



経済経営現象理解とEBPMのための 複数データの融合手法の開発と応用

慶應義塾大学 経済学部 教授
(兼)理研AIP経済経営情報融合分析チームリーダー
星野崇宏
hoshino@econ.keio.ac.jp

EBPMとは？なぜビッグデータが必要？

Evidence Base Policy Making (EBPM) 証拠に基づく政策(立案)意思決定

政策立案者/政治の恣意性を排除

透明性の確保 緊急事態への即応性

我が国で行われてきたアプローチ

ロジックモデル

課題からインパクトをボックスモデルであらわす

必ずしも定量的な評価ではない

世界標準となっているアプローチ(EU,OECD諸国)

計量経済学的手法による厳密な政策効果推定

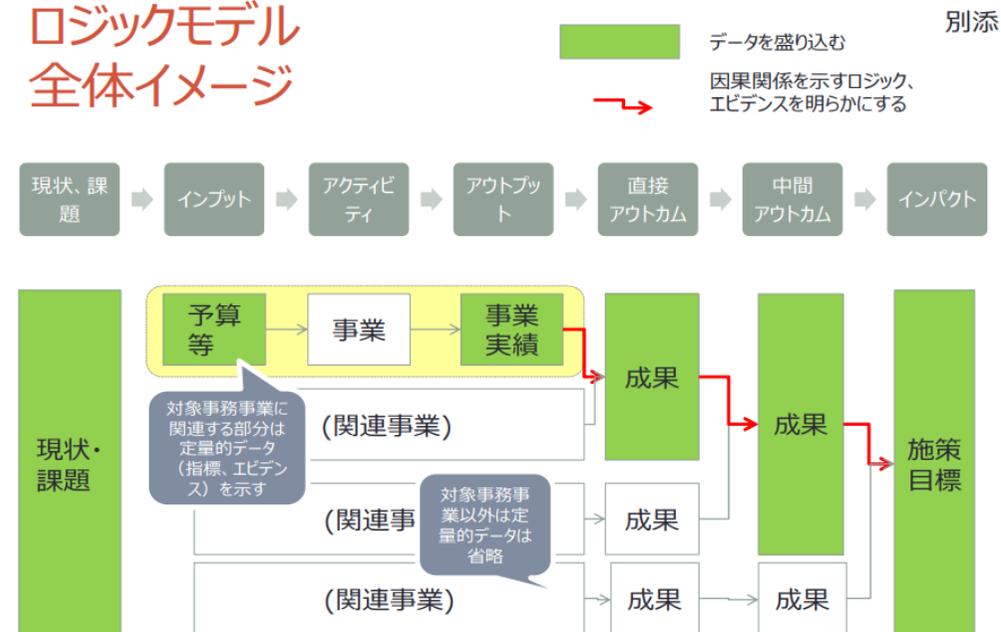
例)内生性・逆因果の排除・一般化可能性

⇒高精度のデータを正しく分析する必要

出典:厚労省HP

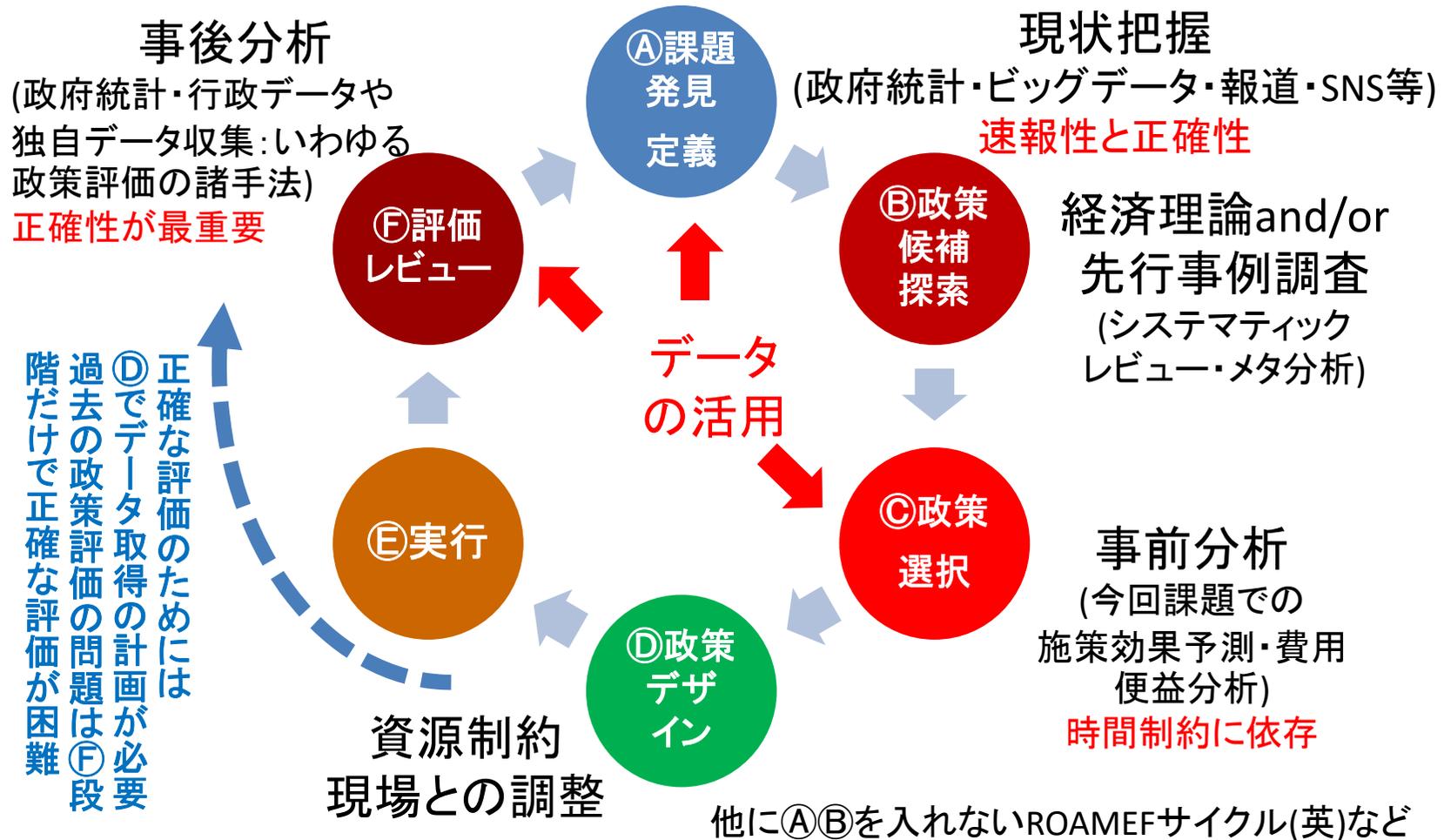
https://www.mhlw.go.jp/jigyo_shiwake/dl_h30/mtg0508_07.pdf

ロジックモデル
全体イメージ



EBPM(エビデンスに基づく政策立案評価)のサイクル

さらに細かく分ける場合もあるがここでは以下の6サイクルとする
過去の”政策評価”は⑥に限定:EBPMは①~⑤も含む幅広い取組



各サイクルでの利用データの要件

①課題発見・定義 における「現状把握」

速報性と正確性の**バランス**

■ 政府統計:速報値であっても2か月程度の遅延

すべてデータがそろえば正確(正解) * 必要な情報があるか?

■ SNSや位置情報などビッグデータ:速報性が高い

⇒**様々なバイアス**(後述) 解決方法は存在

②政策選択肢候補からの選択 における「事前分析」

時間制約による コロナ等急激な情勢変化には速報性が重要

③評価レビュー における「事後分析」

正確性 >> 速報性

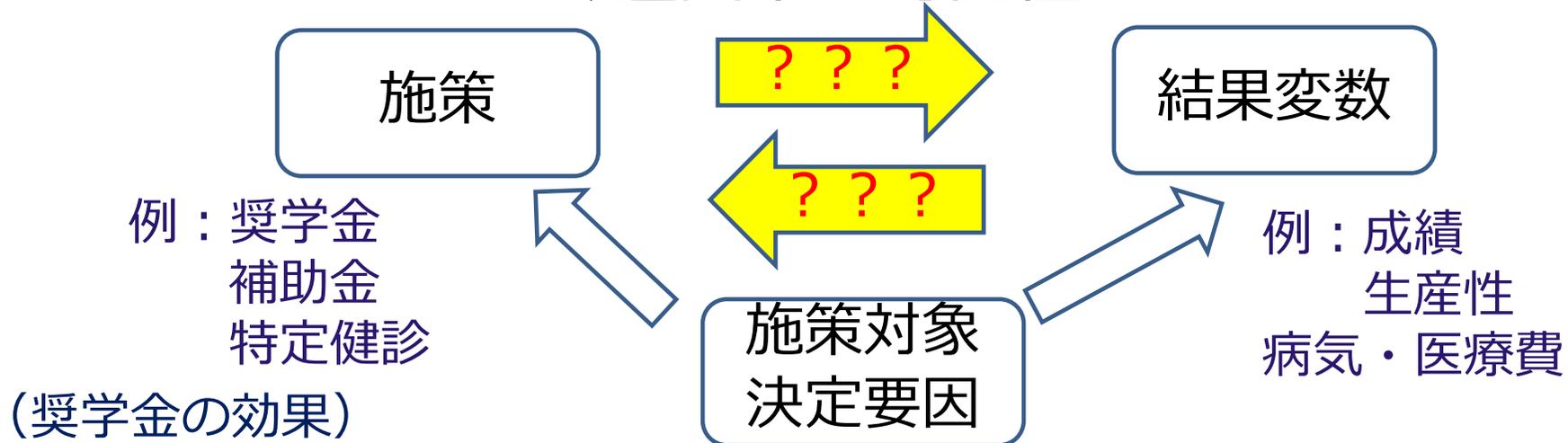
欧米はもとより途上国の政策でも世銀等が“**頑健な評価結果**”要求

評価の方法は政策の内容や対象によりほぼ明確に確定

* 正確な評価のためには④の政策デザインからの計画が必要

いわゆる**因果効果推定を可能にするデザイン**

逆因果の可能性



もともと奨学金をもらえる学生が優秀なのでその後の成績も高いだけ？

(企業の補助金)

