

全国家計構造調査(旧全消)
「年平均推定値」の推定方法について
—10大費目への適用にむけて
ゼロが多い場合—

2021年11月25日

慶應義塾大学大学院経済学研究科

／理化学研究所AIPセンター

松永将志・二荒麟

慶應義塾大学 経済学部／理化学研究所AIPセンター

星野崇宏

前回の「まとめと今後の課題」

【まとめ】全国家計構造調査の年次集計に家計調査を利用する場合、過去の家計調査と全消の乖離の原因を理解する必要

“同じ月の比較で全体として全消の方が大”

①標本の違い 収入 家計 < 全消 $\Rightarrow$ 全消の消費大

②季節性 全消の対象月は↓

③調査継続による効果 家計調査は6か月継続のため↓

本研究では①～③を考慮する方法を提案し合理的な値を推計

前回との違いは「共変量選択」「標準誤差」「検定」の実施

【課題】

(1)まだ一部費目(光熱費)のみ説明不能 $\Rightarrow$ 2019年調査への適用？

(2)各費目の地域ごとの集計値の算出

(3)最終的に何を除去するのか？何を目標にするのか？について

の議論が必要 $\Rightarrow$ 何を考慮した“年間支出”か？

目次

1. 背景・目的
 2. 調査継続効果の多年度比較
 3. 家計構造調査を利用した推定の比較
 4. 地域別の支出額・総支出（2014・2019）
 5. 資料
-

1.背景・目的

- 前提となる2標本の特徴
- 本研究の目的

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

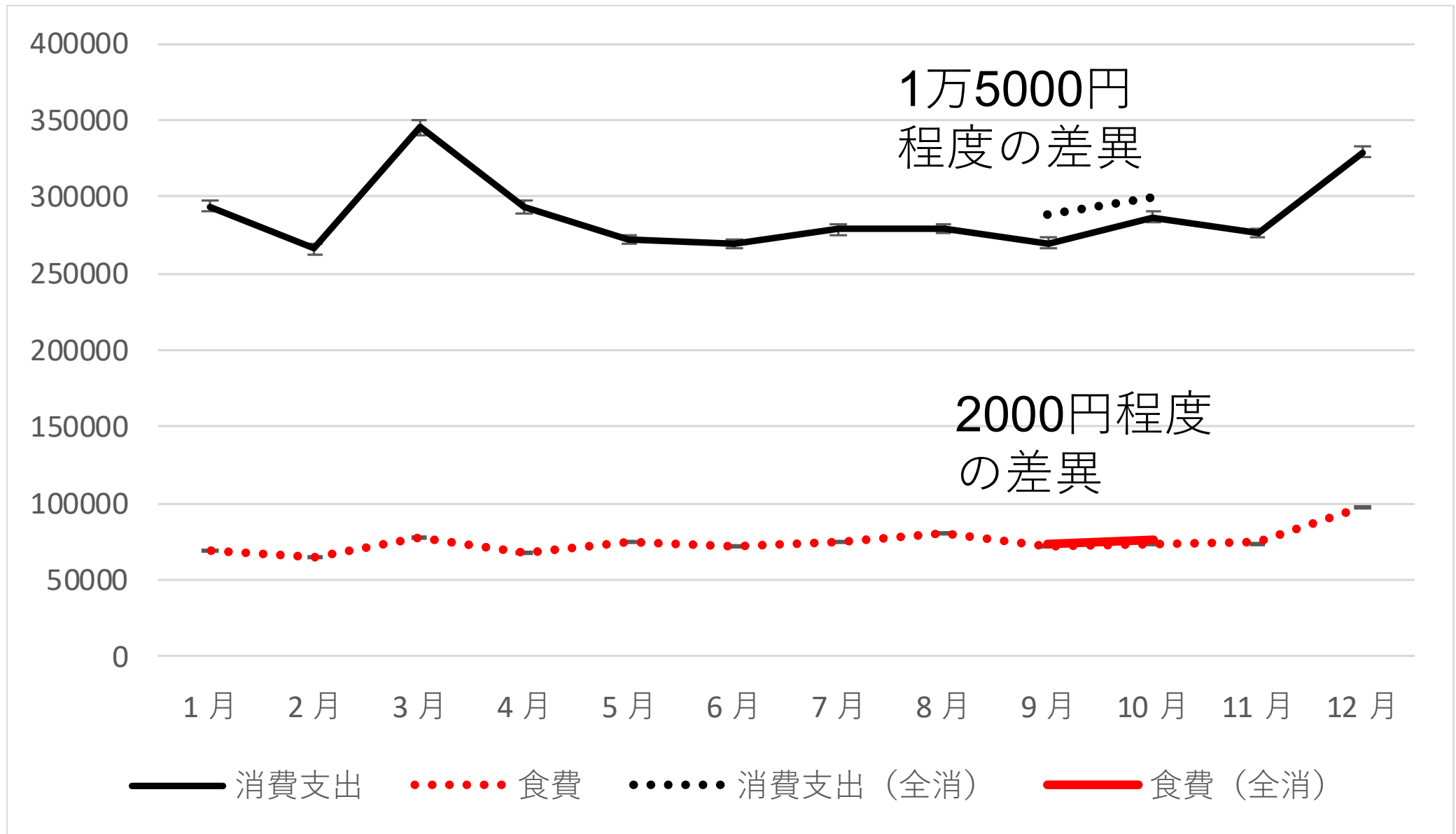
- 通年で結果が得られるローテーションパネル
- 比較的サンプルサイズが小さい(約9,000世帯)

◆ 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)

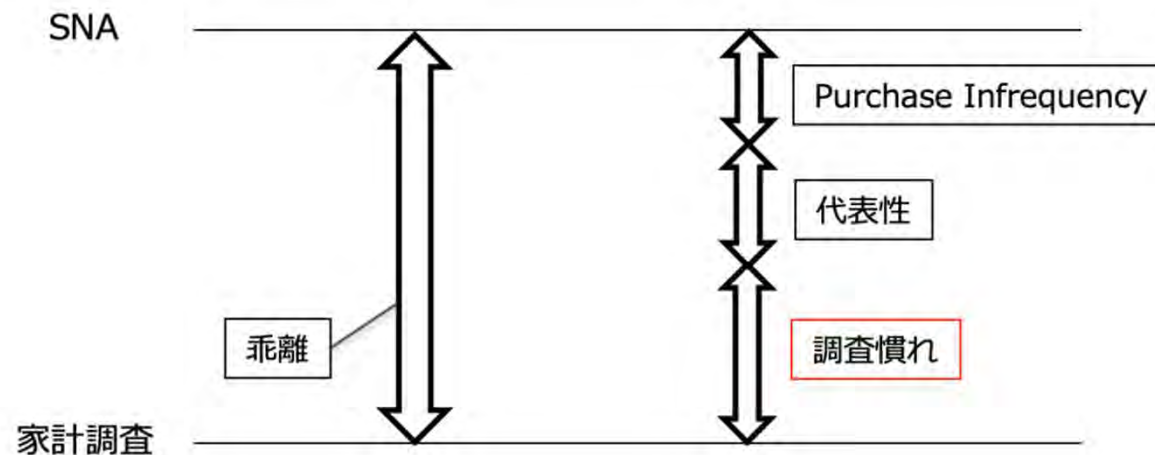


* 消費や食費で家計調査が低いのは過小記入バイアスや倹約化か？
 (過小記入バイアス Deaton&Irish, 1984; 牧, 2007; 倹約化・調査疲れ
 Stephens&Unayama, 2011)

目的

乖離の補正

- 今回は家計調査と全国家計構造調査（旧全消）の乖離に着目する中で調査継続による効果を考慮
- ミクロ統計とマクロ統計の差の乖離を埋め合わせる議論にもつながる



目的

本研究の目的：全国家計構造調査（旧全消）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（以前の全消）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
- 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

属性情報による回帰とTobit-type2 model

家計*i*の当年*t*月のデータ（またはそれ+1の対数）を家計に関する属性情報(子供の数や世帯収入等)*x*、季節ダミー*D*、調査継続ダミー*E*を用いて下記で表現

$$y_{it} = \gamma^T x_i + \sum_{m=1}^{12} \alpha_m D_{it} + \sum_{k=2}^6 \beta_k E_{itk} + \varepsilon_{it}$$

$$\text{但し } D_{it} = \begin{cases} 1 & t = m \\ 0 & \text{otherwise} \end{cases}, \quad E_{it} = \begin{cases} 1 & \text{家計 } i \text{ が } t \text{ 月に } k \text{ か月目} \\ 0 & \text{otherwise} \end{cases}$$

⇒回帰による回答者の違いの考慮

加えて

$$y_{it}^* = \lambda^T x_i + \epsilon_{it}$$

を背後に考え

教育費などゼロが多い項目のことも考慮して*y*ではなく*v*が観測されるとする。

$$v_{it} = \begin{cases} y_{it} & \text{if } y_{it}^* \geq 0 \\ 0 & \text{if } y_{it}^* < 0 \end{cases}$$

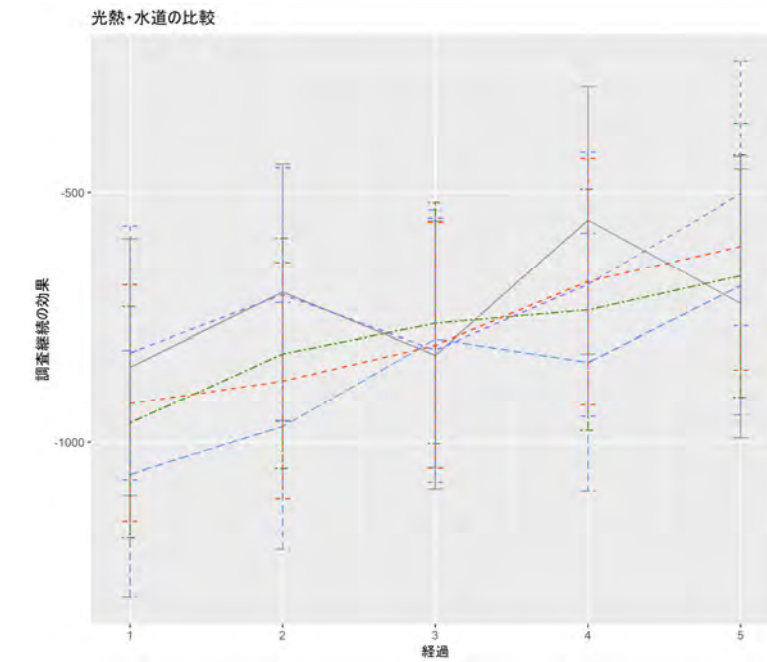
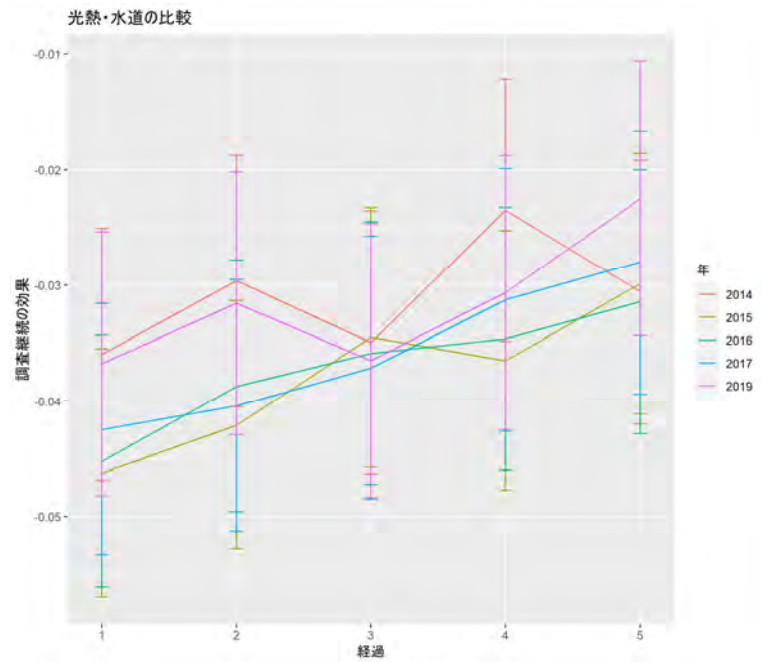
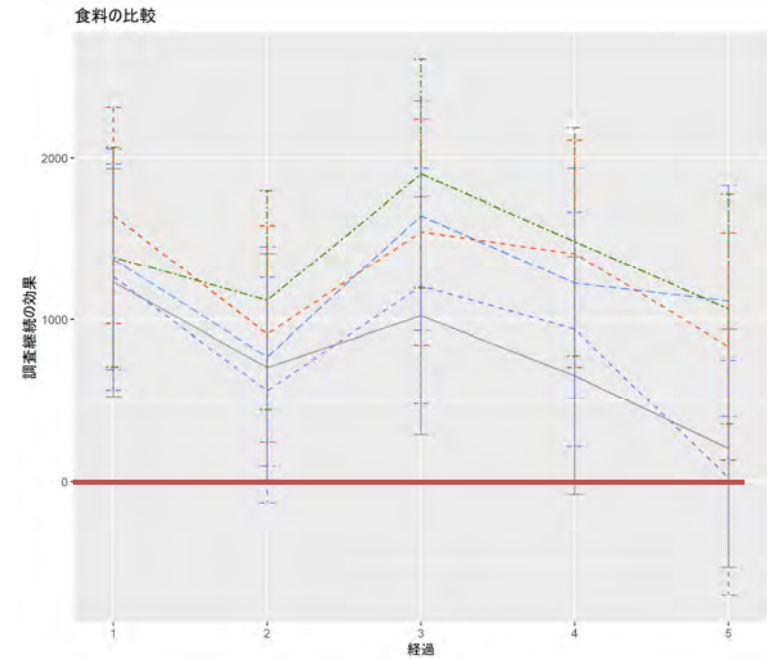
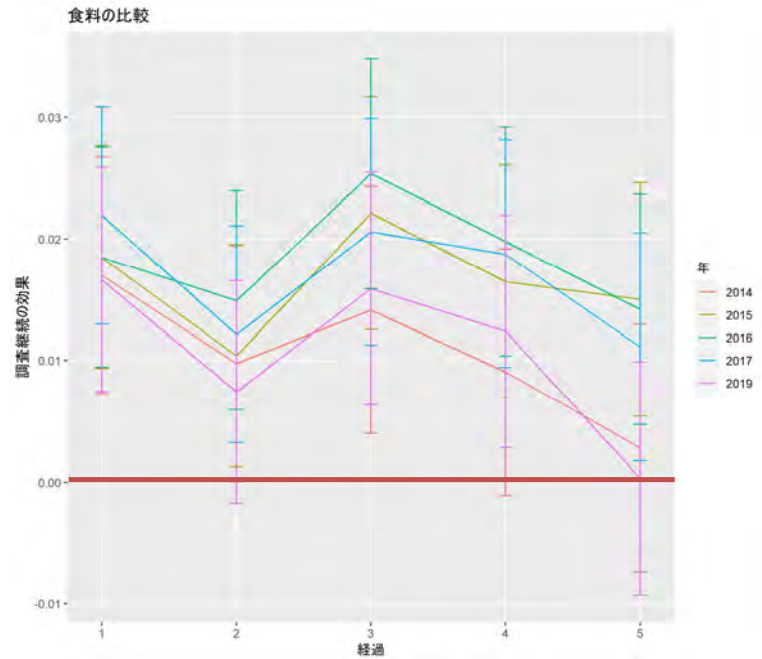
2. 調査継続効果（費目別）

