告を行い、必要に応じて審議を行う。報告に当たっては、第Ⅲ期基本計画で取り組むとされている評価チーム（仮称。以下同じ。）に1つの検討事項として提示し、評価チームで検討することとされた場合には、同チームにフォローアップを任せることとし、同チームの検討事項に採用されない場合には、横断的課題検討部会においてフォローアップを行うこととする。

- 経済産業省生産動態統計調査はシンプルな過去の回答結果をそのまま使用した単一補完 (LOCF: *Last Observation Carried Forward*) を採用
- 経済産業省生産動態統計調査を選択
- 検討の端緒として自動車等の調査票を選択

端緒の分析を総務省統計研究研修所に依頼

◆ 自動車等の調査票を選定した理由

ア 販売金額が比較的高額

イ 当該品目を生産又は販売している事業所数が比較的少ない

ウ 販売金額別の事業所数の分布に偏りがみられる



LOCFの影響を受けやすいと判断

◆ 項目

主要な指標である「販売金額」をシミュレーションの対象

◆ 分析対象期間

リーマンショック前後を含む2003年1月分～2017年12月分までの15年分（180か月）とした。

分析に使用した品目一覧

名称

軽乗用車

小型乗用車

普通乗用車

バスシャシー（完成車を含む） - 小型バス

バスシャシー（完成車を含む） - 大型バス

トラックシャシー（完成車を含む） - 軽自動車