補助金の対象になるのは体力のない企業

対象にならない企業より生産性が低いのは当たり前？

(無料の特定健康診断)

健康診断を受けるのは健康に不安がある人？

もっと健康を気にしている人は人間ドックを受ける？

真の効果を知りたい！ 施策の因果的な効果を調べる政策評価手法が利用

補助金の効果検証

Decreamerら(Small Bus Econ)

ベルギーのフランドール地方の中小企業への
補助金施策の効果

- 特定の閾値前後で施策対象が決まる
- 計量経済学の方法(回帰不連続デザイン)で検証
- 特に10人以下の小企業には補助金の効果

⇒データベース化により効果検証は容易

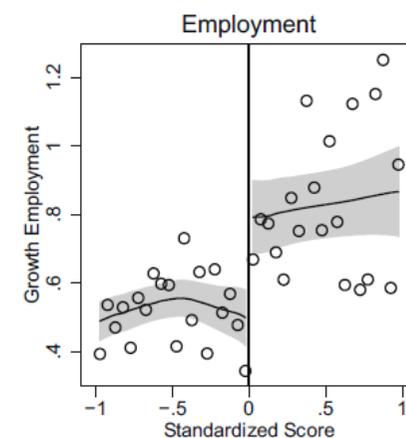
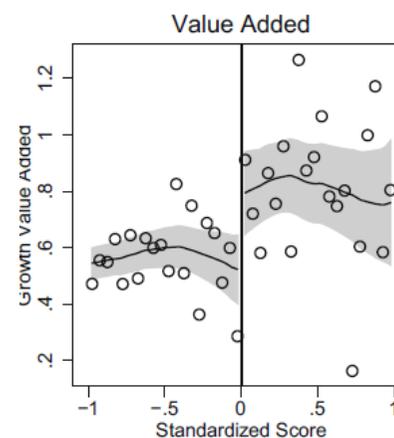
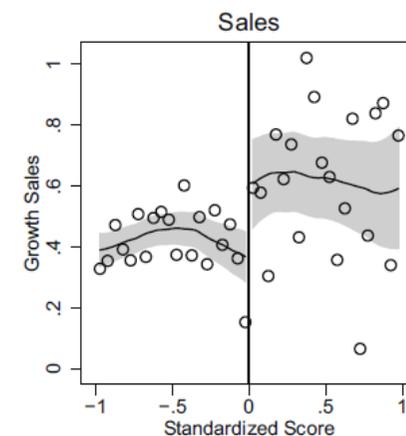
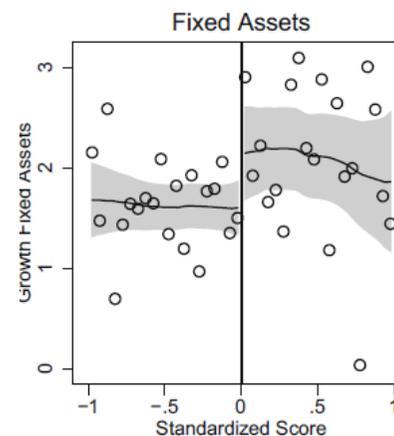
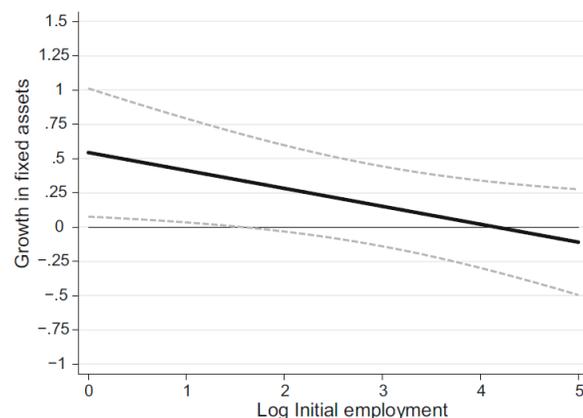
下記図は初期雇用人数での異質性

⇒施策への示唆

国内では経産省

⇒デジタル庁移管の

GビズIDに期待



どのようなデータが利用できるか

集めるデータ: 事前に計画して収集したデータ

集まるデータ: 実務など別目的で収集された(ビッグ)データ

	代表性	速報性	コスト	取り方	例
集めるデータ	高い	低い	高い	質問	市場調査
集まるデータ	低い	高い	低い	ログ(事実)	Web閲覧ログ

【企業が意図的に収集するデータ】 集めるデータ

顧客調査・市場調査・マーケティングパネルデータ(SCIなど)

【企業活動で収集されるデータ】 集まるデータ

POSデータ(レジでの売り上げ情報)・検索履歴データ

SNSの書き込み情報・位置情報 など

【政府統計や公的統計調査】 集めるデータ

国民や企業を抽出して回答を依頼

【行政データ】 集まるデータ

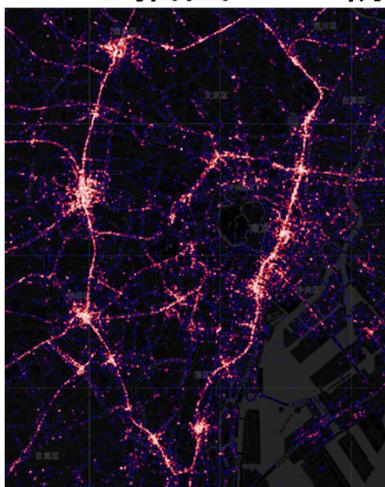
行政記録(税務データ、国勢調査、マイナンバーなど) 診療報酬情報

* 網羅的(全数)・情報取得目的では収集されていない(変数定義の問題)

(集まる)ビッグデータの例

例として講演者が実際に分析しているもの
位置情報データ

GPSから描画 朝



夕方



POS(Point of purchase)データ

年間40億件
国内7%
程度補足



口座連動型の家計簿アプリ(1000万人) 診療報酬情報(レセプト1000万人)



公的統計分野へのビッグデータ活用

【オーストラリア統計局】

オーストラリア政府統計での消費者物価指数

家計調査に加えてPOSデータ・一部耐久消費財でWebスクレイピングも

* 日本では家計調査と小売物価統計調査

GIS・位置情報と衛星、環境データを使って産業活動と環境統計作成

⇒2年で100億ぐらいかけて種々のデータを整備

【国内の例】

総務省 消費動向指数研究協議会

経産省「ビッグデータを活用した新指標開発事業」

METI POS小売販売額指標

METI × NOMURA コンシューマーセンチメント・インデックス

ナウキャスト:JCBカードデータでの消費動向指数作成

サイバーエージェントAI Lab 森脇氏

ハローワークへの訪問をGPSを使って労働統計の予測

労働力調査は2か月程度のラグ

米 Opportunity Insightsの取り組み

ハーバード大学Raj Chettyらのグループのデータベース

コロナ禍前から「労働経済学・教育経済学」の観点から全米ZIP単位でデータ整備

クレジットカードデータ(消費・企業決済)・給与計算システム／オンライン人材募集／スケジュール管理ツール(労働)・クラウド会計(事業・企業取引)＋公開された公的統計

出典: Opportunity Insights HPより

(いくつかの発見)