- 調査継続による効果の結果（2014~2019年度）
 - * 但しデータ整備の都合上2018年除く
- 調査継続効果の検定

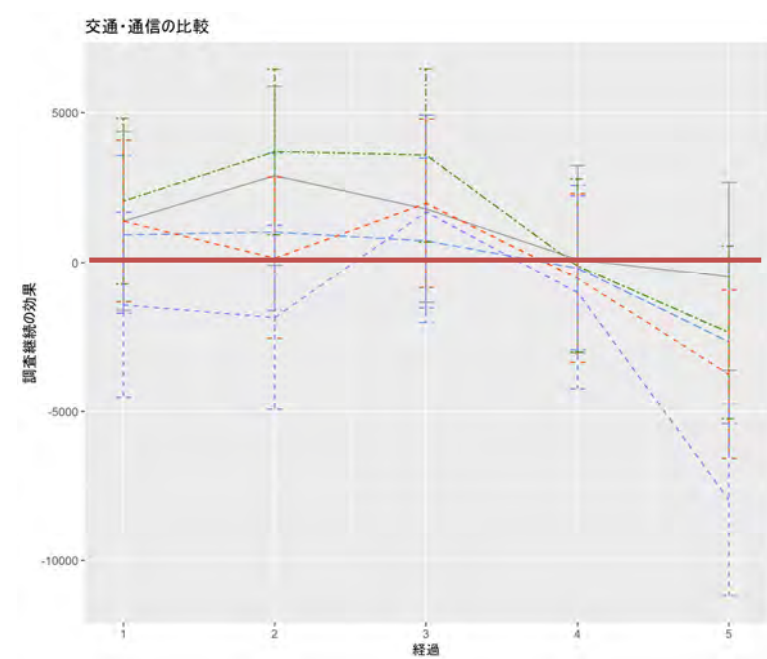
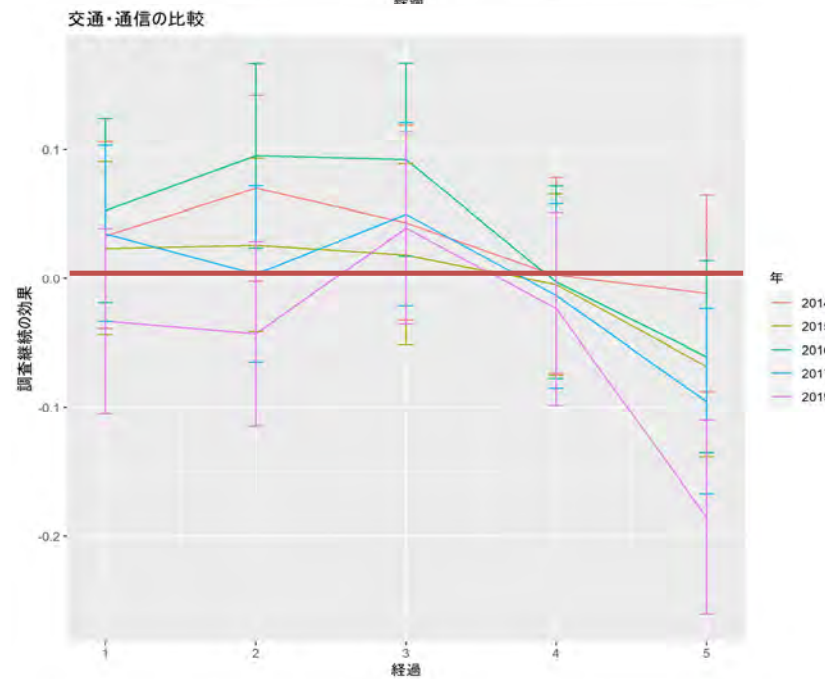
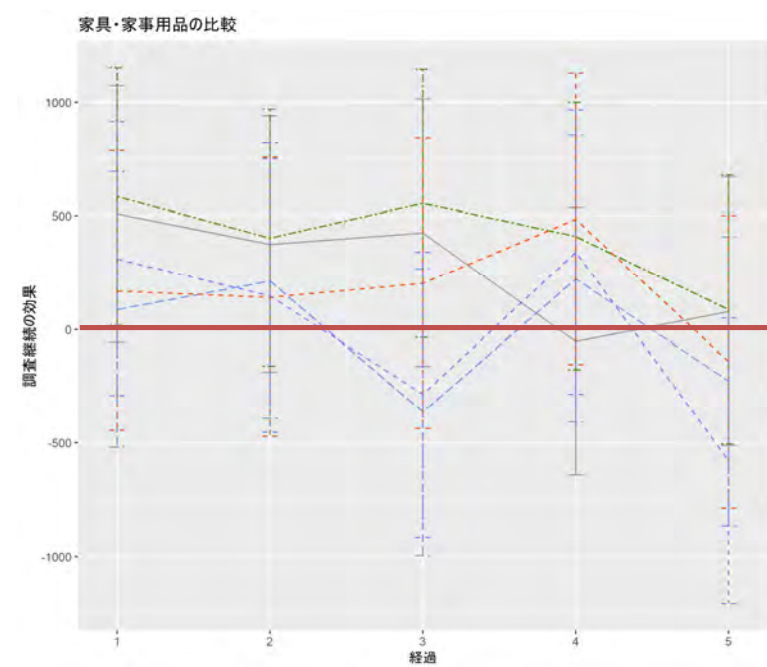
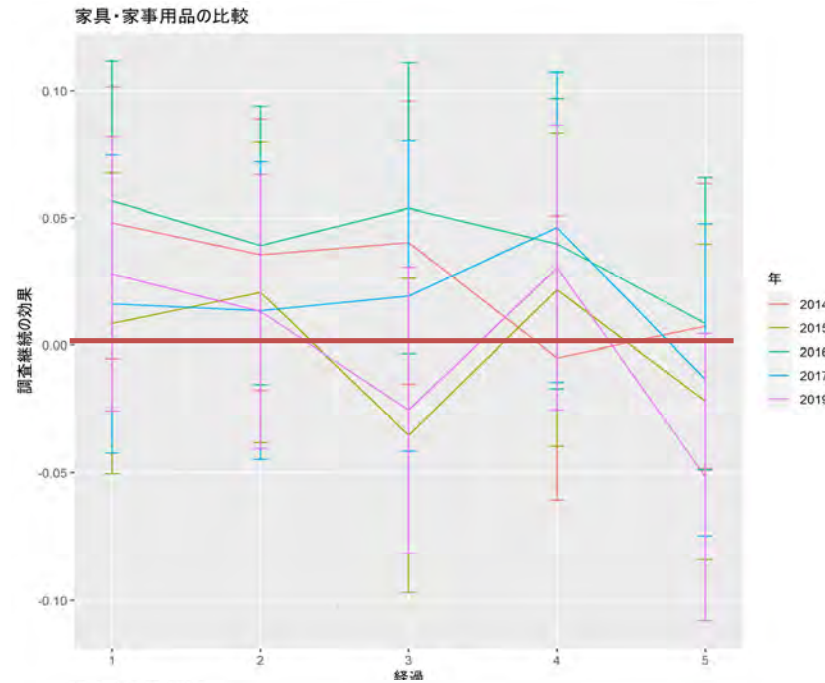
調査継続による効果

- 季節効果とは別に、調査継続に伴った影響も6か月まで今回の分析ではみることができる（家計調査のみを使用）
 - ◆ 例) 2～3か月目については増加、その後減少など
 - 季節ダミーも別途あるため継続効果には影響なし
 - 多年度での比較の結果、費目ごとに一定の再現性あり
 - ◆ 調査開始から1か月程度で実施される記入方法の指導等が原因か
 - さらに費目ごとに検定を実施（2014年,2019年）
 - ◆ Tobit-Type2はWald検定、重回帰はF検定
-

調査継続効果 (食料、光熱・水道)