トラックシャシー（完成車を含む） - 小型自動車 - ガソリン車

トラックシャシー（完成車を含む） - 小型自動車 - ディーゼル車

トラックシャシー（完成車を含む） - 普通自動車 - ガソリン車

トラックシャシー（完成車を含む） - 普通自動車 - ディーゼル車

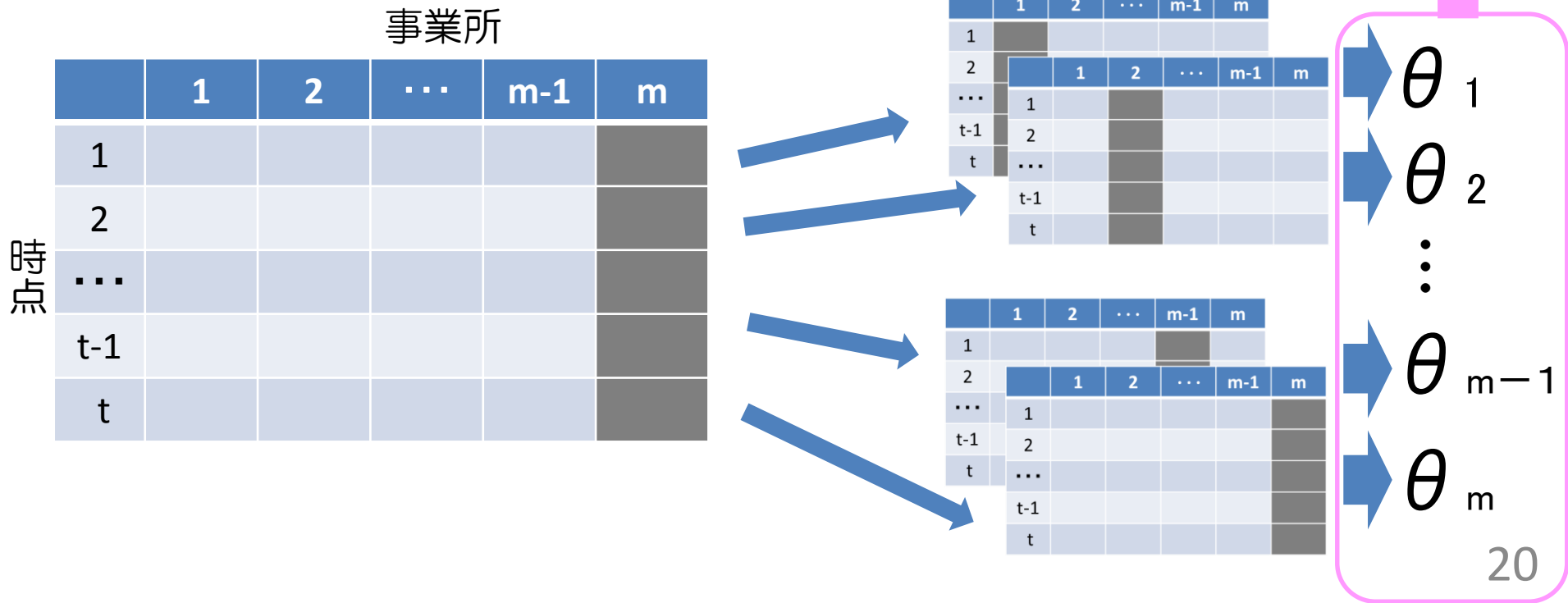
トラックシャシー（完成車を含む） - けん引車

特殊自動車

トレーラ

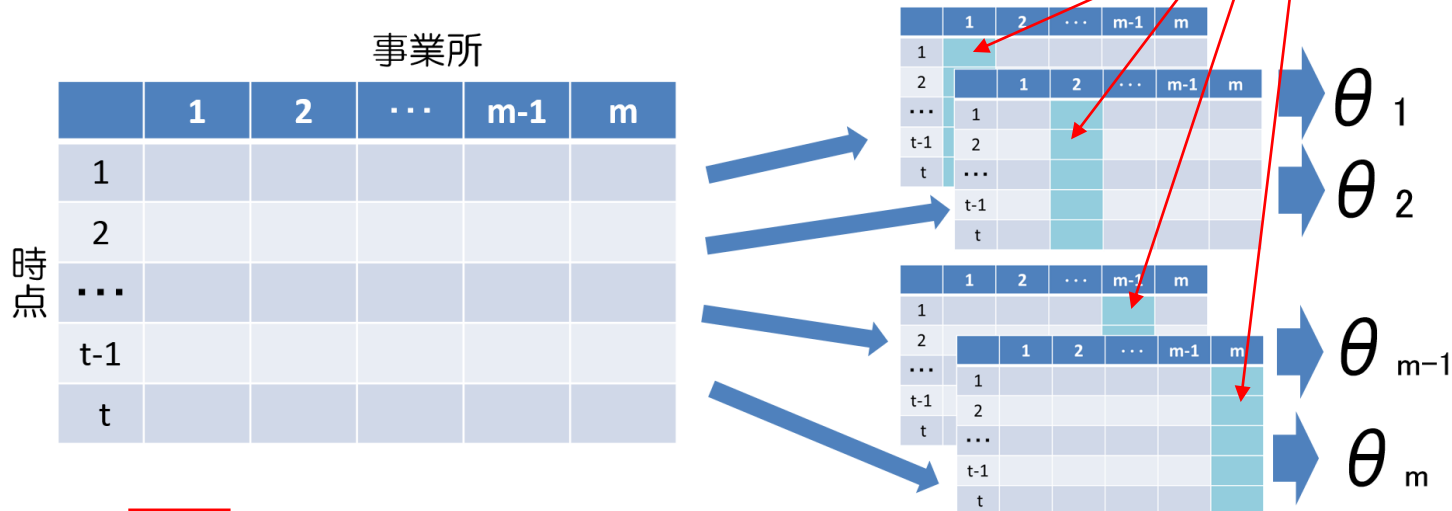
シミュレーションによるクロスバリデーション

- ① ロテーションで1事業所のデータを全期間欠測させた複数(m個)のデータセットを作成し、各データセットの平均や総和、分散などの統計量 (θ) を計算
- ② 上記①で求めたm個の統計量から母数($\hat{\theta}$)や標本誤差(S_{θ})等を計算



シミュレーションによるクロスバリデーション

- ① ロテーションで1事業所のデータを全期間欠測させた複数(m個)のデータセットを作成し、各データセットの平均や総和、分散などの統計量 (θ) を計算
- ② 上記①で求めたm個の統計量から母数($\hat{\theta}$)や標本誤差(S_{θ})等を計算



θ \longleftrightarrow $\left[\theta_1 \ \theta_2 \ \dots \ \theta_{m-1} \ \theta_m \right]$