①2020年前半のロックダウンがどう

経済に影響を与えたのか？

⇒富裕層がリアル消費を抑え

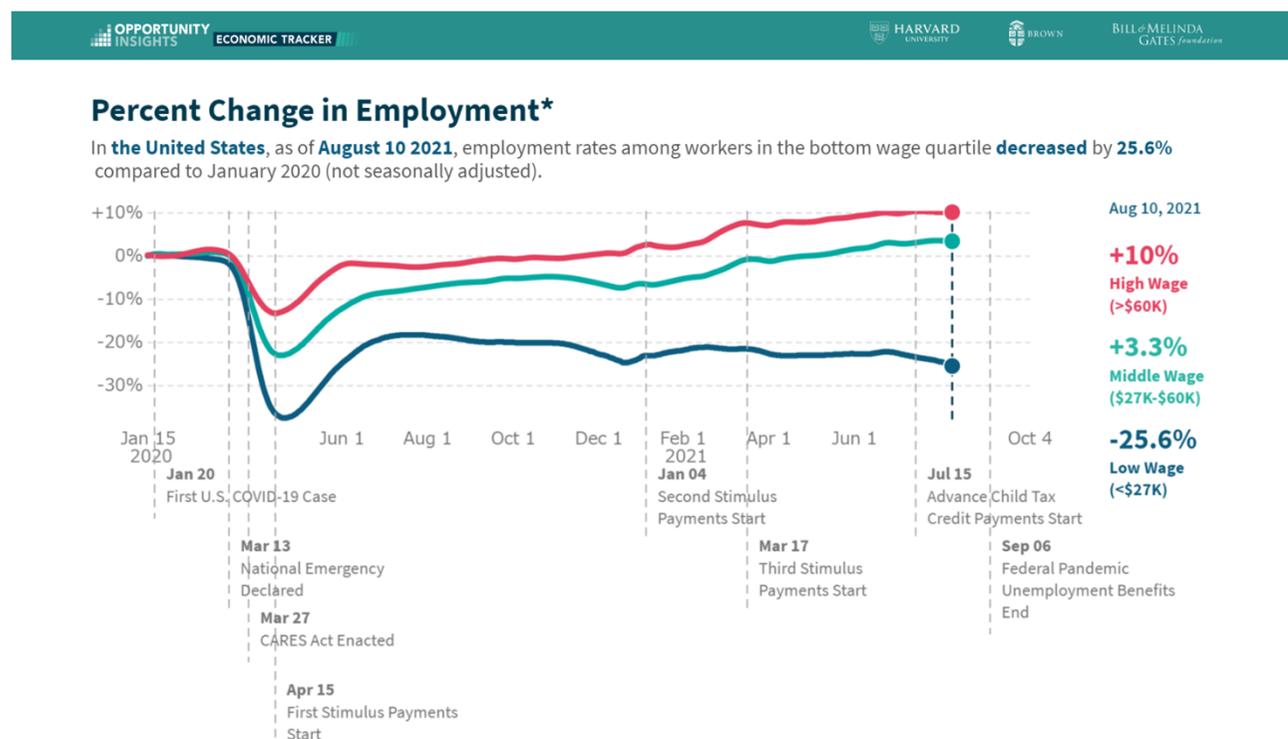
他の層の雇用に波及

②米のコロナ特定給付の効果

③コロナ禍での景気のK型回復

④ワクチン接種率の州間格差と

経済への影響の関連



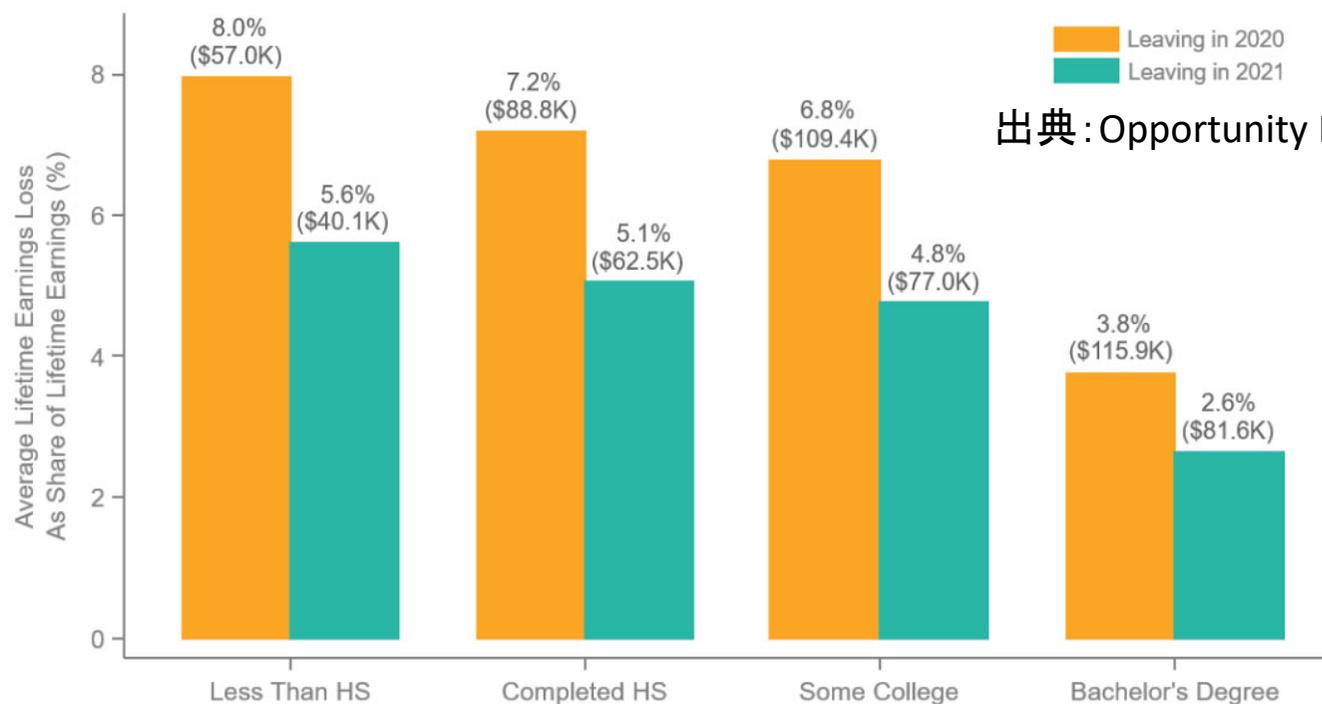
米 Opportunity Insightsの取り組み

コロナ禍において卒業・就職した学生の生涯賃金の損失分

就職初期での経験(人的資本)が形成されないことが大きな損失に

特に低学歴程影響が大きい 短大・高卒でも7%程度に

STUDENTS LEAVING SCHOOL AND ENTERING THE LABOR MARKET DURING THE RECESSION WILL EXPERIENCE LASTING WAGE LOSSES



出典: Opportunity Insights HPより

Data Source: [Opportunity Insights Economic Tracker](#), ACS, and SIPP

(企業に集まる)ビッグデータの注意点

ビッグデータには特定のバイアスがつきもの

速報性があるからと単純に利用すると大きな問題(Ⓐの現状把握でもダメ)

(1) 政府統計と測定値の定義が違う【測定誤差】

例) 別府市の滞在人口増減率(右図)

* すでに総務省統計委員会担当室WP(2019)

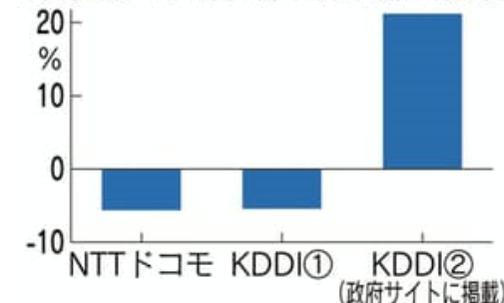
「東京都における流動人口データの有効性の検証」

で“電車や車での通過”も入れてしまう問題が指摘

例) クレジットカードでコロナの消費予測

現金払いが無視 高額支払いに偏る 収入分布?

別府では市長が抗議する事態に
(平時と比べた4月27日の滞在人口の増減率)



(注) NTTドコモは滞在時間を基に加重平均、
KDDI①は15分以上滞在した人口、KDDI
②は15分未満の滞在も含んだ人口
(政府サイトに掲載)

(2) 自社ユーザーしか取れないバイアス【代表性の問題】日経6/14”位置データのクセに注意
「密」集計、実態とズレも”