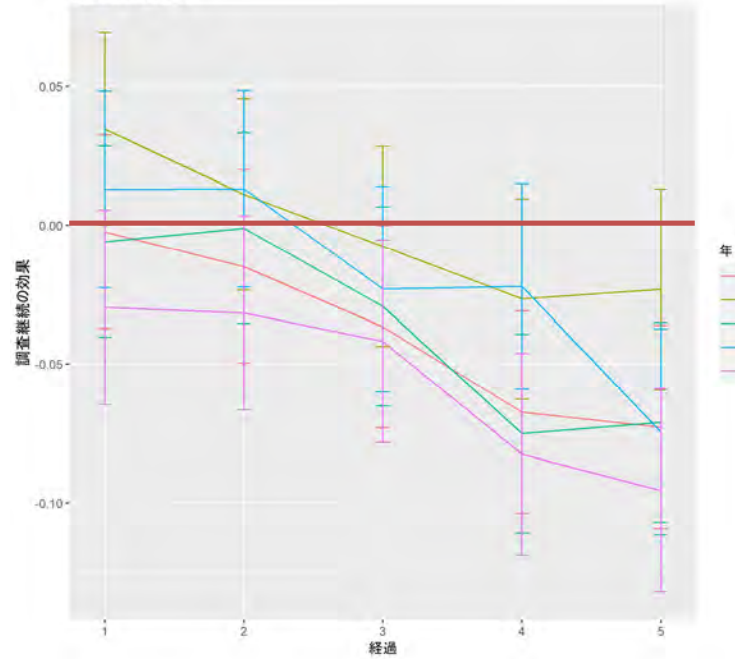


調査継続効果（家具・家事用品、交通・通信）

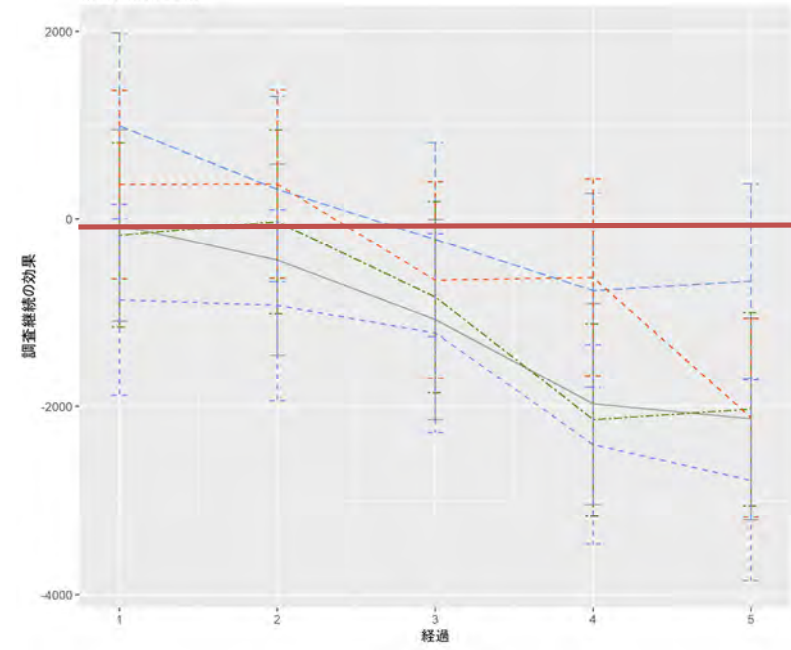


調査継続効果（教養娯楽、その他の消費支出）

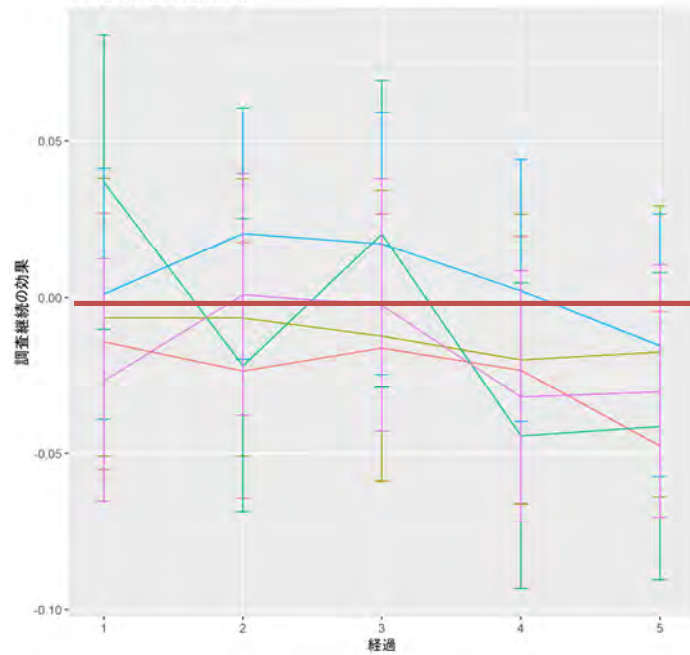
教養娯楽の比較



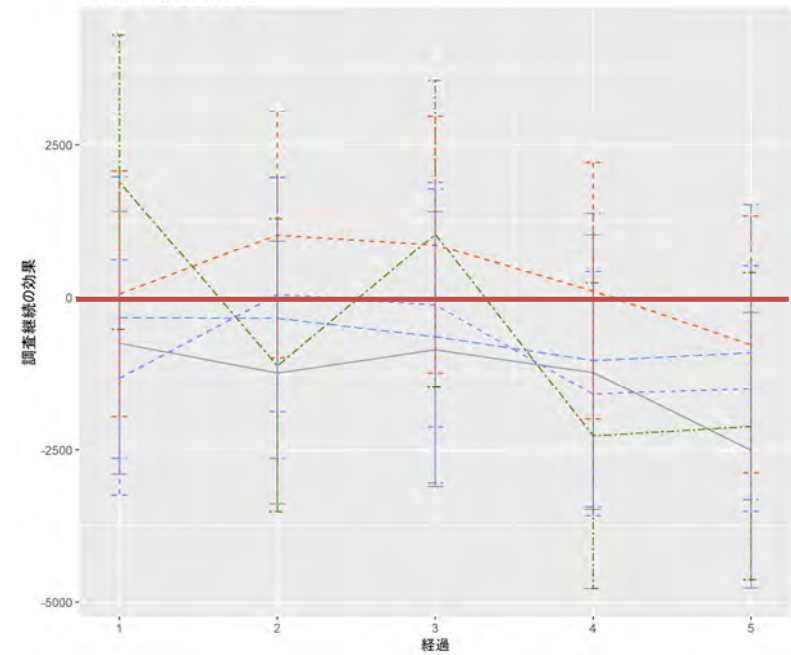
教養娯楽の比較



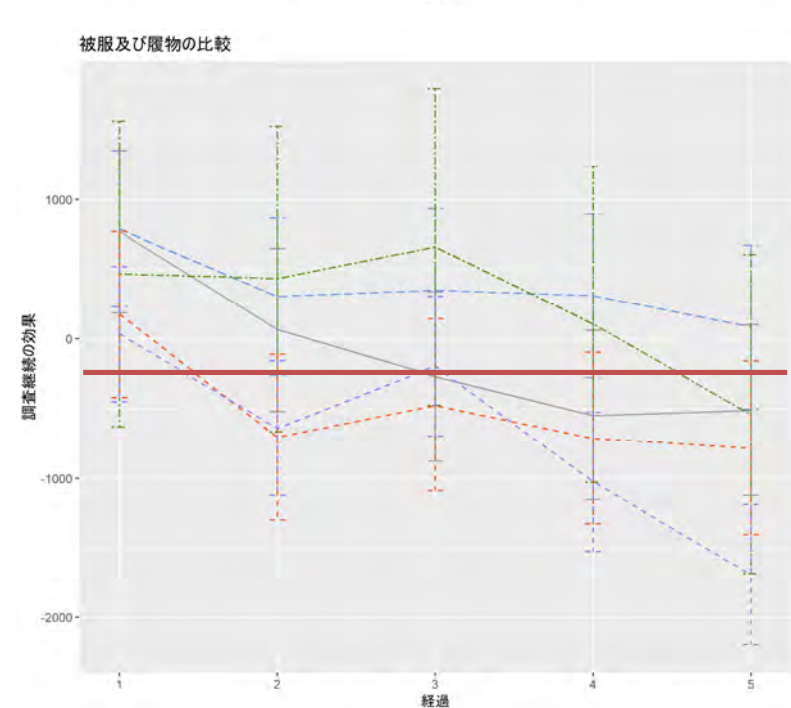
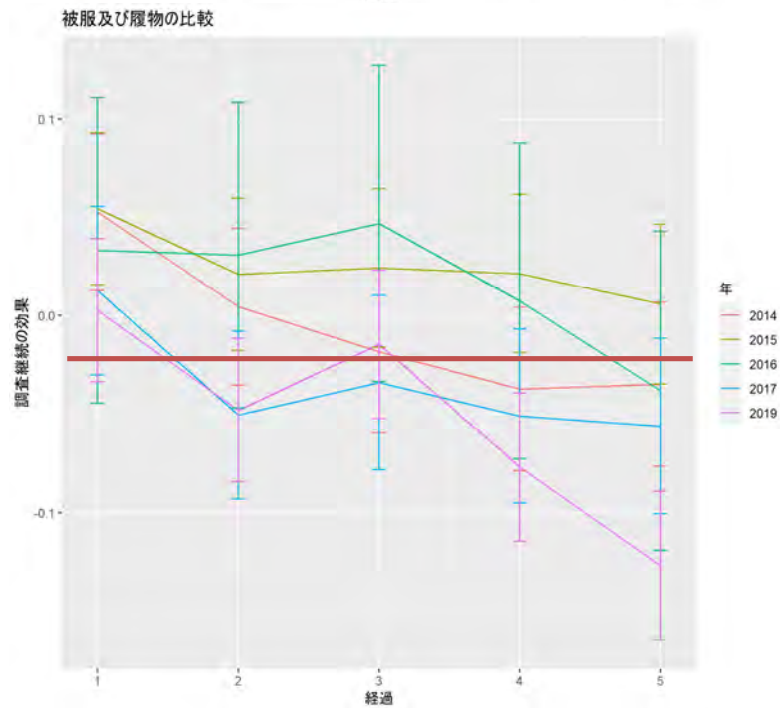
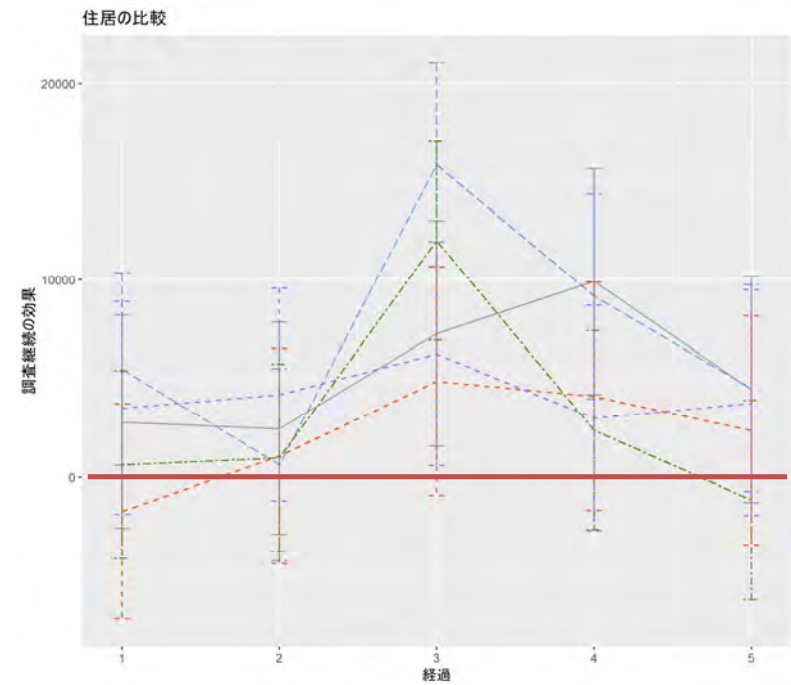
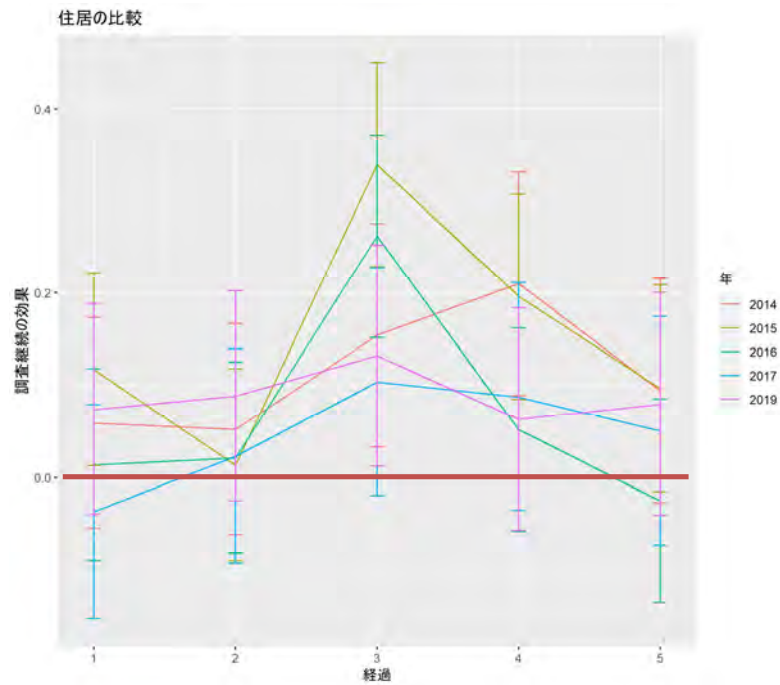
その他の消費支出の比較



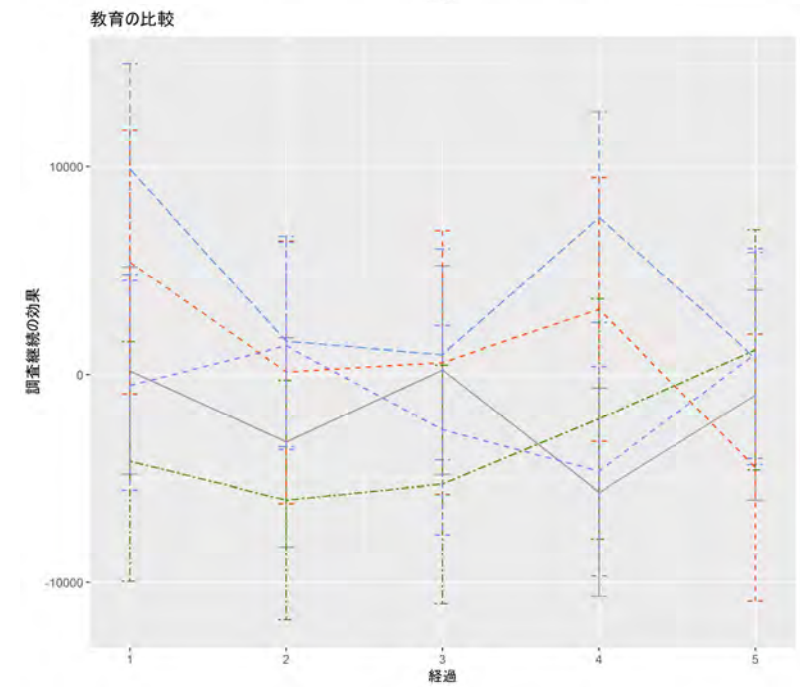
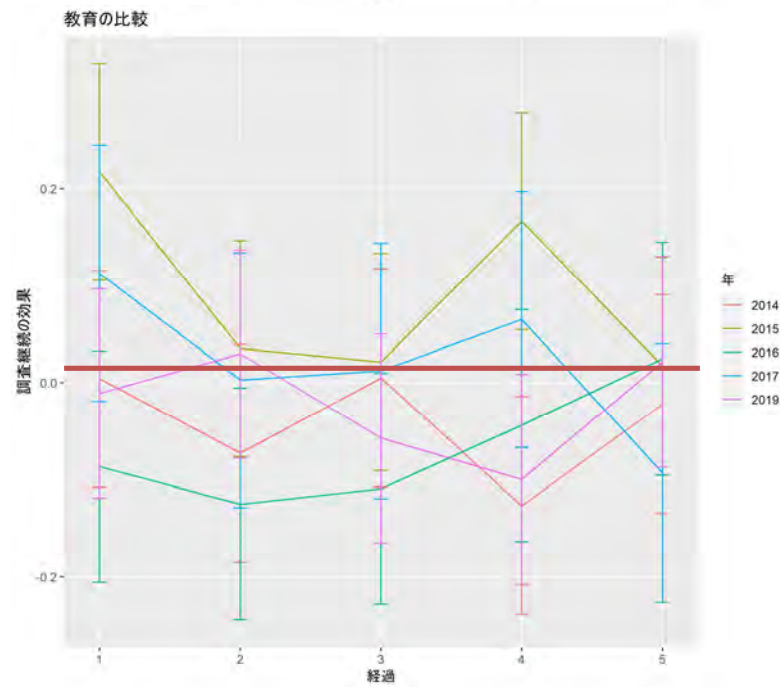
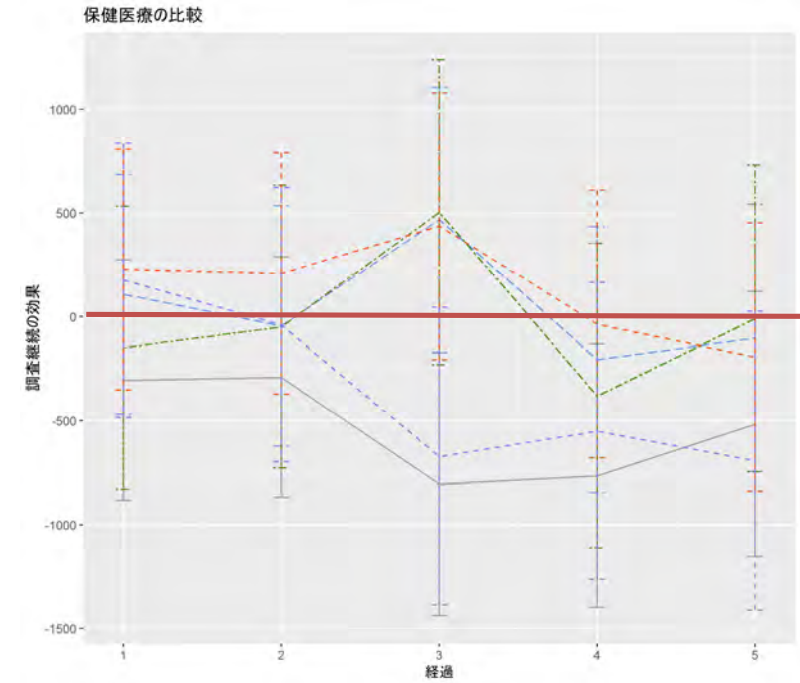
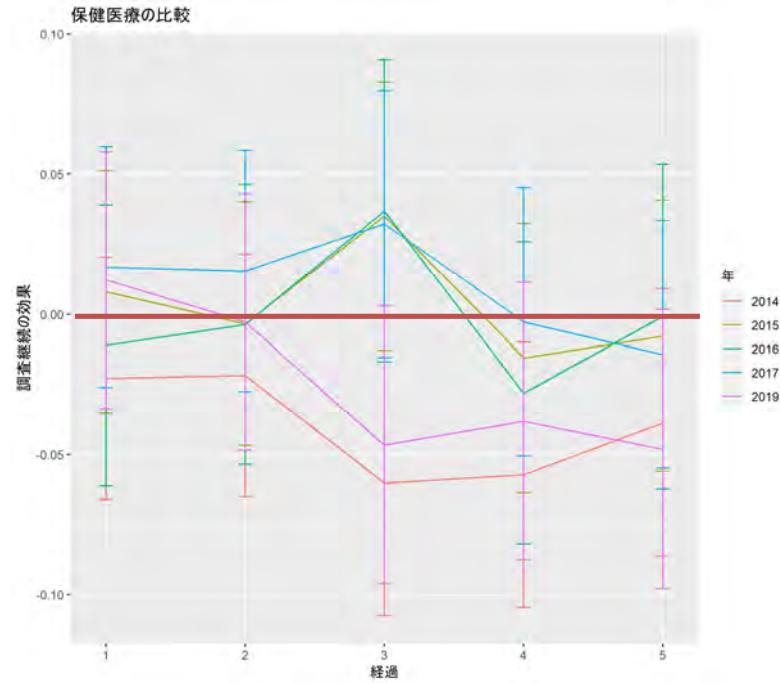
その他の消費支出の比較