この差を比較することで、補完のパフォーマンスを検証

パフォーマンスを確認した補完方法

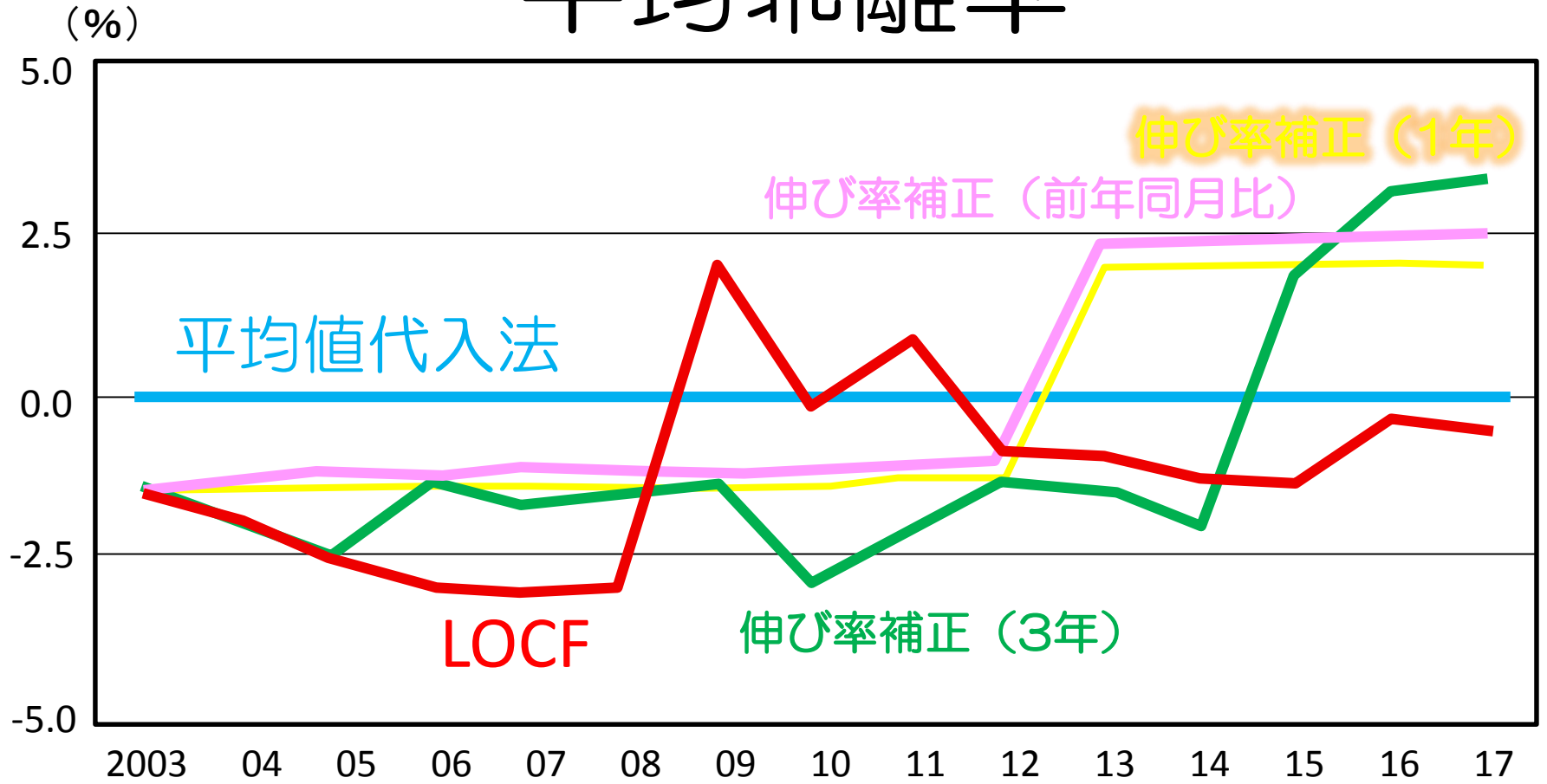
- (1) LOCF（現状）
- (2) 平均値代入法
- (3) 年間の伸び率によるLOCFの補正
- (4) 3年ごとの伸び率によるLOCFの補正
- (5) 前年同月比を使用したLOCFの補正