例) NTTドコモは企業契約端末が多く個人の割合低い

格安携帯は年齢構成が若い、など

例) iPhoneとAndroid, Amazonと楽天 利用者層が異なる

(3) 属性等が不明なことが多い 例) 位置情報単体では年齢家族構成不明

いくらボリュームが多くても偏ったデータ

ビッグデータのバイアスを除く努力

	代表性	速報性	コスト	取り方	例
集めるデータ	高い	低い	高い	質問	市場調査
集まるデータ	低い	高い	低い	ログ(事実)	Web閲覧ログ

集まるビッグデータの利点: 低コスト・速報性・実際の行動であること

欠点: 代表性

政府統計など調査の利点: 代表性

欠点: 質問であり精度問題 速報性が低い

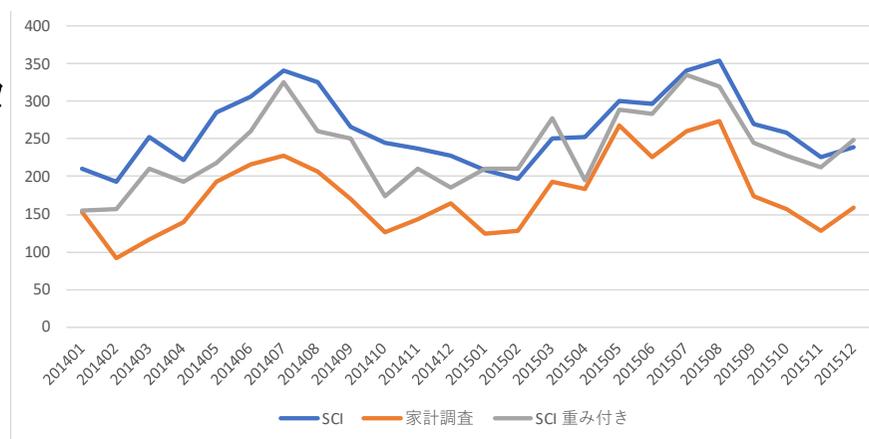
⇒統計手法により両者を融合: 代表性あり精度が高く速報性がある

発想は政府統計側の属性(年齢・収入等)の分布に合うように補正する

例) 家計調査 + 購買スキャンデータの統計的融合

* 家計調査は過小報告バイアス

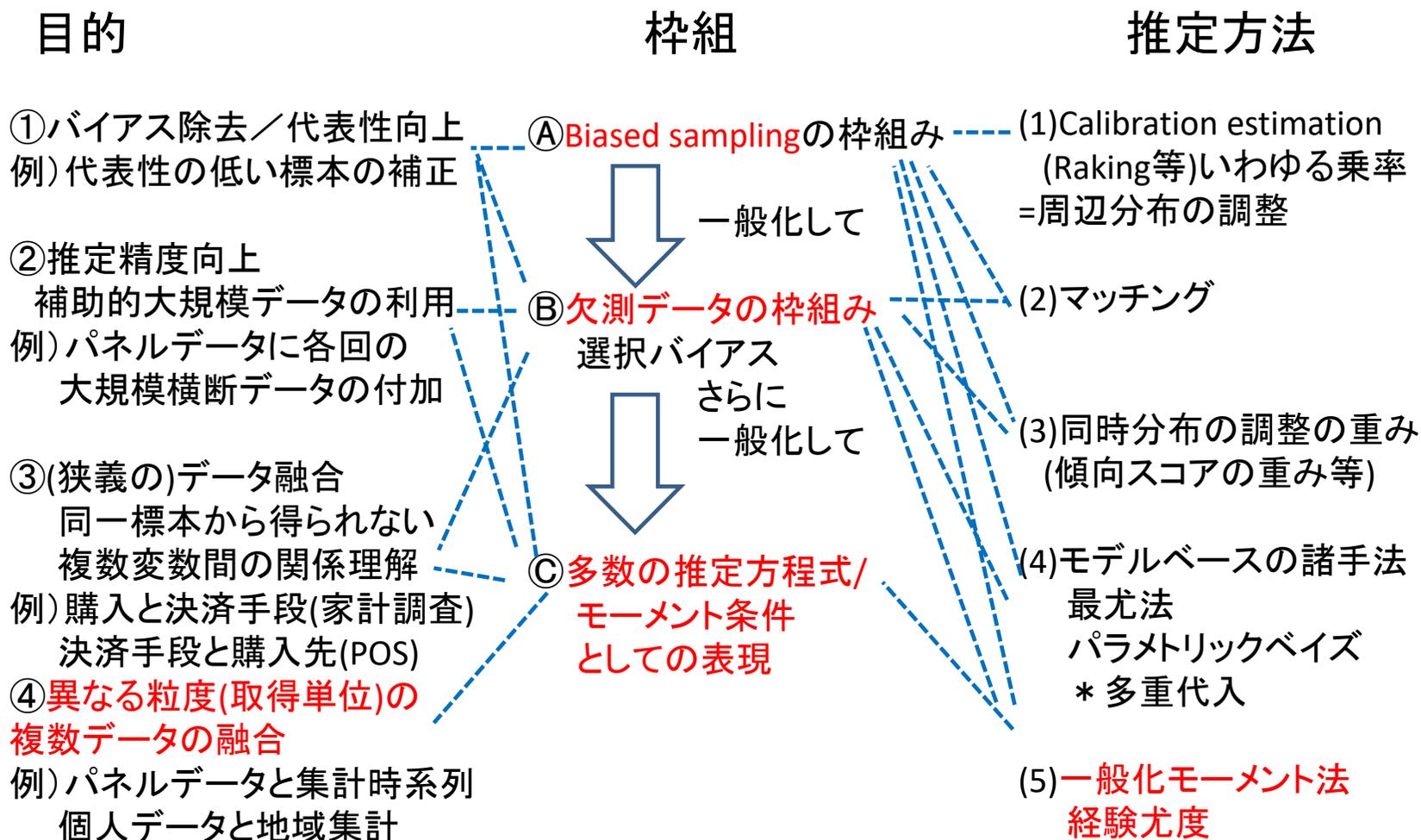
* 購買スキャンは対象者バイアス



複数データ融合の
欠測データとしての
統一的枠組み

および我々の方法論開発

複数データを融合する目的/枠組み/手法



注意) ④でバイアスなどは考慮しない/バイアス発生の特定化が可能な場合には尤度表現(階層モデル)も可能(仮定強い)

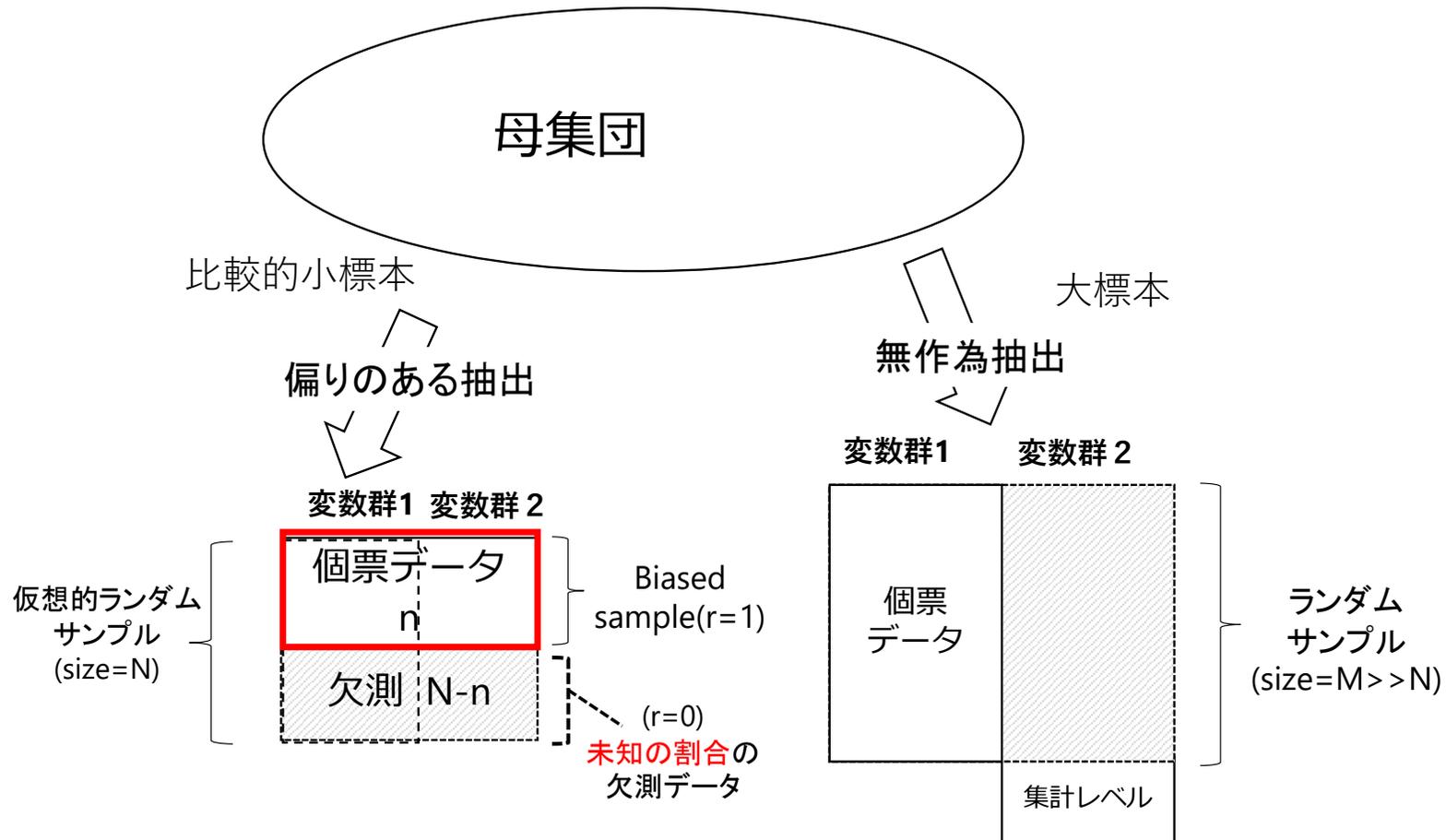
最近行った方法論開発

- 外部補助情報がある場合の因果推論(Takahata & Hoshino,2019)
- Biased sampling状況での推測(Shimizu & Hoshino,2019)
- NMAR状況でのガウス過程を用いたデータ融合(Mitsuhiro & Hoshino, under rev.)
- ミクロデータとマクロデータの融合(Igari & Hoshino,2018;in submission) 個票ではなく分布情報・モーメント情報を利用

- 不完全なデータ設定での因果推論
特に近年は機械学習分野で研究されている状況設定
 - いわゆるSeparate Labelでの因果効果推定(Shinoda & Hoshino,2021)
 - Positive Unlabeledデータでの因果効果推定と生存時間モデル(Toyabe, Hasegawa & Hoshino,2021; in submission)

Biased samplingの欠測データ表現

Biased samplingは欠測データ表現が可能
但しどの部分がランダムサンプルであるかを十分考慮する必要



政府統計等複数データの融合のための方法論開発1

【外部周辺情報が利用可能である場合の母数推定についての理論開発】

因果推論の手法開発

共変量と未介入時の結果変数の周辺情報が利用可能であるとき、**因果効果の異質性を識別・推定**することが可能なことを証明しベイズ推定法を開発 (Takahata&Hoshino, 2019)

Biased sampling下の推測

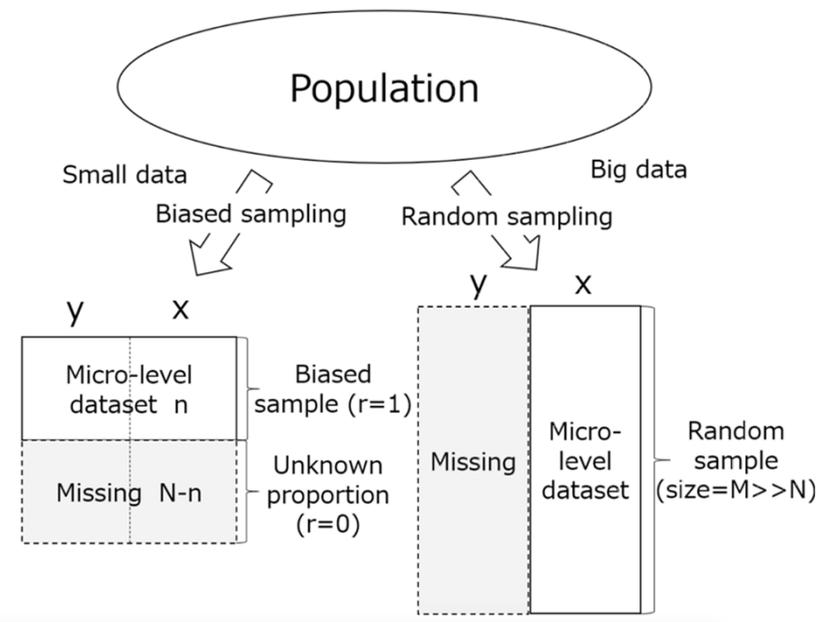
元データ(共変量+結果変数)が母集団からのランダムサンプルではないが

共変量のみ観測されているサイズのサンプルが利用可能

⇒それらを融合して**biasを補正**する手法の開発

(Shimizu&Hoshno, 2019)