調査継続効果（住居、被服及び履物）



調査継続効果（保健医療・教育）



調査継続による効果に関する検定 (2014)

10大費目	推定方法	検定方法	F値或いはWald統計量	P値
食料	重回帰分析	F検定	1295.00	0
住居	Tobit-Type2	Wald検定	42.04	5.79E-08
光熱、水道	重回帰分析	F検定	958.30	0
家具、家事用品	重回帰分析	F検定	105.22	0
被服及び履物	Tobit-Type2	Wald検定	23.84	0.00023
保険医療	Tobit-Type2	Wald検定	17.63	0.00345
交通、通信	重回帰分析	F検定	68.11	0
教育	Tobit-Type2	Wald検定	10.45	0.06339
教養娯楽	重回帰分析	F検定	270.70	0
その他の消費支出	重回帰分析	F検定	292.60	0

推定された調査継続効果パラメーターと調査継続効果パラメーター（5つのダミー変数）を「0」とした場合が有意に異なるかどうかの検定：

- $F値 = ((SSE(R) - SSE(F)) / K) / (SSE(F) / (n - K))$

* SSE: 残差平方和, K: 説明変数の数, n: サンプルサイズ

R: 「調査継続効果パラメーター=0」のモデル

F: 推定された調査継続効果パラメーターによるモデル

- $Wald統計量 = (\theta(F) - \theta(R))' V(\theta(F))^{-1} (\theta(F) - \theta(R))$

* $\theta(R)$, $\theta(F)$: それぞれのモデルにおける推定値

調査継続による効果に関する検定（2019）

10大費目	推定方法	検定方法	F値或いはWald統計量	P値
食料	重回帰分析	F検定	1546.00	0
住居	Tobit-Type2	Wald検定	4.28	5.10E-01
光熱、水道	重回帰分析	F検定	896.90	0
家具、家事用品	重回帰分析	F検定	87.02	0
被服及び履物	Tobit-Type2	Wald検定	45.51	0.00000
保険医療	Tobit-Type2	Wald検定	10.61	0.05978
交通、通信	重回帰分析	F検定	66.89	0
教育	Tobit-Type2	Wald検定	3.09	0.68578
教養娯楽	重回帰分析	F検定	294.10	0
その他の消費支出	重回帰分析	F検定	247.40	0

推定された調査継続効果パラメーターと調査継続効果パラメーター（5つのダミー変数）を「0」とした場合が有意に異なるかどうかの検定：

– $F値 = ((SSE(R) - SSE(F)) / K) / (SSE(F) / (n - K))$

* SSE: 残差平方和, K: 説明変数の数, n: サンプルサイズ

R: 「調査継続効果パラメーター=0」のモデル

F: 推定された調査継続効果パラメーターによるモデル

– $Wald統計量 = (\theta(F) - \theta(R))' V(\theta(F))^{-1} (\theta(F) - \theta(R))$

* $\theta(R)$, $\theta(F)$: それぞれのモデルにおける推定値

2調査の差と調査継続による効果の大きさ(2014)

19

10大費目	推定方法	9.10月での平均の 2調査の乖離(絶 対値)	共変量の群間 差による説明 分	調査継続効果 の最小値最大 値の差	家計調査 の年平均	0でない割 合(%)	継続効果バ イアス(%)
食料	重回帰分析	3039	2009.34	970.32	72471	100.000	1.339
住居	Tobit-Type2	1002	169.46	2889	17947	37.526	6.041
光熱、水道	重回帰分析	1173	568.01	307.53	22808	97.708	1.317
家具、家事用品	重回帰分析	1196	315.09	524.599	10307	97.448	4.960
被服及び履物	Tobit-Type2	368	25.83	938	11874	81.860	6.467
保険医療	Tobit-Type2	538	33.50	798	12118	90.561	5.964
交通、通信	重回帰分析	5831	1204.99	3241.94	40013	98.158	7.953
教育	Tobit-Type2	2294	178.37	1096	9793	22.915	2.565
教養娯楽	重回帰分析	1803	1075.00	1945.45	28791	98.449	6.652
その他の消費支 出	重回帰分析	5483	1976.38	1613.14	51618	98.002	3.063

食費などは乖離が大きいがこれは「2調査の収入と世帯人員の差」で説明できる
一方それ以外では調査継続効果による説明の可能性

注1) 共変量の群間差による説明分 = ゼロでない割合/100

$\times \sum (\text{共変量}i\text{の係数} \times (\text{全消の共変量}i\text{の平均} - \text{家計の共変量}i\text{の平均}))$

ただし、重回帰分析の共変量*i*は年間収入,世帯人員,世帯主の性別,世帯主の年齢,
Tobit-Type2の共変量*i*は年間収入,世帯人員.

注2) 調査継続効果によるバイアス = ゼロでない割合/100

$\times (\text{調査継続効果の最大値と最小値の差} / \text{家計調査の年平均})$

2調査の差と調査継続による効果の大きさ(2019)

20

10大費目	推定方法	10,11月での平均の2 調査の乖離(絶対値)	共変量の 群間差	調査継続効果の最 小値最大値の差	家計調査の 年平均	0でない 割合	継続効果 バイアス (%)
食料	重回帰分析	8695.45	-1939.17	1243.32	76024.73	100	1.64
住居	Tobit-Type2	762.76	-206.07	1700	17290.35	38.87	3.82
光熱、水道	重回帰分析	2561.93	-854.15	316.4986091	21291.67	97.19	1.44
家具、家事 用品	重回帰分析	2033.59	-117.24	876.0465478	10954.69	96.72	7.73
被服及び 履物	Tobit-Type2	934.01	-28.68	939	10380.98	76.47	6.92
保険医療	Tobit-Type2	753.83	-184.55	750	13252.73	90.4	5.12
交通、通信	重回帰分析	3473.72	-362.44	9480.544822	41399.64	97.29	22.28
教育	Tobit-Type2	2270.50	-265.70	397	9772.728	19.42	0.79
教養娯楽	重回帰分析	3709.80	88.07	1874.234431	28490.5	97.69	6.43
その他の 消費支出	重回帰分析	4549.77	905.50	1580.038992	48606	97.49	3.17

食費などは乖離が大きいがこれは「2調査の収入と世帯人員の差」で説明できる
一方それ以外では調査継続効果による説明の可能性

注1) 共変量の群間差による説明分 = ゼロでない割合/100

$\times \sum (\text{共変量}i\text{の係数} \times (\text{家計構造の共変量}i\text{の平均} - \text{家計の共変量}i\text{の平均}))$

ただし、重回帰分析の共変量*i*は年間収入,世帯人員,世帯主の性別,世帯主の年齢,
Tobit-Type2の共変量*i*は年間収入,世帯人員.

注2) 調査継続効果によるバイアス = ゼロでない割合/100

$\times (\text{調査継続効果の最大値と最小値の差} / \text{家計調査の年平均})$

調査継続効果のまとめ

- 各費目で比較すると、水道・光熱を除き、(2か月目に上昇したうえで)数か月後に減少する傾向が2014～2019年で一定程度再現されている
 - ◆ 例外として教育費（月次で議論するのが難しい）
 - ◆ 費目ごとに調査の経過によって記入に対する態度が一定程度同じであることが推測できる
 - ◆ 水道・光熱は簡単に削減することが難しい費目であるからか
 - 調査継続効果
 - = 「初月から2か月めに掛けての記入指導(過小or過大?)」
 - + 「調査慣れ(記入が簡便化されるor一部詳細化?)」
 - + 「家計のモニタリング効果(節約が起きる)」
- と考えるとある程度理解可能か？
- 2014に比べて2019で調査継続効果の減少はウェブ入力等の効果か？
- * ウェブ入力世帯に限った分析などの必要性
-

3. 全国家計構造調査を利用した推定

- 2014年・2019年の費目ごとの推定値比較

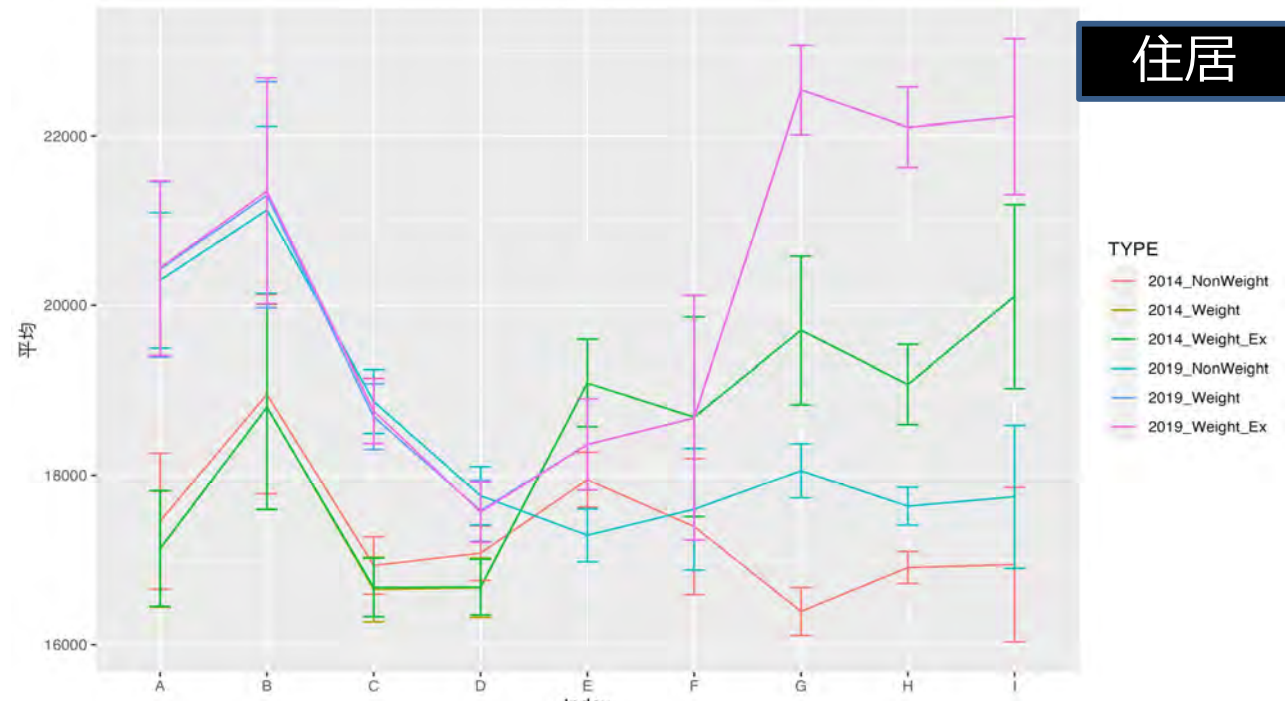
年平均の推定・標準誤差(Tobit-type2 model)