平均乖離率

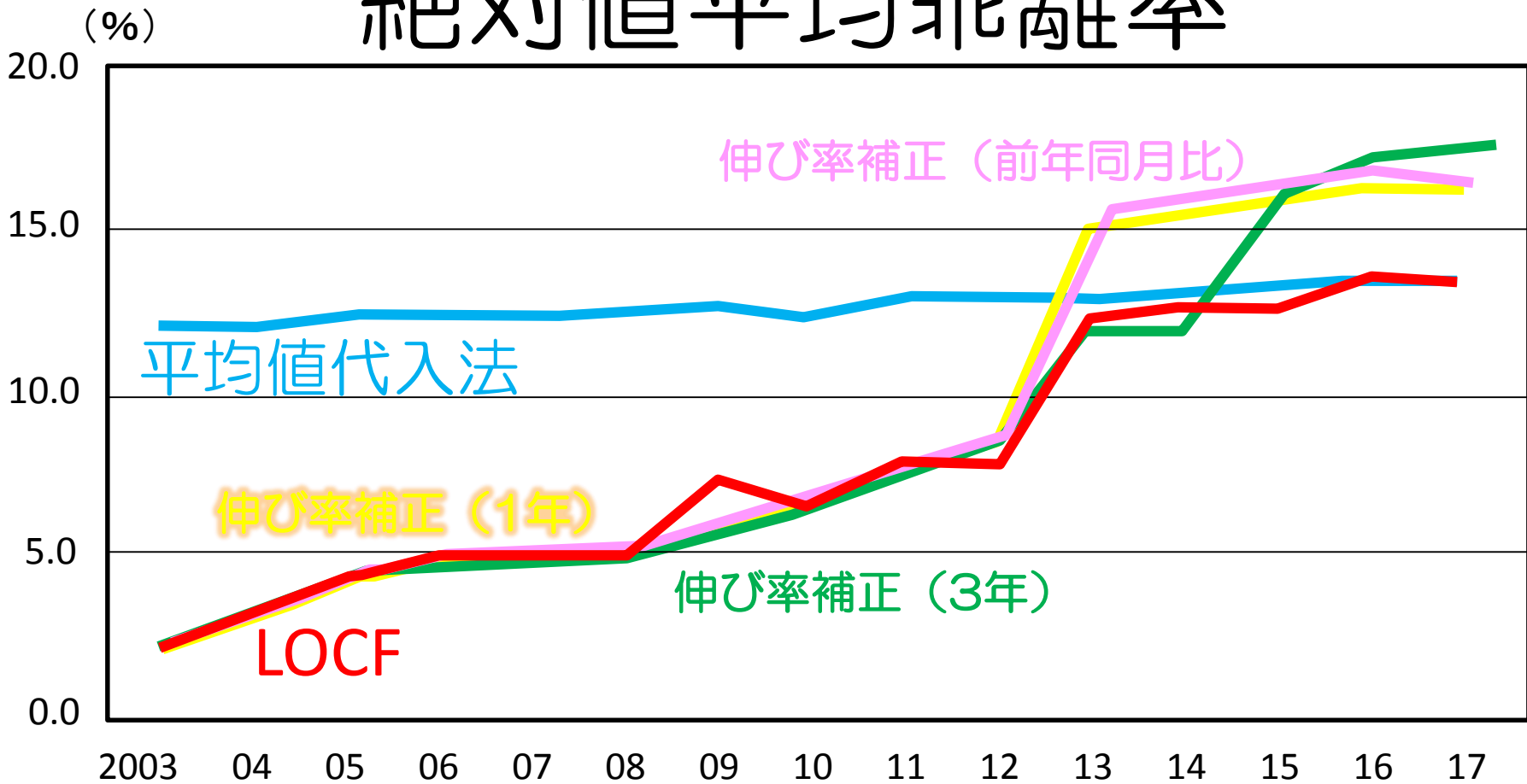
絶対値平均乖離率

でパフォーマンスを確認

平均乖離率



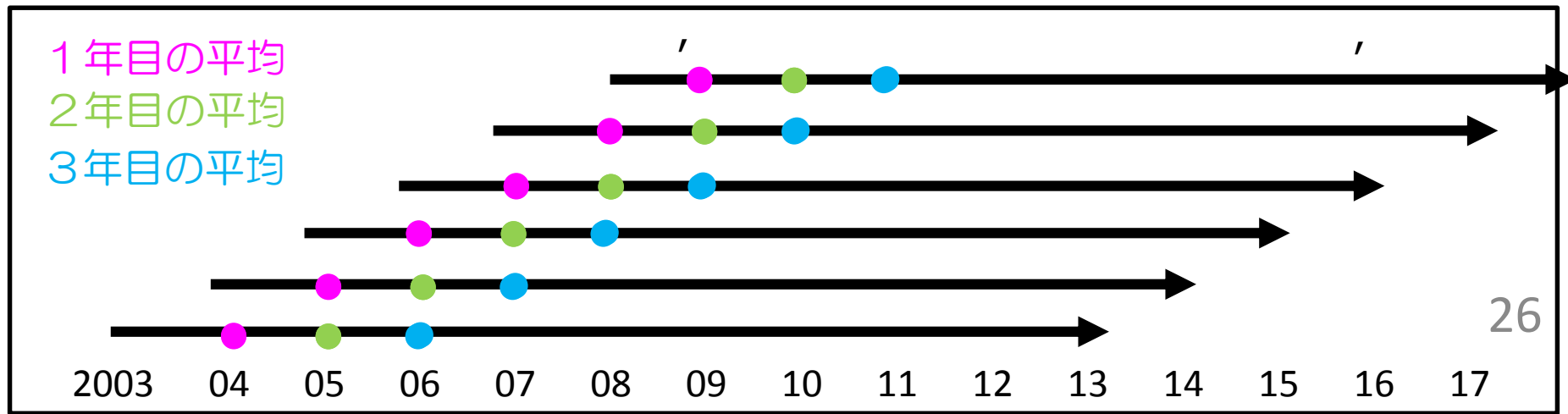