政府統計や外部ビッグデータとの融合への応用について現在実施中



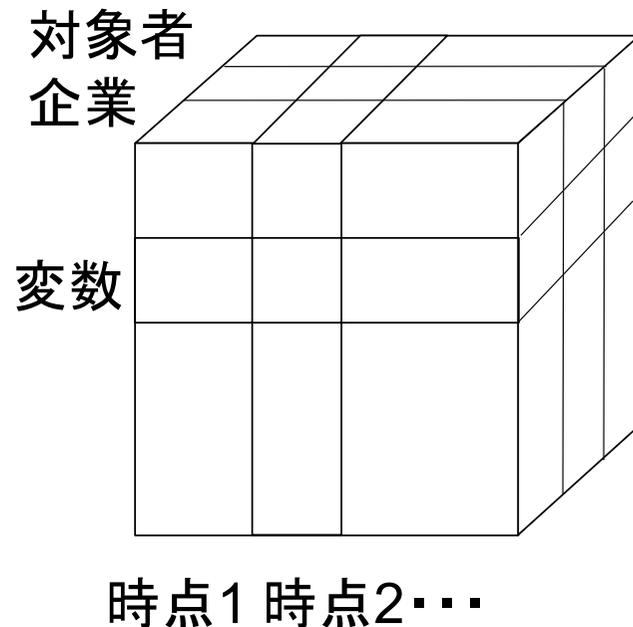
異なる粒度のデータ間の統合

粒度が異なるデータを統合するニーズも存在

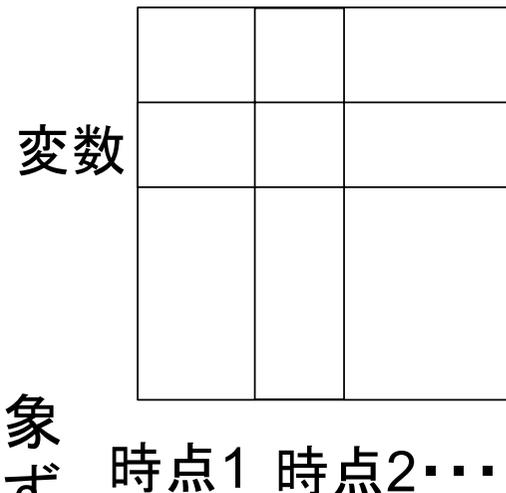
この場合には一般にモーメント情報・推定方程式のセットとして考慮する必要

⇒GMM(Generalized method of moment)やEmpirical likelihoodの利用例)生産や出荷側の時系列は網羅的 調査はバイアスがある

パネルデータ(バイアスあり)



集計時系列データ
(バイアスなし)



調査対象
一致せず

代表性ある集計統計と偏った大規模データの融合

政策の効果（例：キャッシュレス施策, 投稿中）

「どんな消費者が」「どの業態が」恩恵を受ける？

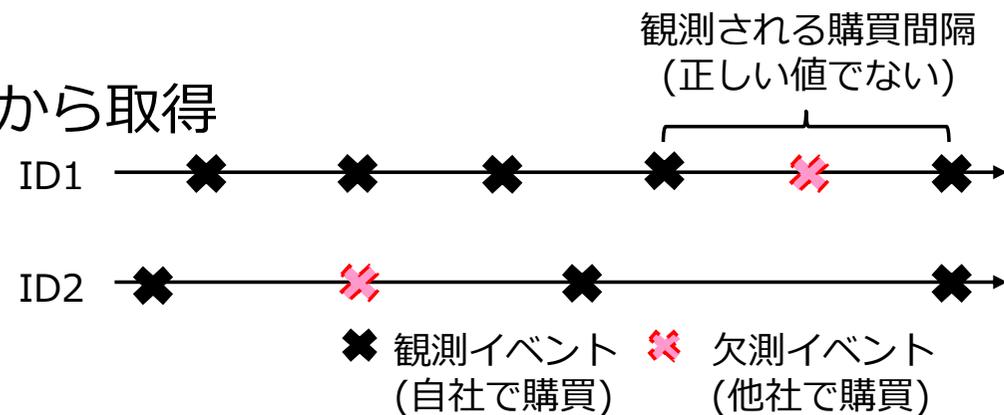
ビッグデータは小売り側から取得

（他社での利用は不明）

偏りがあるが

支払先わかる

* 家計調査では不明



マクロ情報は代表性のある標本から：ただし**集計情報**

例) 小売企業の販売シェア、決済比率情報など

偏ったミクロのビッグデータだけでは政策効果を大きく過小評価

⇒ミクロとマクロの融合の必要性

政府統計等複数データの融合のための方法論開発2

【中間欠測のある継続時間分析への応用】

中間欠測では欠測インディケータ自体が観測されない

→ 観測データのみから推定を適切に行うことは難しいことから、**マクロデータを融合**することで適切に母数を推定する

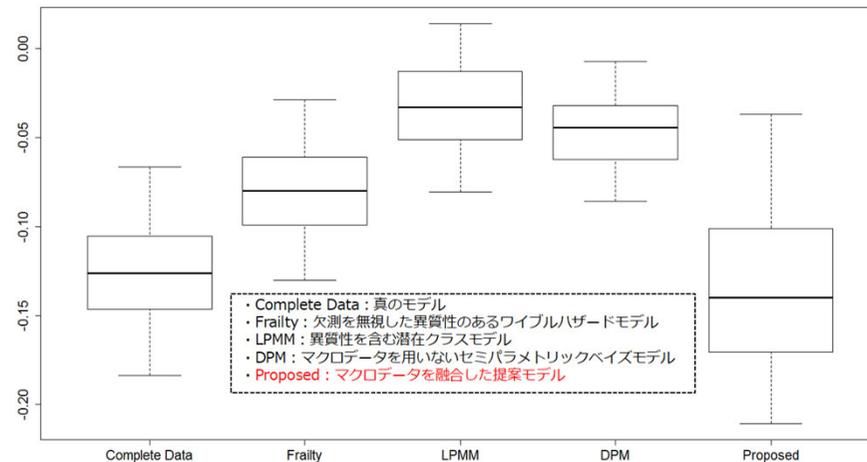


$$q(\theta|y) \propto \underbrace{p(y|\theta)}_{\text{尤度}} \underbrace{\exp\left\{-\frac{n}{2}L_n(\theta)\right\}}_{\text{モーメント制約}} \underbrace{p(\theta)}_{\text{事前分布}}$$

= ミクロデータ部分 = マクロデータの融合

尤度とGMMの良いところを組み合わせ、データ部分は尤度関数で表現し、理論制約のみをGMMで表現する(Hoshino&Igari,2018) 特にこの方法論を現在一般化(Hoshino&Igari, in submission)

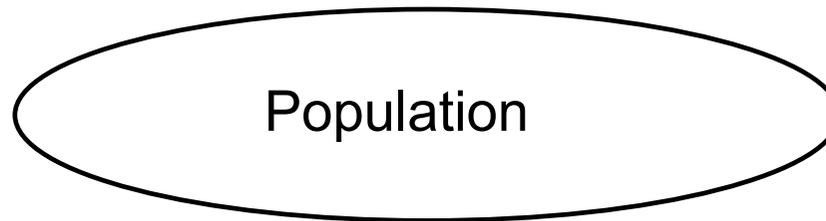
解析例：「価格」が購買間隔に与える影響 (Igari&Hoshino, 2018)
 バイアスを無視すると価格の影響を3分の1程度に過小評価
 これ以外に企業の意思決定や政府統計への応用を現在実施中



データ融合の別の例: **Separate Label**問題

アウトカムYと介入Zが同時に観測できない

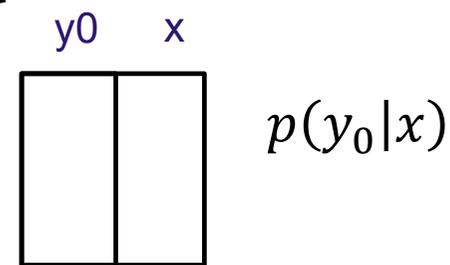
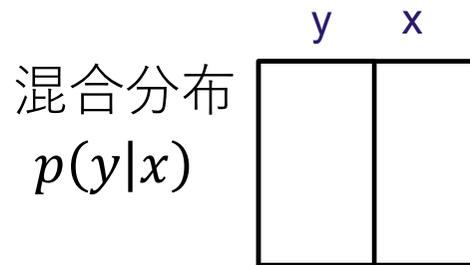
$$p(r|y_1, y_0, z, x) = p(r|x)$$



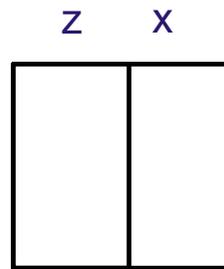
Random sampling (with inclusion probability dependent on x)

介入実施群(r=1)

介入無実施群(r=0)



データまたは計算規則 $p(z|x)$



実務的には実施は容易
(介入の一部実施)

Shinoda & Hoshino (AAAI2022)