23

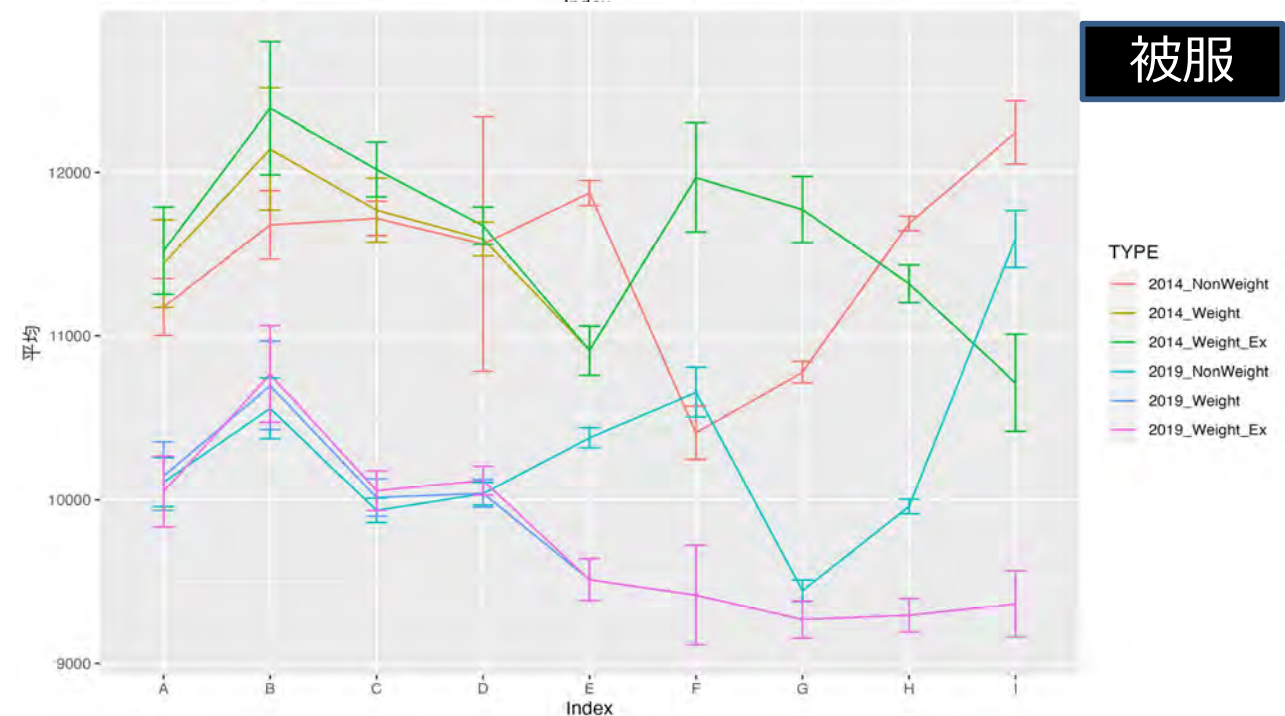
バーは95%信頼区間

- (A)年推定 (調査継続効果平均)①
- (B)年推定 (調査継続効果最小)②
- (C)調査継続効果1,2か月目での
年推定(回帰利用)③
- (D)調査継続効果1,2か月目での
年推定(平均利用)④
- (E)年平均 (家計)
- (F)9,10 月平均 (家計)
- (G)9,10 月平均 (全消)
- (H)年平均 (合計 = 家計と全消混合)
- (I)年平均
(家計 + 全消の効果 =
 $E + (G - F)$)

住居



被服

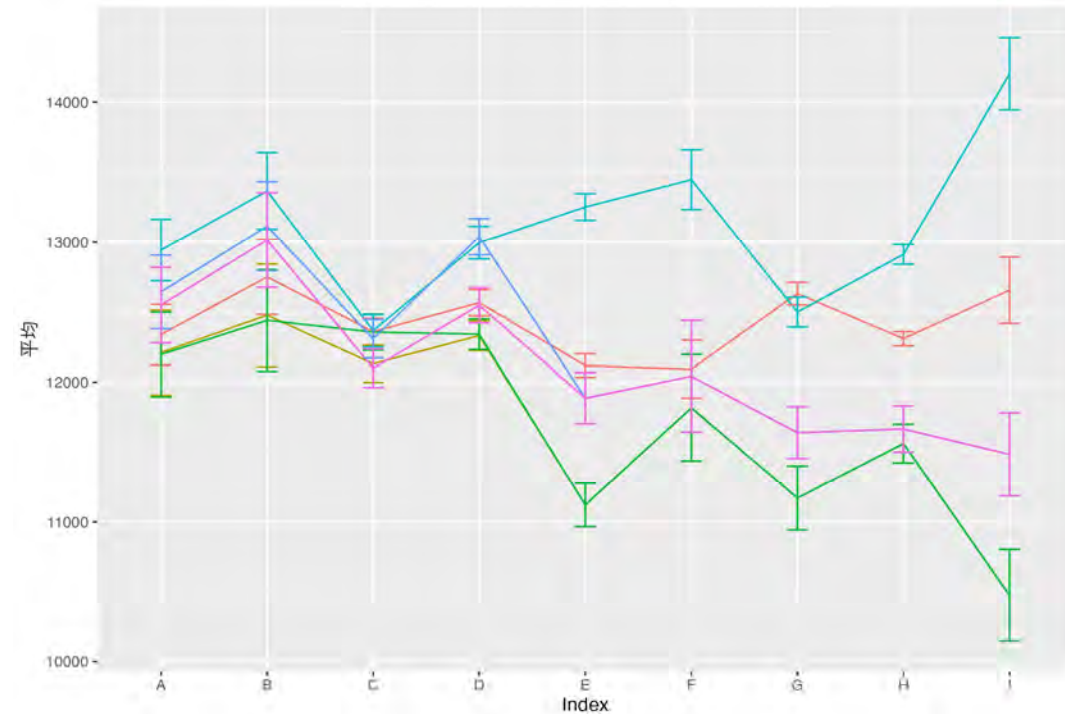


年平均の推定・標準誤差(Tobit-type2 model)

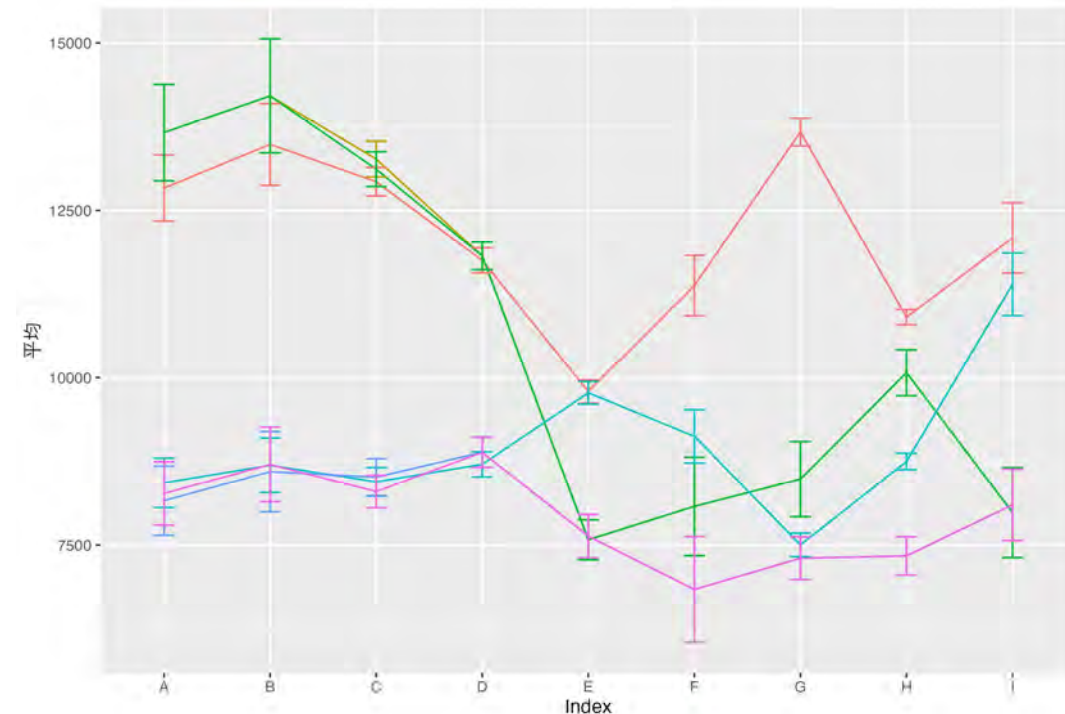
24

バーは95%信頼区間

- (A)年推定 (調査継続効果平均)①
- (B)年推定 (調査継続効果最小)②
- (C)調査継続効果1,2か月目での
年推定(回帰利用)③
- (D)調査継続効果1,2か月目での
年推定(平均利用)④
- (E)年平均 (家計)
- (F)9,10 月平均 (家計)
- (G)9,10 月平均 (全消)
- (H)年平均 (合計 = 家計と全消混合)
- (I)年平均
(家計 + 全消の効果 =
 $E + (G - F)$)



保健
医療



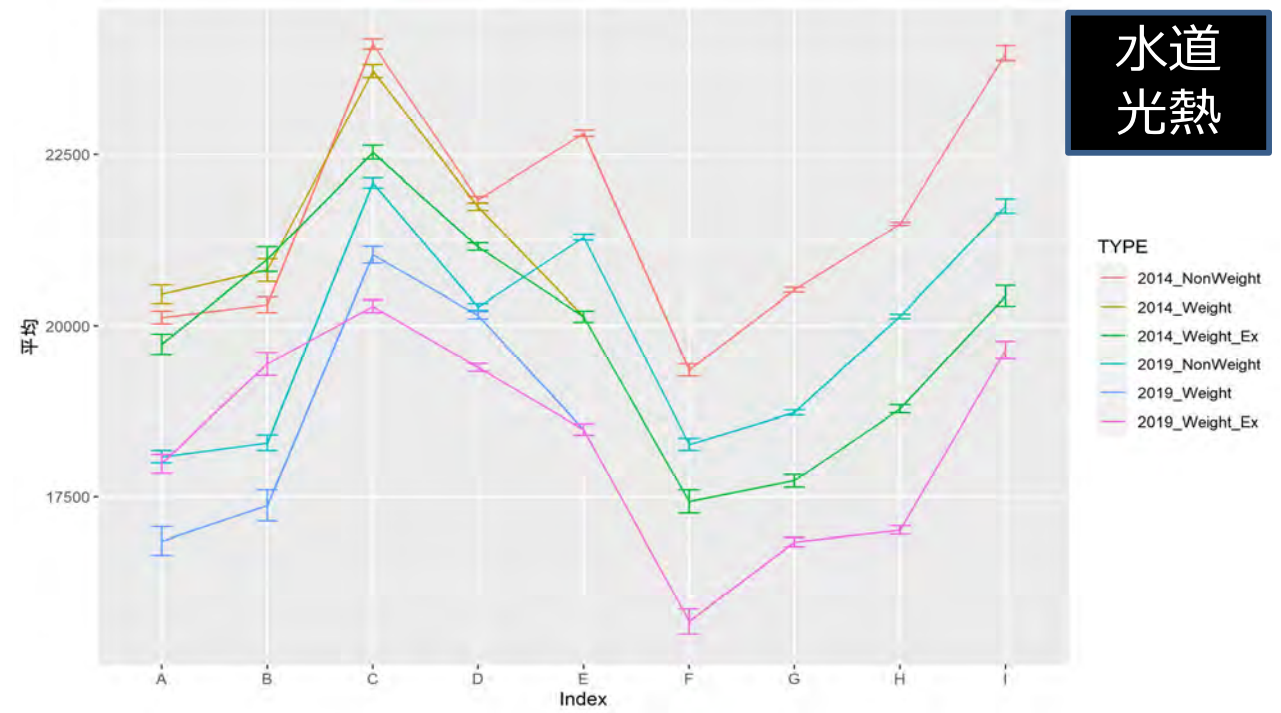
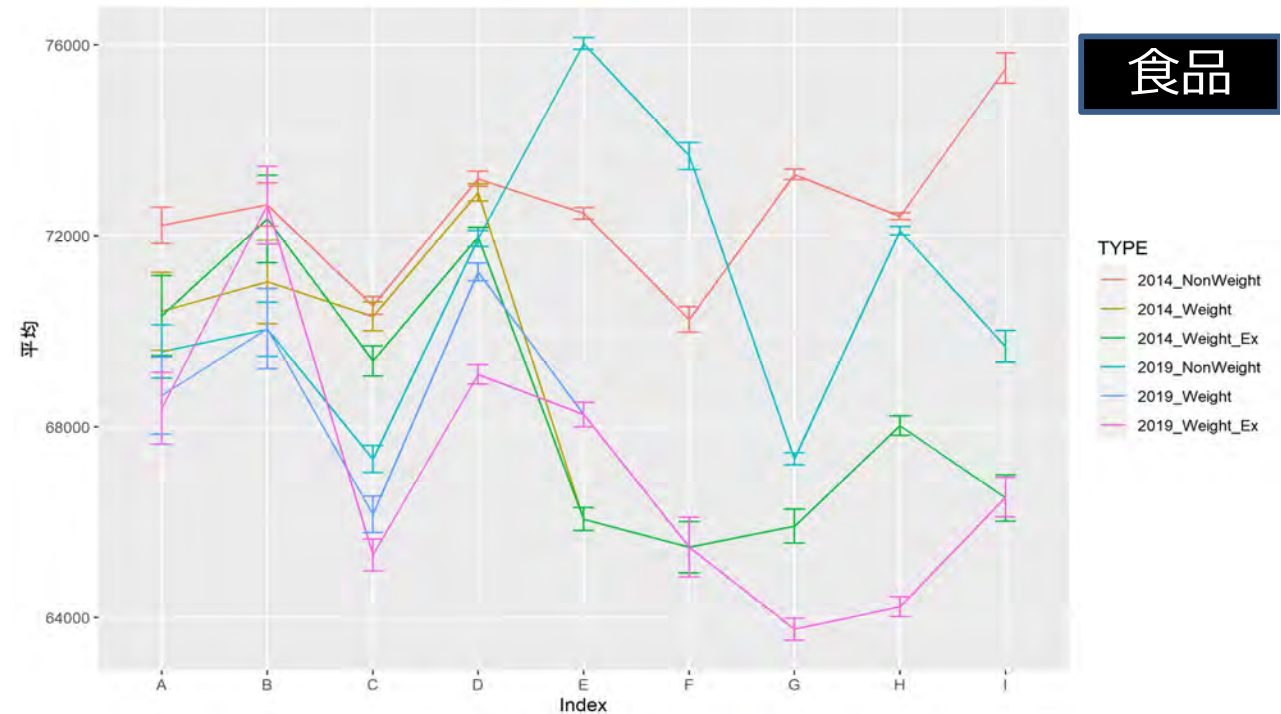
教育

年平均の推定・標準誤差(重回帰)