絶対値平均乖離率



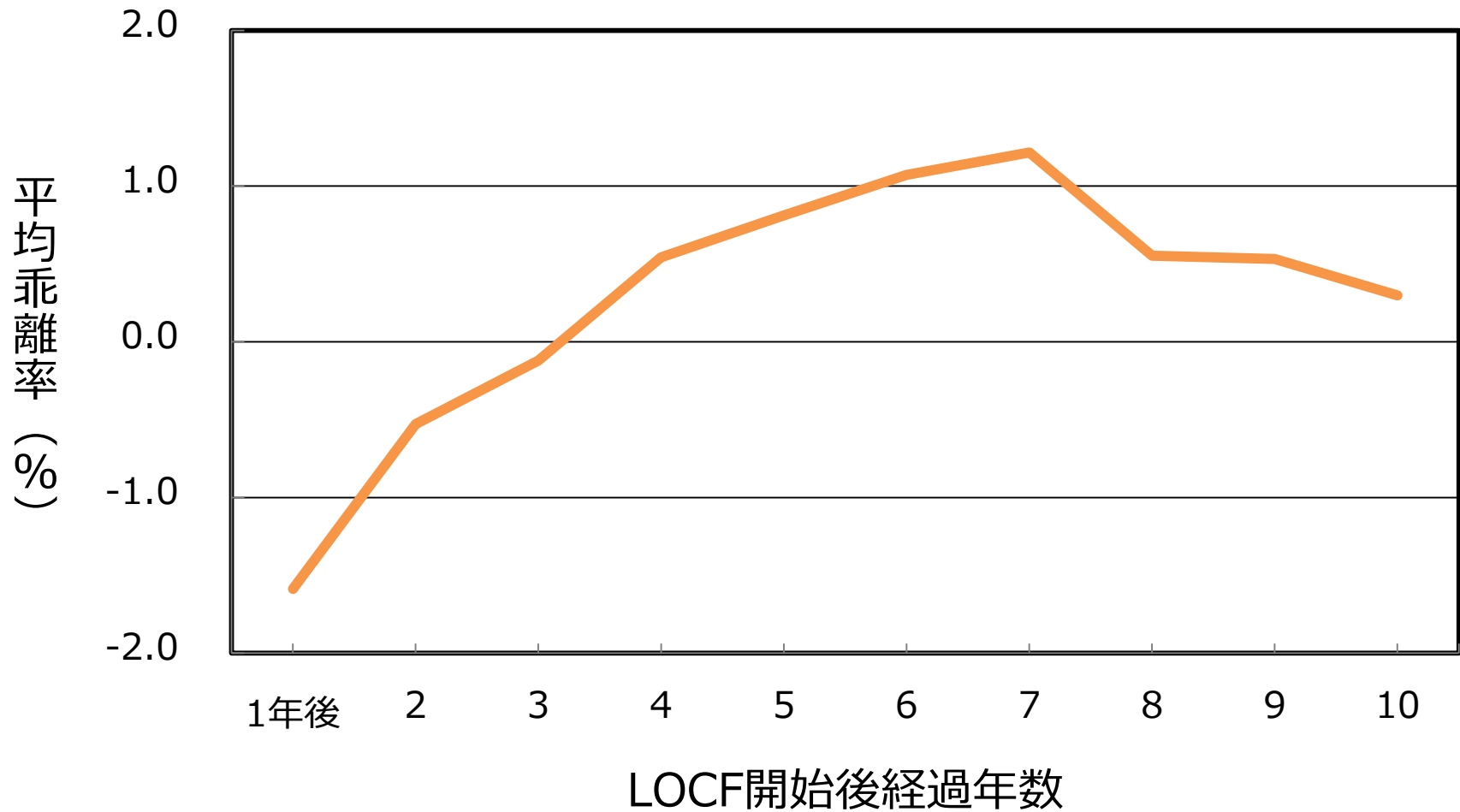
- 自動車等の調査票情報について、現在採用されているLOCFは、他の方法と比べてもパフォーマンスは劣らない
- 品目ごとに見ても、おおむね同様の結果
- ただし、急激に金額の水準が変化している一部品目では、LOCFのパフォーマンスが低下する傾向及び補正を行うことでLOCF開始時の水準までパフォーマンスが改善