Separate labelの問題で現実的なのは”こちらが割り当てても実際にやってくれるかはわからない”ノンコンプライアンスの場合
例) 失業者に職業訓練を割り当ててるが実際には受けない
例) 広告のターゲットに入札するが提示されるのはRTBで勝った時
この場合の推定すべき関数は比の形式

$$\mu(\mathbf{X}) = \frac{E[Y^{(1)}|\mathbf{X}] - E[Y^{(0)}|\mathbf{X}]}{E[D^{(1)}|\mathbf{X}] - E[D^{(0)}|\mathbf{X}]}$$

新規性

- ①識別: 本来は (Y, D, X) が必要だが、 $P(D=1|X)$ の確率が違うデータが2つあれば (Y, X) と $D=1$ の X だけで推定できる
- ②推定: 別々に推定すると精度が低下(Vapnik則)
一方傾向スコアの重みづけ的な方法は不安定
⇒直接重みづけ推定法(Direct Least Square)の開発

Shinoda & Hoshino (AAAI2022)

Shape	n	q_x	DWLS	IWLS	SEP	DLS
h_{con}	10K	1	0.65 ± 1.54	5.04 ± 16.5	1.80 ± 2.38	0.15 ± 0.20
		5	0.40 ± 0.91	1.30 ± 3.21	3.46 ± 2.71	0.93 ± 1.17
		10	0.65 ± 1.00	4.96 ± 14.4	4.18 ± 3.81	1.81 ± 2.82
	50K	1	0.22 ± 0.46	53.5 ± 372	1.04 ± 0.93	0.08 ± 0.13
		5	0.06 ± 0.07	0.35 ± 0.88	2.13 ± 2.00	0.41 ± 0.67
		10	0.10 ± 0.11	1.22 ± 6.07	3.52 ± 1.92	0.41 ± 0.52
h_{lin}	10K	1	0.09 ± 0.15	0.51 ± 1.25	0.25 ± 0.30	0.39 ± 0.30
		5	0.14 ± 0.08	1.32 ± 6.72	0.58 ± 0.41	0.88 ± 0.52
		10	0.61 ± 0.27	4.73 ± 22.4	1.39 ± 0.65	1.53 ± 0.75
	50K	1	0.02 ± 0.03	6.79 ± 53.1	0.11 ± 0.11	0.36 ± 0.33
		5	0.04 ± 0.03	0.12 ± 0.16	0.36 ± 0.24	1.04 ± 0.53
		10	0.32 ± 0.10	11.9 ± 112	0.76 ± 0.26	1.61 ± 0.89
h_{log}	10K	1	0.19 ± 0.26	0.72 ± 1.82	0.30 ± 0.31	0.32 ± 0.29
		5	0.29 ± 0.23	12.1 ± 113	0.58 ± 0.40	0.97 ± 0.56
		10	0.72 ± 0.31	2.76 ± 7.65	1.11 ± 0.49	1.40 ± 0.72
	50K	1	0.06 ± 0.08	0.30 ± 0.88	0.22 ± 0.16	0.32 ± 0.30
		5	0.10 ± 0.04	0.18 ± 0.13	0.56 ± 0.31	0.75 ± 0.62
		10	0.30 ± 0.10	55.3 ± 550	0.86 ± 0.28	1.20 ± 0.99

Table 1: The mean and standard deviation of the MSE over 100 trials with $\gamma = 0$. The results are multiplied by 100 (constant), 10 (linear) and 10 (logistic), respectively. The bold face denotes the best and comparative results according to the two-sided Wilcoxon signed-rank test at the significance level of 5%.

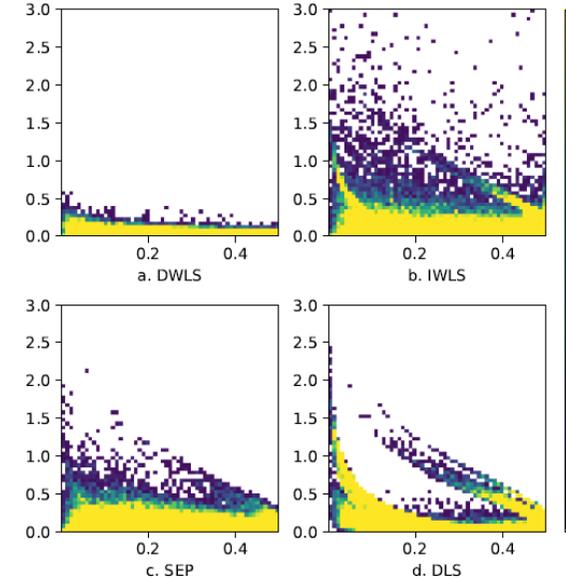


Figure 2: Relation between the squared error (Y-axis) and the PSD (X-axis) when $\mu(\mathbf{X})$ is linear, $q_x = 5$ and $n = 10000$. Each point is coloured according to the spatial density.

Shinoda & Hoshino (2021) “Estimation of Local Average Treatment Effect by Data Combination” arXiv

**複数調査の統計的融合のため
の段階推定法の開発と応用
家計構造調査の年次推定への応用**

**慶應義塾大学大学院経済学研究科
松永将志・二荒麟・清水祐弥・
総務省統計局 柴田卓也
との共同研究**

前提となる標本の特徴

全国家計構造調査（旧全国消費実態調査）の長所・短所

- サンプルサイズが大きい(約90,000世帯)
→詳細な区分での分析が可能
- 2か月(前回3か月) 5年に一度

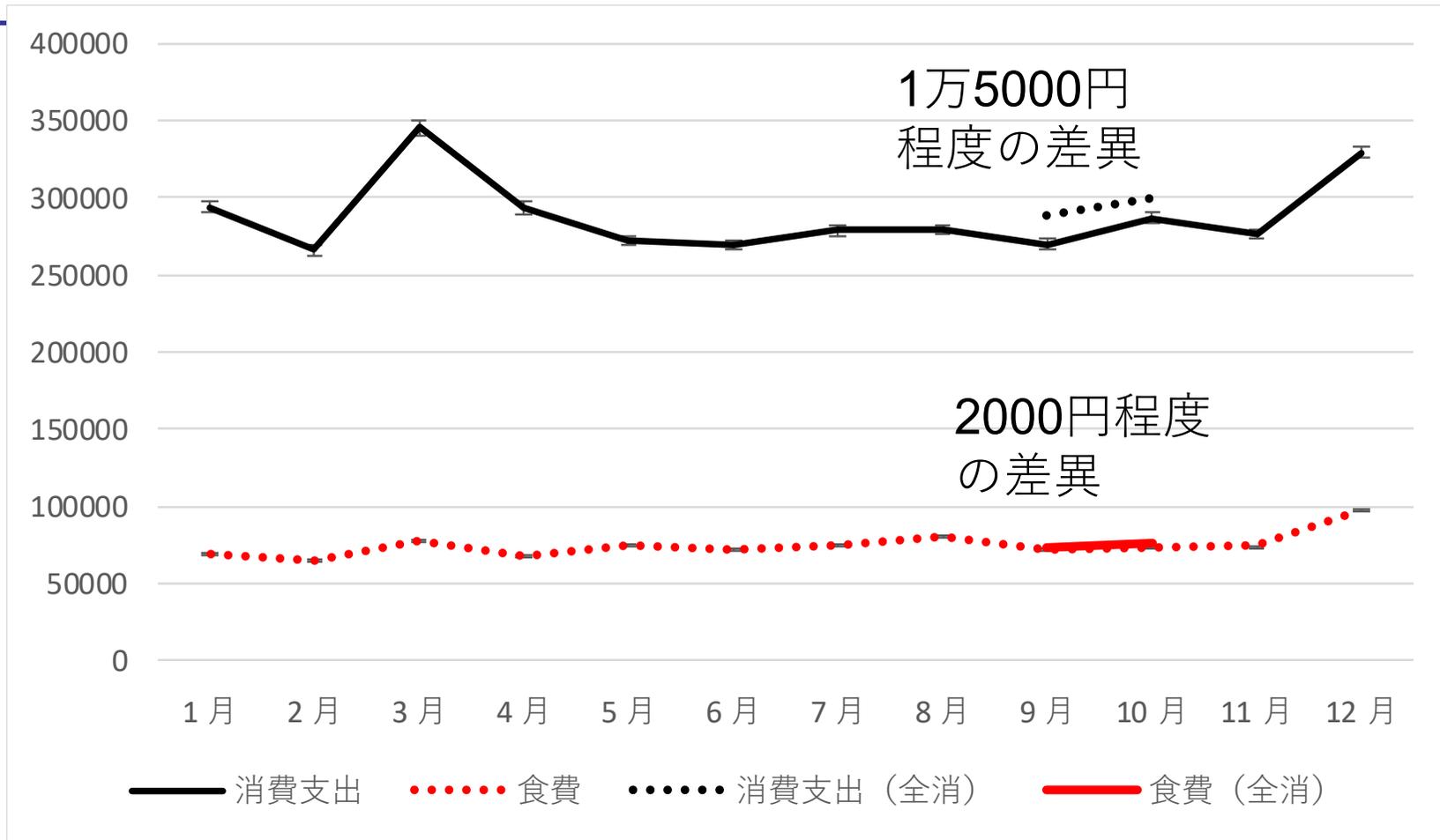
家計調査の長所・短所

- 通年で結果が得られるローテーションパネル
 - 比較的サンプルサイズが小さい(約9,000世帯)
- ◆ 目的：食費など月次の変動が大きいものについて両者を融合させて通年での推測

○個票データでの補完 ×時系列データとしての補完

⇒両調査の調査モードや標本の違いを反映し精度の高い推定

例：月次での支出の変動(2014年)



* 消費や食費で家計調査が低いのは過小記入バイアスや儉約化か？
(過小記入バイアス Deaton&Irish, 1984; 牧, 2007; 儉約化・調査疲れ Stephens&Unayama, 2011)