25

バーは95%信頼区間

- (A)年推定 (調査継続効果平均)①
- (B)年推定 (調査継続効果最小)②
- (C)調査継続効果1,2か月目での
年推定(回帰利用)③
- (D)調査継続効果1,2か月目での
年推定(平均利用)④
- (E)年平均 (家計)
- (F)9,10 月平均 (家計)
- (G)9,10 月平均 (全消)
- (H)年平均 (合計 = 家計と全消混合)
- (I)年平均
(家計 + 全消の効果 =
 $E + (G - F)$)

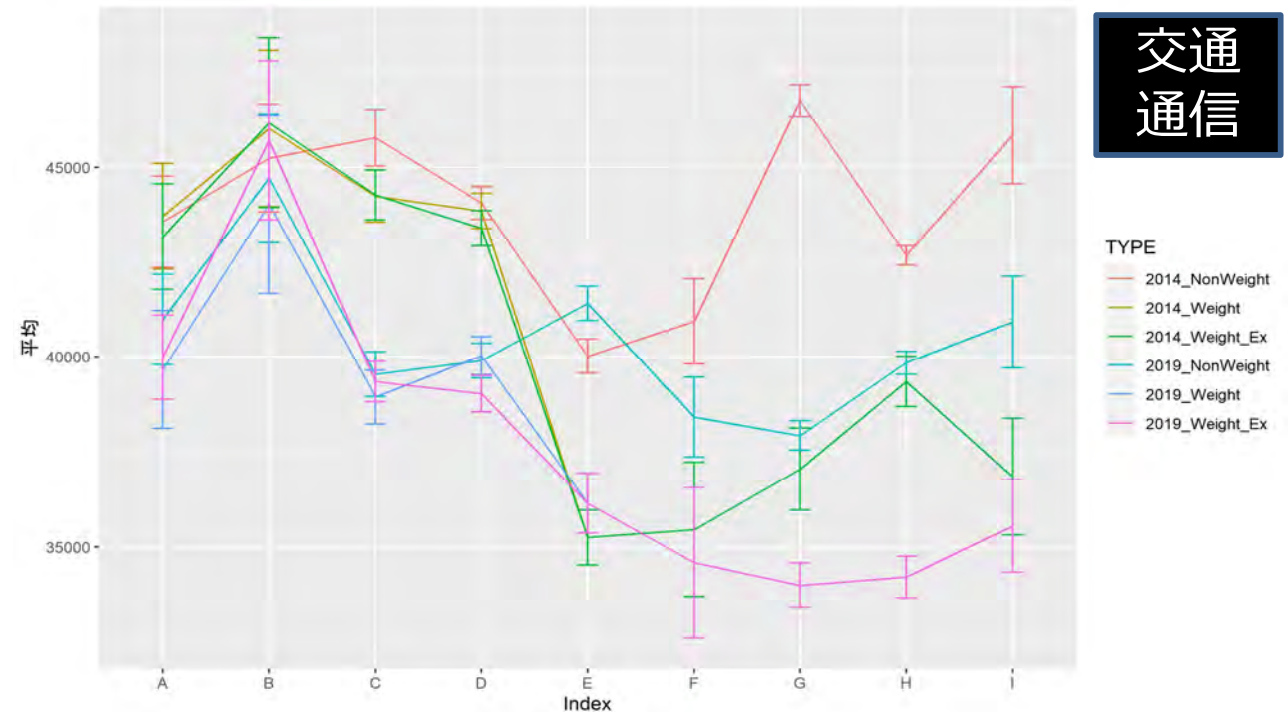
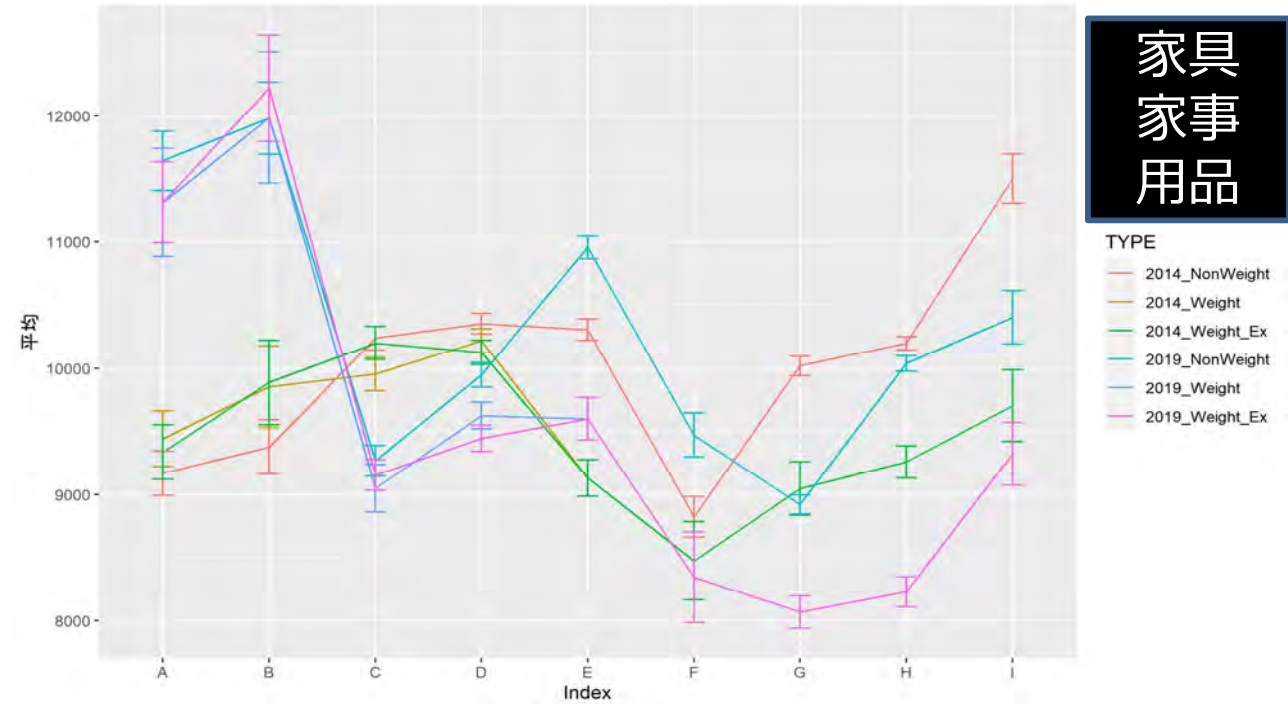


年平均の推定・標準誤差(重回帰)

26

バーは95%信頼区間

- (A)年推定 (調査継続効果平均)①
- (B)年推定 (調査継続効果最小)②
- (C)調査継続効果1,2か月目での
年推定(回帰利用)③
- (D)調査継続効果1,2か月目での
年推定(平均利用)④
- (E)年平均 (家計)
- (F)9,10 月平均 (家計)
- (G)9,10 月平均 (全消)
- (H)年平均 (合計 = 家計と全消混合)
- (I)年平均
(家計 + 全消の効果 =
 $E + (G - F)$)

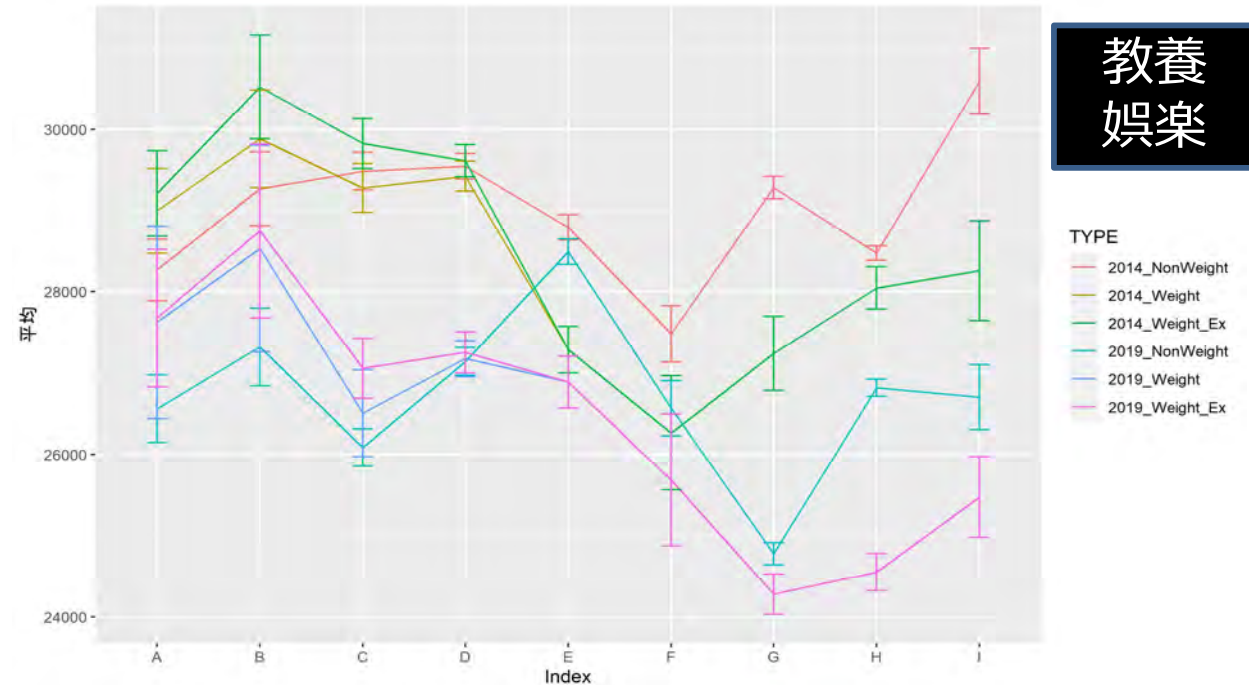


年平均の推定・標準誤差(重回帰)

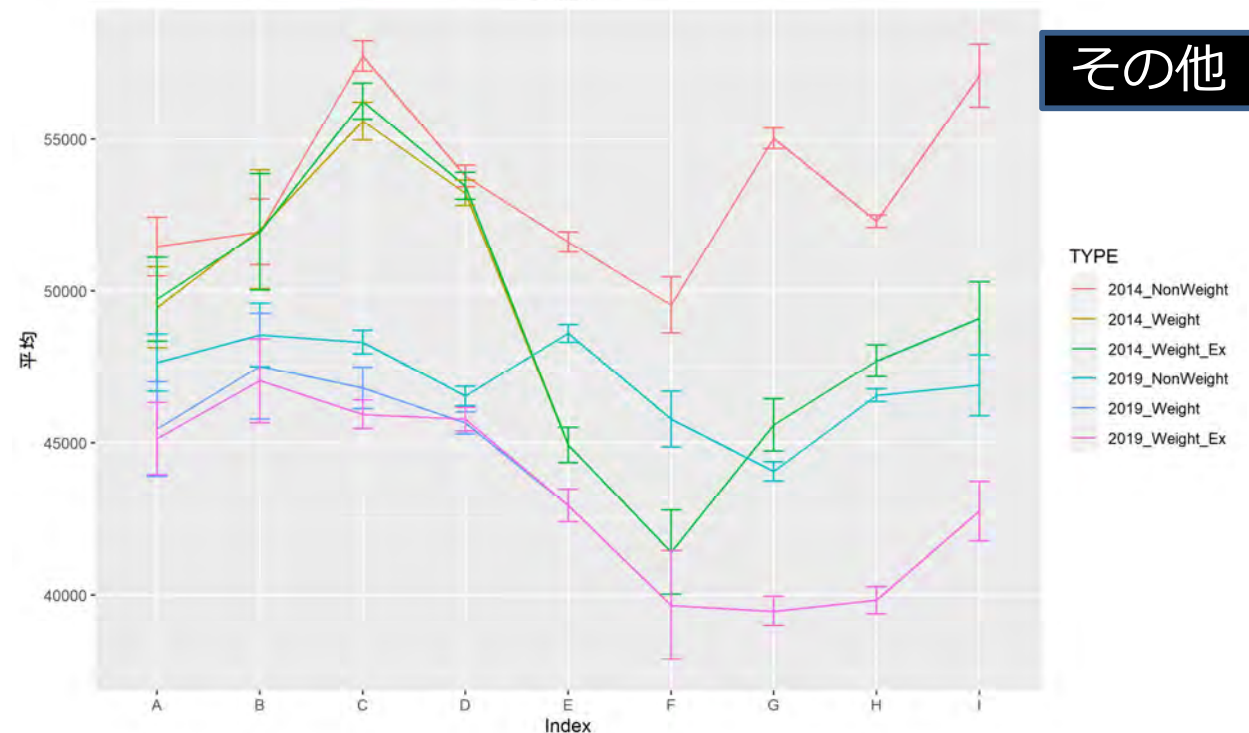
27

バーは95%信頼区間

- (A)年推定 (調査継続効果平均)①
- (B)年推定 (調査継続効果最小)②
- (C)調査継続効果1,2か月目での
年推定(回帰利用)③
- (D)調査継続効果1,2か月目での
年推定(平均利用)④
- (E)年平均 (家計)
- (F)9,10 月平均 (家計)
- (G)9,10 月平均 (全消)
- (H)年平均 (合計 = 家計と全消混合)
- (I)年平均
(家計 + 全消の効果 =
 $E + (G - F)$)



教養
娯楽



その他

4. 地域別の支出額・総支出

- 10地域別の支出額（2019年）
- 総支出（2014年・2019年）

2019年の推定・標準誤差(Tobit-type2 model)

29

住居

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

(G)9,10 月平均 (全消)