ウィンドウ分析

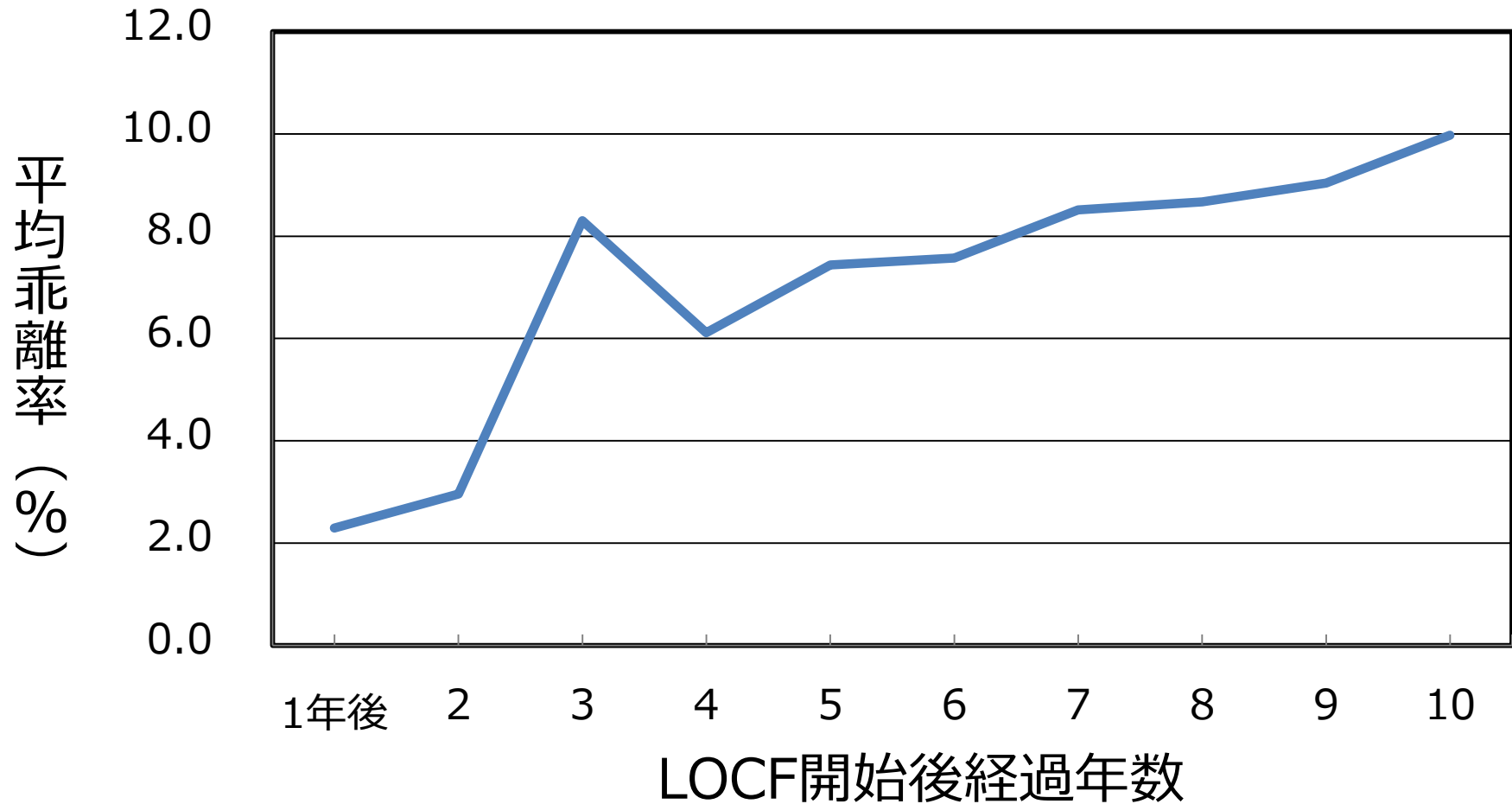
LOCFの開始年 $a = \{2003, 2004, 2005, 2006, 2007, 2008\}$ を1年ずつずらした10年間のシミュレーションを、2003年開始から2008年開始まで6セット行い、LOCFの開始後の経過年数ごとの平均乖離率の平均を求めた。



ウィンドウ分析（符号付き平均乖離率）



ウィンドウ分析（絶対値平均乖離率）



- 時間が経過すれば、乖離は拡大するものの、一部の特殊な品目を除き長期的に乖離が極端に拡大する傾向は観測されなかった。
- 数年程度であれば、一部の特殊な品目を除き、LOCFは有効に機能

考えられること

- i) 自動車等に含まれる品目のうち、金額の変動が大きい品目であって補完期間が5年に至るものについて、5年毎に実施される経済センサス-活動調査を活用し、経済センサス-活動調査の当該事業所の伸び率等でLOCF値を補正するなどの対応を行う。
- ii) 自動車等に含まれる品目のうち、複数年に渡り金額の大きな変動が続いている品目については、2年以内に同じ品目内の伸び率等を乗じてLOCF値を補正し、併せて、5年毎の経済センサス-活動調査の当該事業所の結果を用いて補正するなどの対応を行う。

考えられること

- 今回の検証は、検討の端緒として自動車等の調査票に限ってクロスバリデーション法を応用。したがって厳密には他品目にこの成果は応用できない。
- 一方、経済産業省生産動態統計調査の全調査票に対し検証を行うには時間がかかる。そして、経済産業省生産動態統計調査の結果作成は絶えず行われており、自動車以外の他の調査票に対しても、期限を定めないLOCFが継続されている状況にある
- 自動車等の検証結果は、常識的かつ包括的な考え方での対応

検証を行っていない品目についても、当該ルールを暫定的に導入して対応することが考えられる

ご審議をお願いしたい論点

- 安定した推移をしている間、LOCFの値を5年毎に見直すことでよいか
- 自動車以外の他の品目についても、安定した推移をしている品目について、暫定的に5年毎に見直すことでよいか
- 欠測値補完について、何かアドバイスはないか