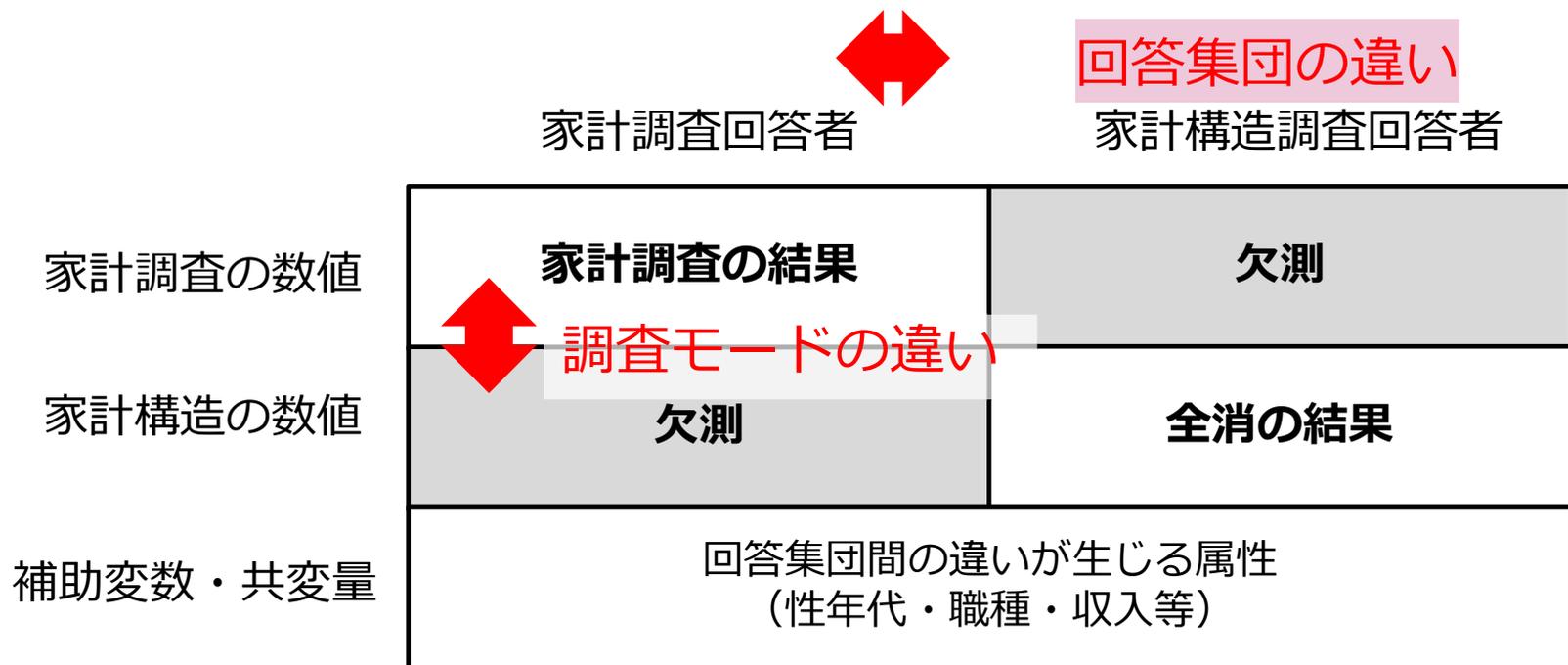
前提となる標本の特徴

家計調査と家計構造(旧全消)の違い

“調査・データ取得モードの違い” = 取り方の違い(今回は質問)

⇒両者の違いが混ざっているので分離して議論したい

* 但し今回の両者の違いは非標本誤差ではなく標本誤差



目的

本研究の目的：全国家計構造調査の年平均の推定

このために家計調査を補助情報として利用する

【家計調査と全国家計構造調査（以前の前消）の乖離の原因】

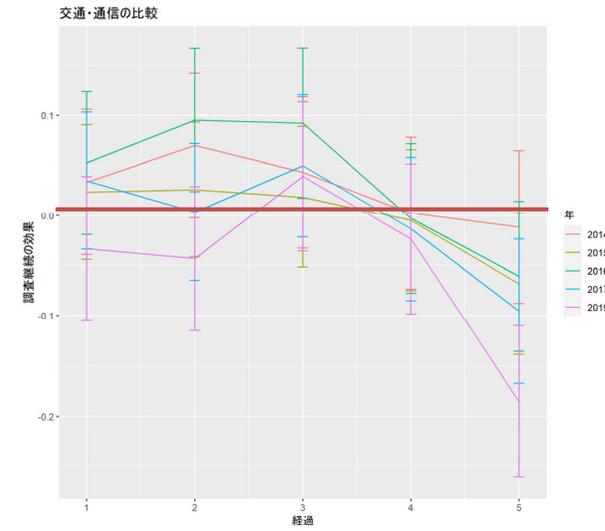
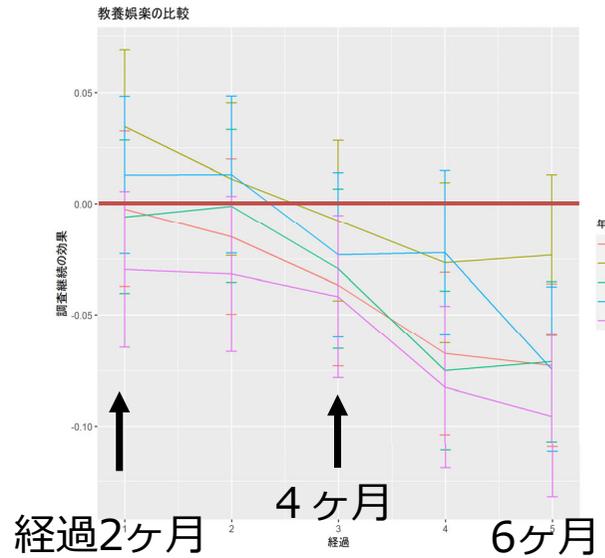
同じ月の比較で全体として全消の方が大

- ①標本の違い
- ②季節性
- ③調査継続による効果（節約効果＋記入漏れの増大）
- ④調査項目対象(財・サービスの対象)

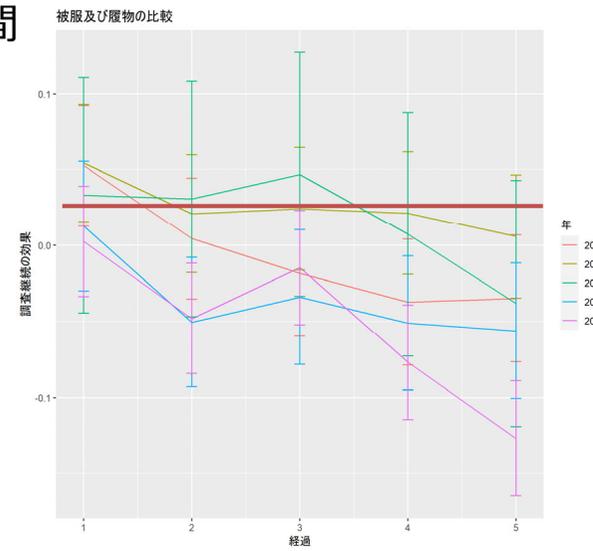
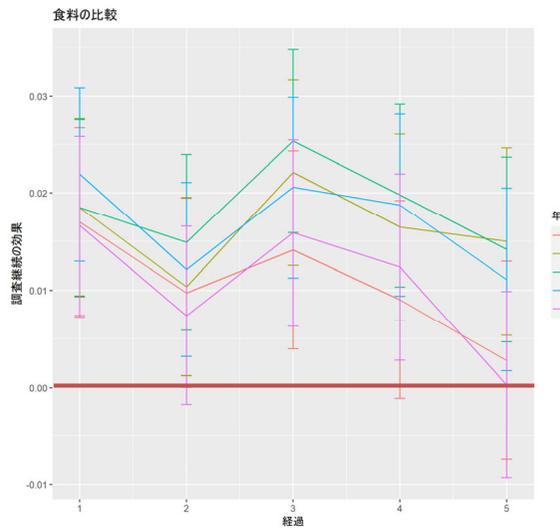
今回は②と③を調整