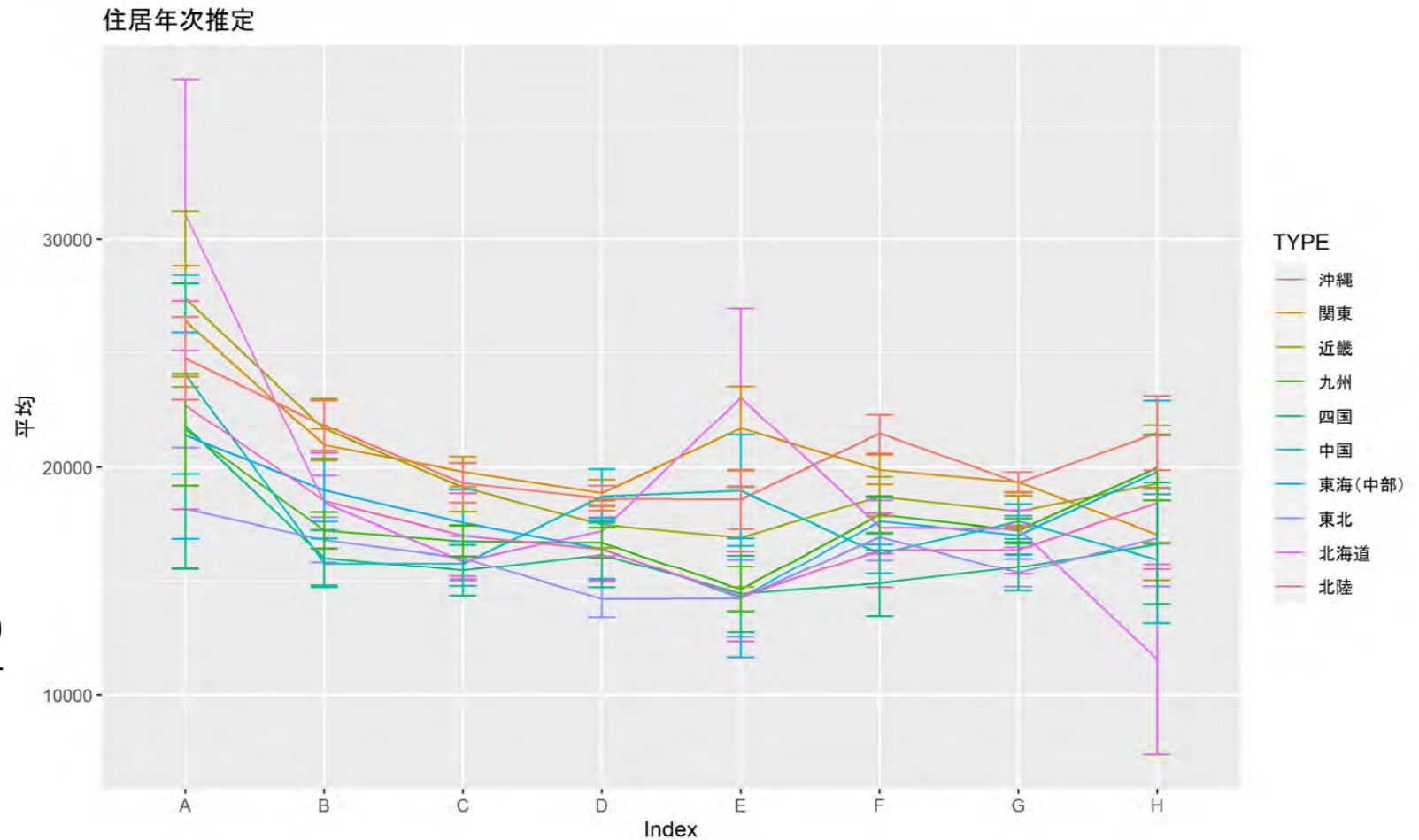
(H)年平均 (合計 = 家計
と全消混合)

(I)年平均

(家計 + 全消の効果

=

$$E + (G - F)$$



2019年の推定・標準誤差(Tobit-type2 model)

30

教育

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

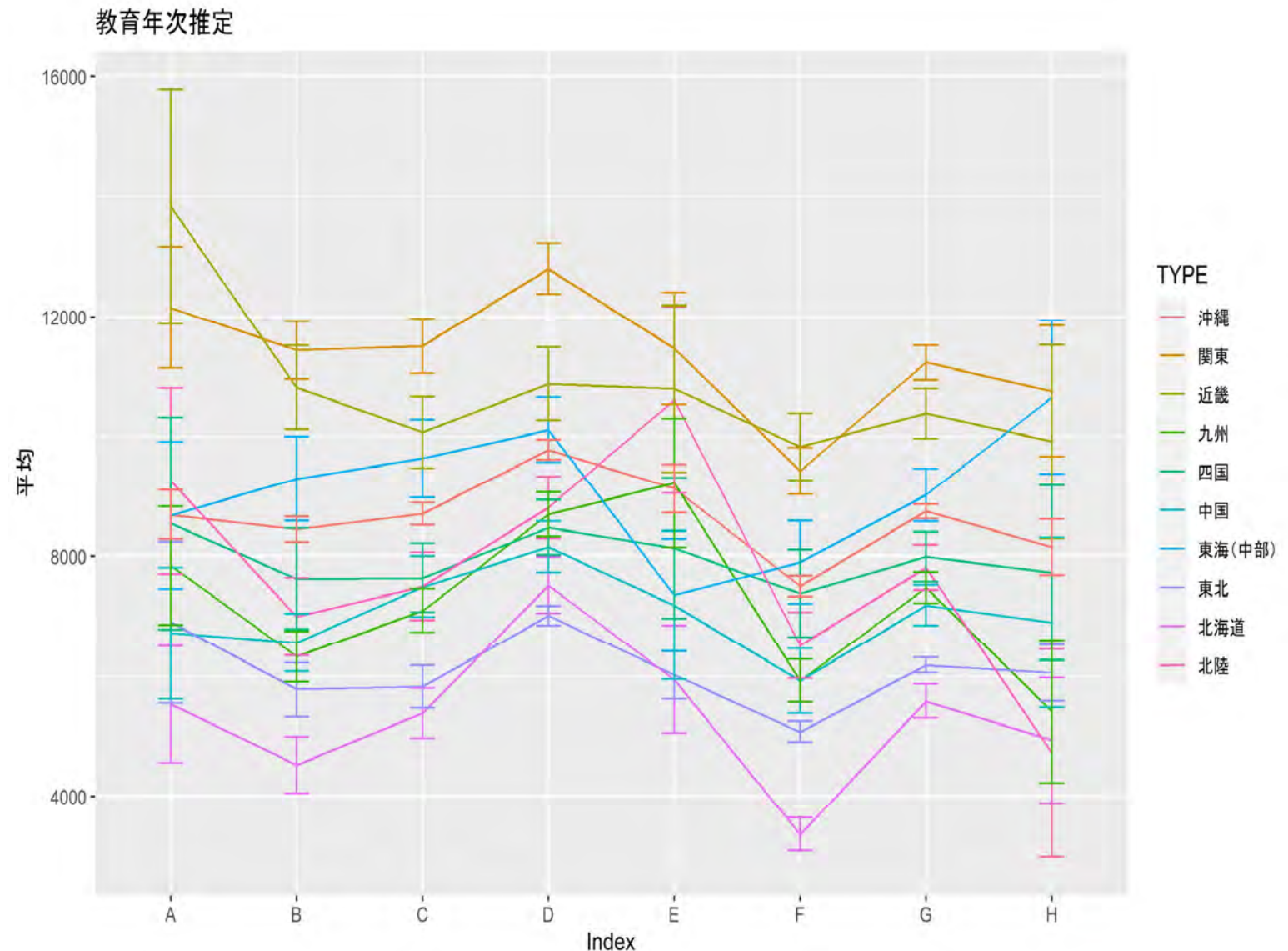
(G)9,10 月平均 (全消)

(H)年平均 (合計 = 家計
と全消混合)

(I)年平均
(家計 + 全消の効果

=

$$E + (G - F)$$



2019年の推定・標準誤差(Tobit-type2 model)

31

保険医療

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

(G)9,10 月平均 (全消)

(H)年平均 (合計 = 家計
と全消混合)

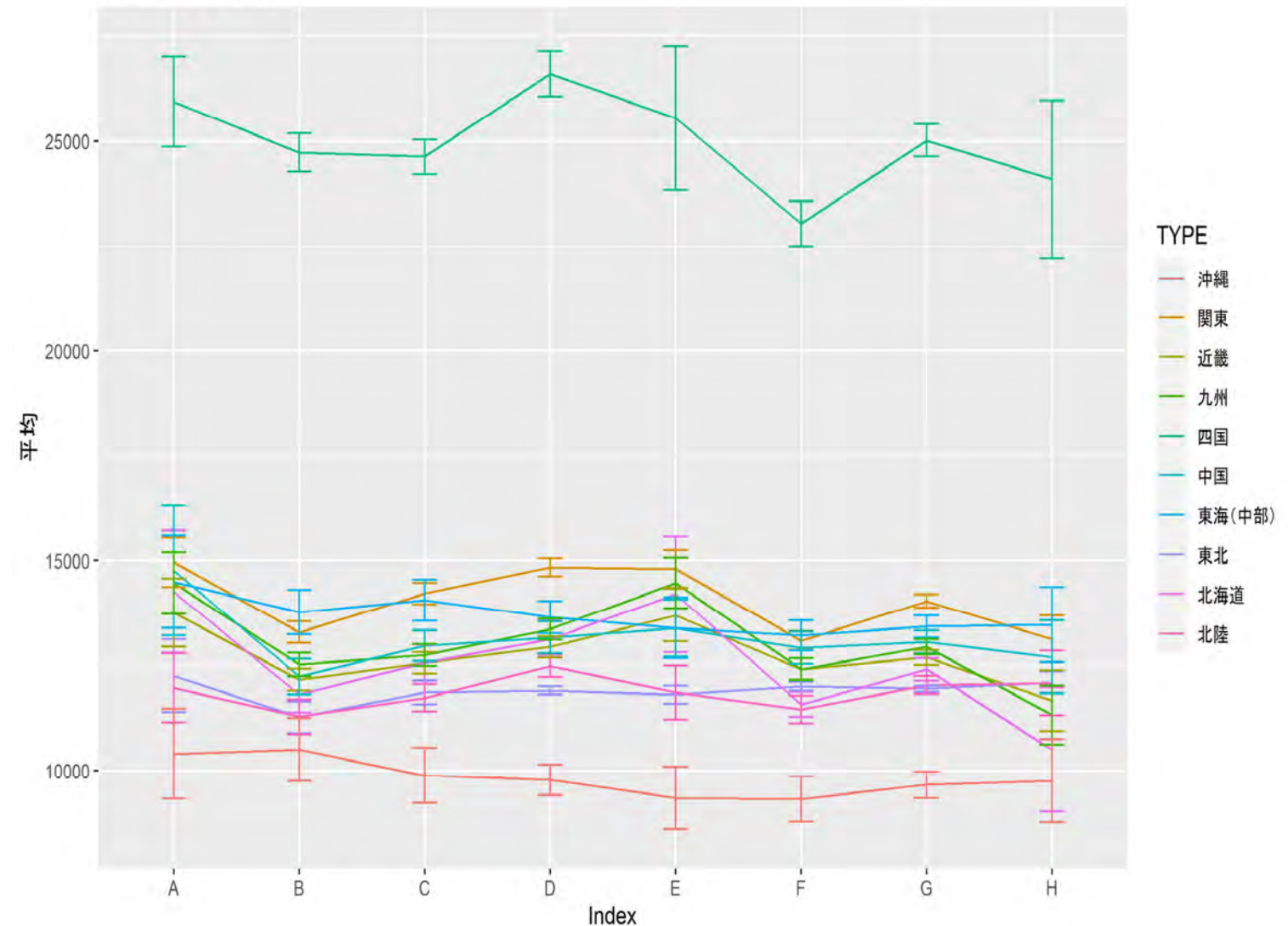
(I)年平均

(家計 + 全消の効果

=

$$E + (G - F)$$

保険医療年次推定



2019年の推定・標準誤差(Tobit-type2 model)

被服及び履物

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での

年推定(回帰利用)③

(D)調査継続効果1,2か月目での

年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

(G)9,10 月平均 (全消)

(H)年平均 (合計 = 家計と全消混合)

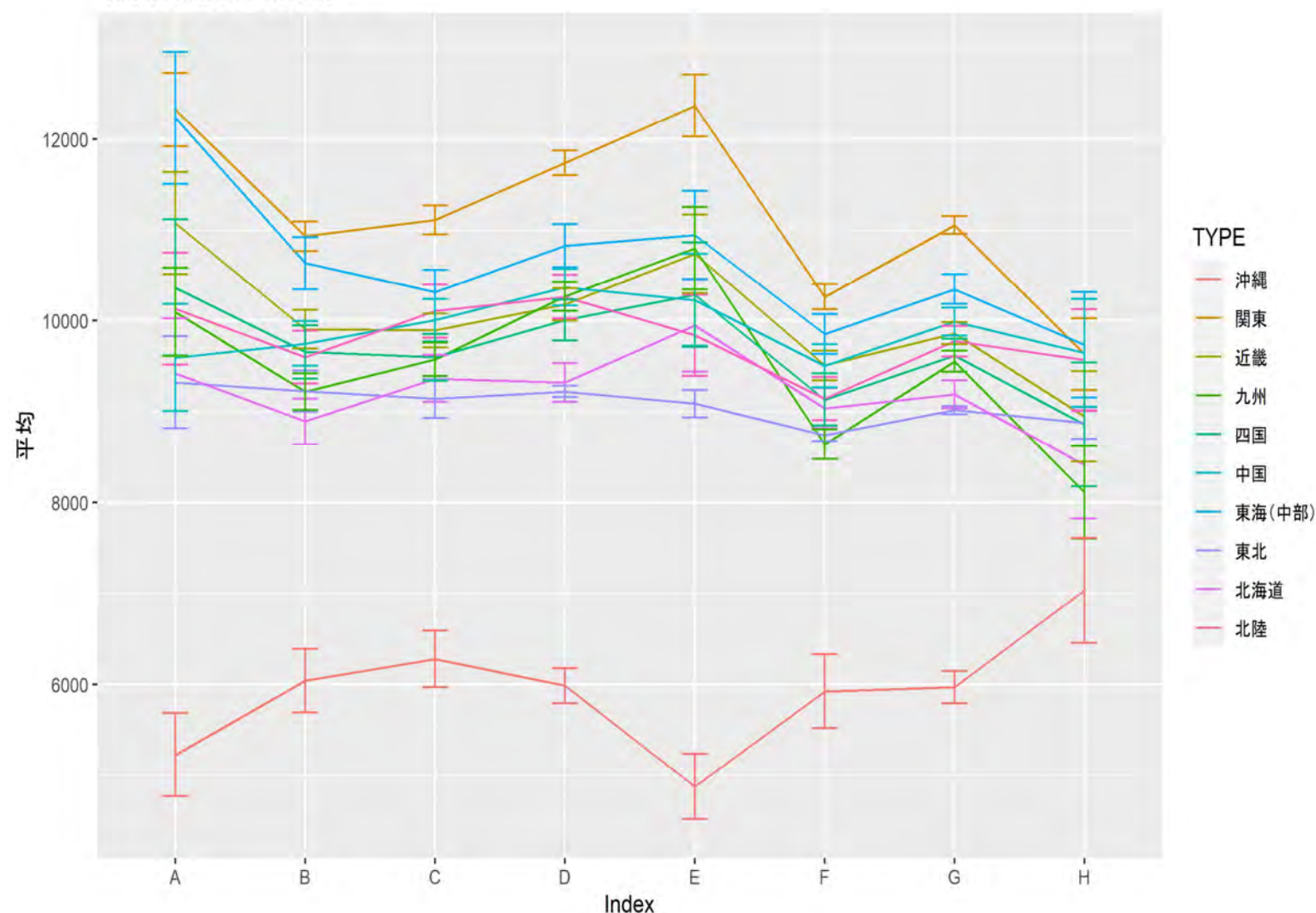
(I)年平均

(家計 + 全消の効果

=

$$E + (G - F)$$

被服及び履物年次推定



2019年の推定・標準誤差(Tobit-type2 model)