* 厳密には回帰によって①も一定程度考慮する

各支出費目での調査継続効果



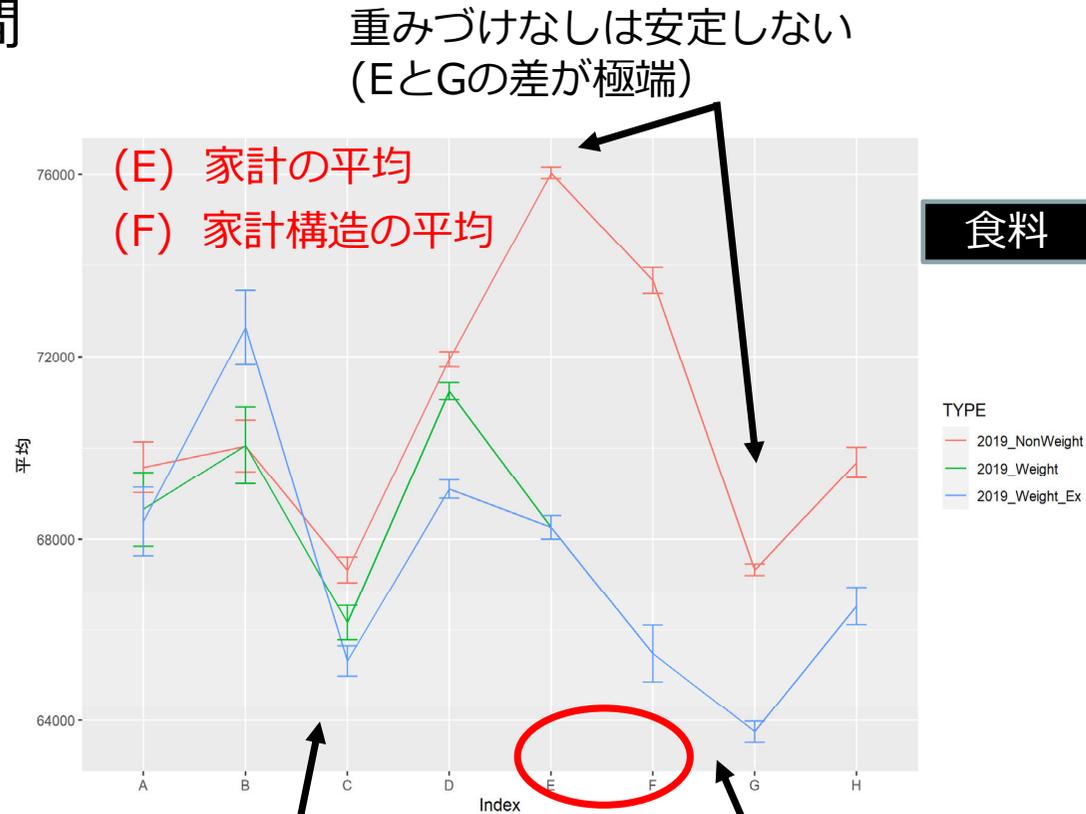
赤線はゼロ
バーは
95%信頼区間



年平均の推定・標準誤差の例

バーは95%信頼区間

- (A)年推定 (調査継続効果平均)
- (B)年推定 (調査継続効果最小)
- (C)調査継続効果1,2か月目での年推定(回帰利用)
- (D)調査継続効果1,2か月目での年推定(平均利用)
- (E)年平均 (家計)
- (F)10,11 月平均 (家計)
- (G)10,11 月平均 (家計構造)
- (H)年平均
(家計 + 構造の効果 = $E + (G - F)$)



A~Dの中では安定

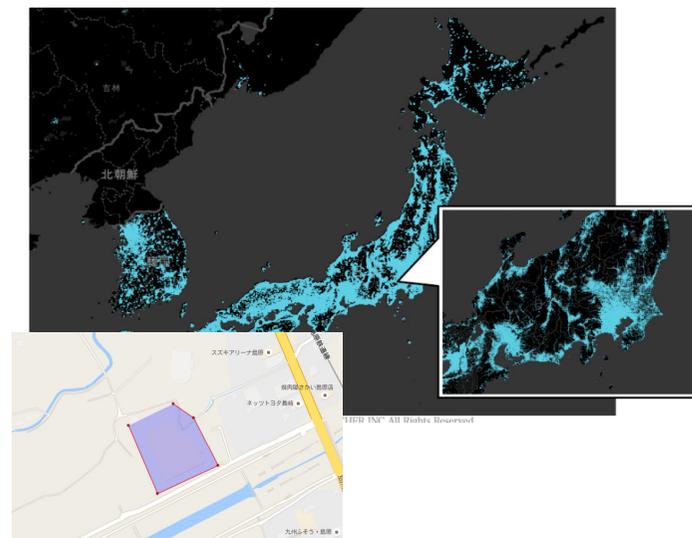
重みづけがあると安定 (E~Iの差がなだらか)

こちらの分野の研究についての今後

(1) 利用できるデータの拡大

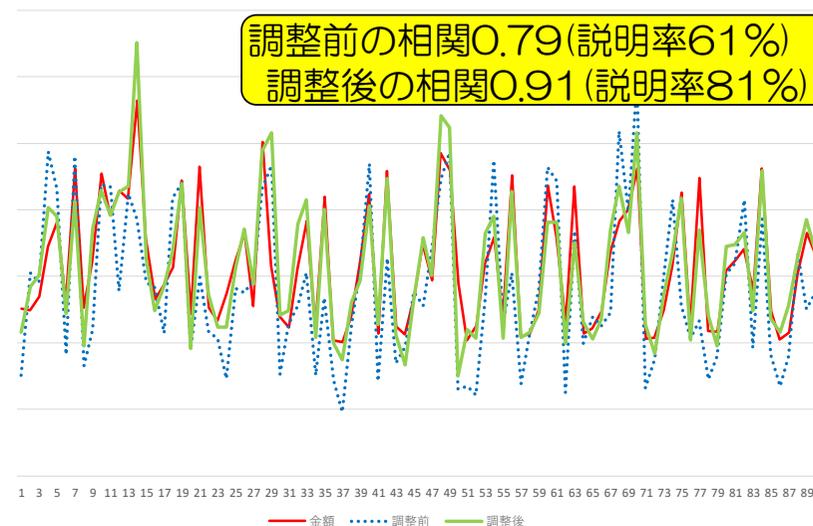
例) 位置情報

- * 売上情報と位置情報データから他社予測(Kato&Hoshino,2020)
- * 失業者予測(Moriwaki et,al,)



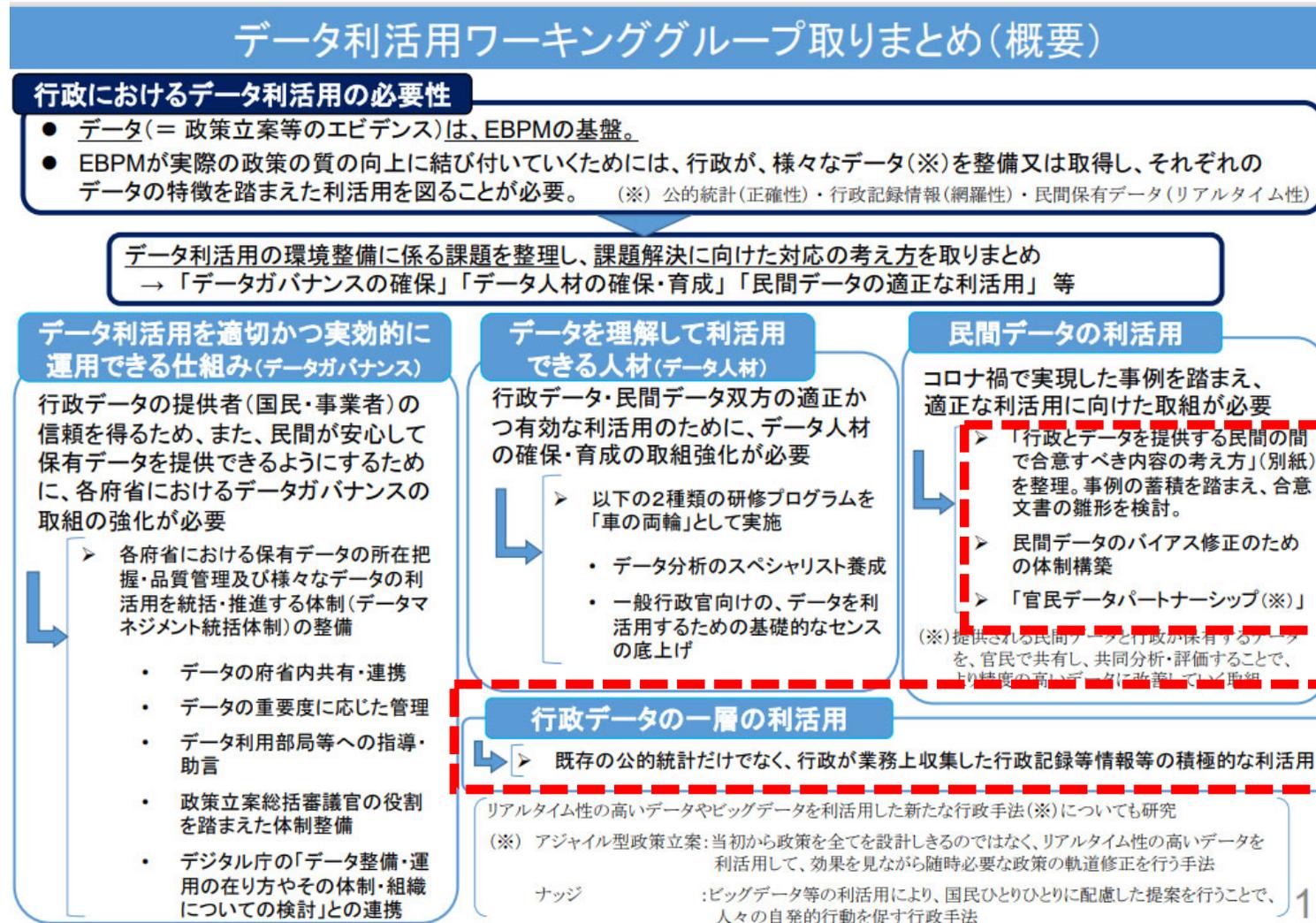
(2) 企業からどのようなデータを 取得するので十分な精度が 担保されるか？

例えば政府統計による母集団
情報に従った集計時系列など



EBPM委員会ビッグデータ活用WGのとりまとめ

出典:内閣官房HP <https://www.kantei.go.jp/jp/singi/it2/ebpm/dai7/siryoushu1-2.pdf>



まとめ

- 民間は膨大なビッグデータを活用してビジネス上の意思決定を実施:ただしビッグデータにバイアスがあることも理解されつつある
- EBPMの中で速報性のある「現状把握」にはビッグデータ活用可能

- その場合でも“測定誤差”と“代表性”に注意しないと誤った解釈

- 完全な1つのデータで議論することは困難・複数のデータを統合的に利用する必要
- 複数データの統合の目的・枠組み・方法を
“Biased sampling” “欠測データ” “複数モーメント情報の融合”
の観点から整理

- まずは政府統計内のデータ融合の成果を公表として実現させた