住居

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

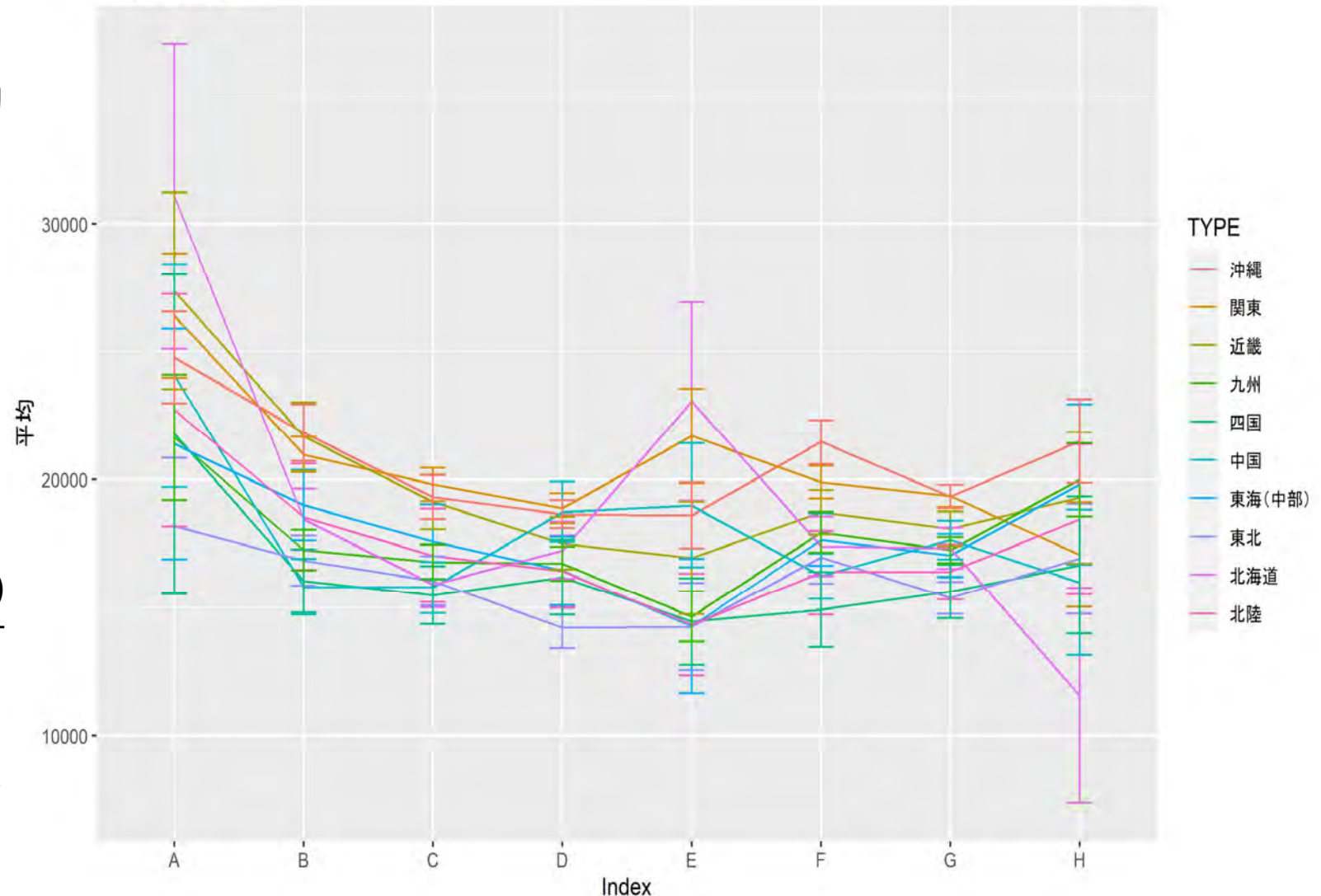
(G)9,10 月平均 (全消)

(H)年平均 (合計 = 家計
と全消混合)

(I)年平均
(家計 + 全消の効果

=
 $E + (G - F)$

住居年次推定



2019年の推定・標準誤差(重回帰)

食料

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

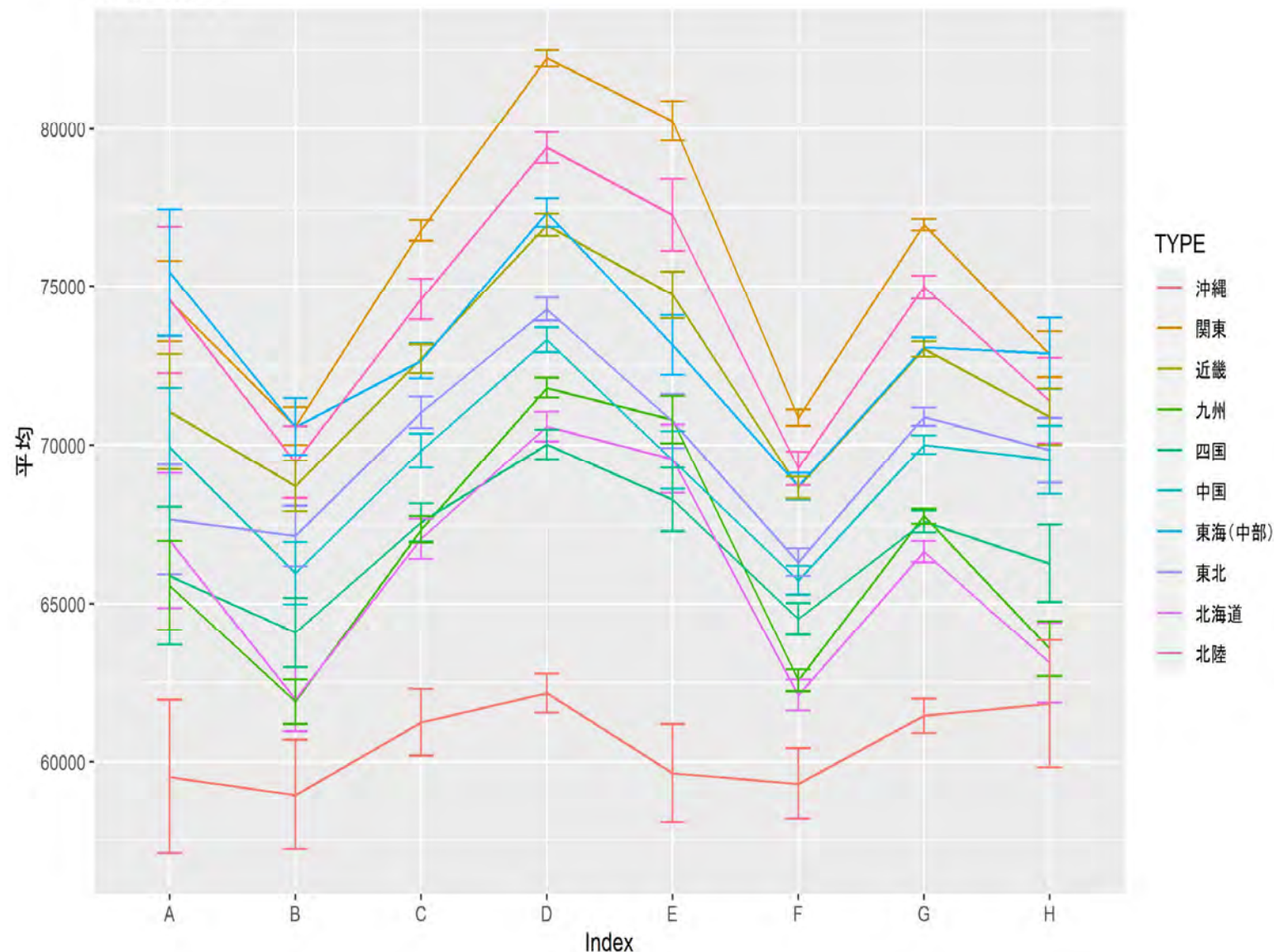
(G)9,10 月平均 (全消)

(H)年平均 (合計 = 家計
と全消混合)

(I)年平均
(家計 + 全消の効果

=
 $E + (G - F)$)

食料年次推定



2019年の推定・標準誤差(重回帰)

光熱水道

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での

年推定(回帰利用)③

(D)調査継続効果1,2か月目での

年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

(G)9,10 月平均 (全消)

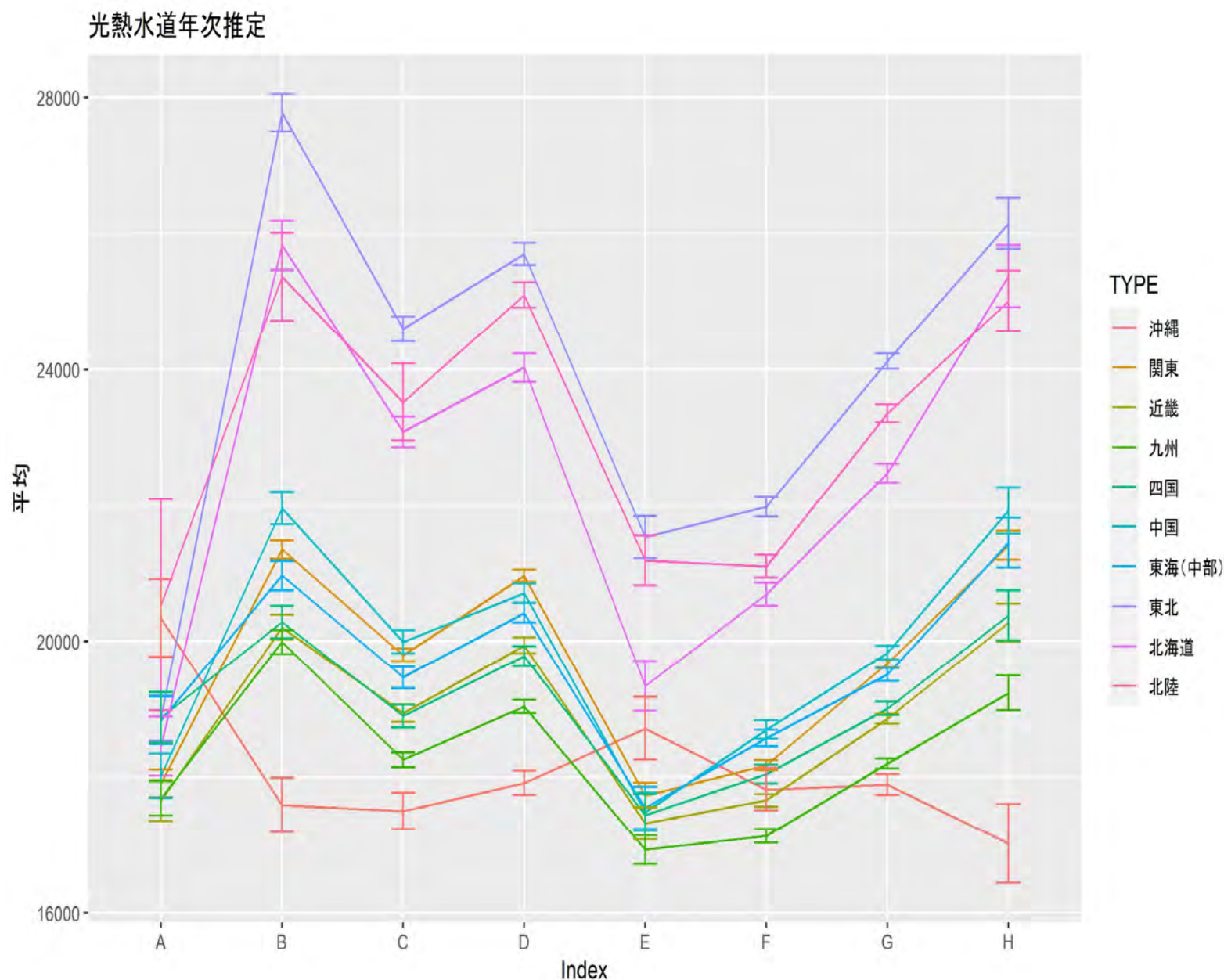
(H)年平均 (合計 = 家計と全消混合)

(I)年平均

(家計 + 全消の効果

=

$$E + (G - F)$$



2019年の推定・標準誤差(重回帰)

交通通信

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

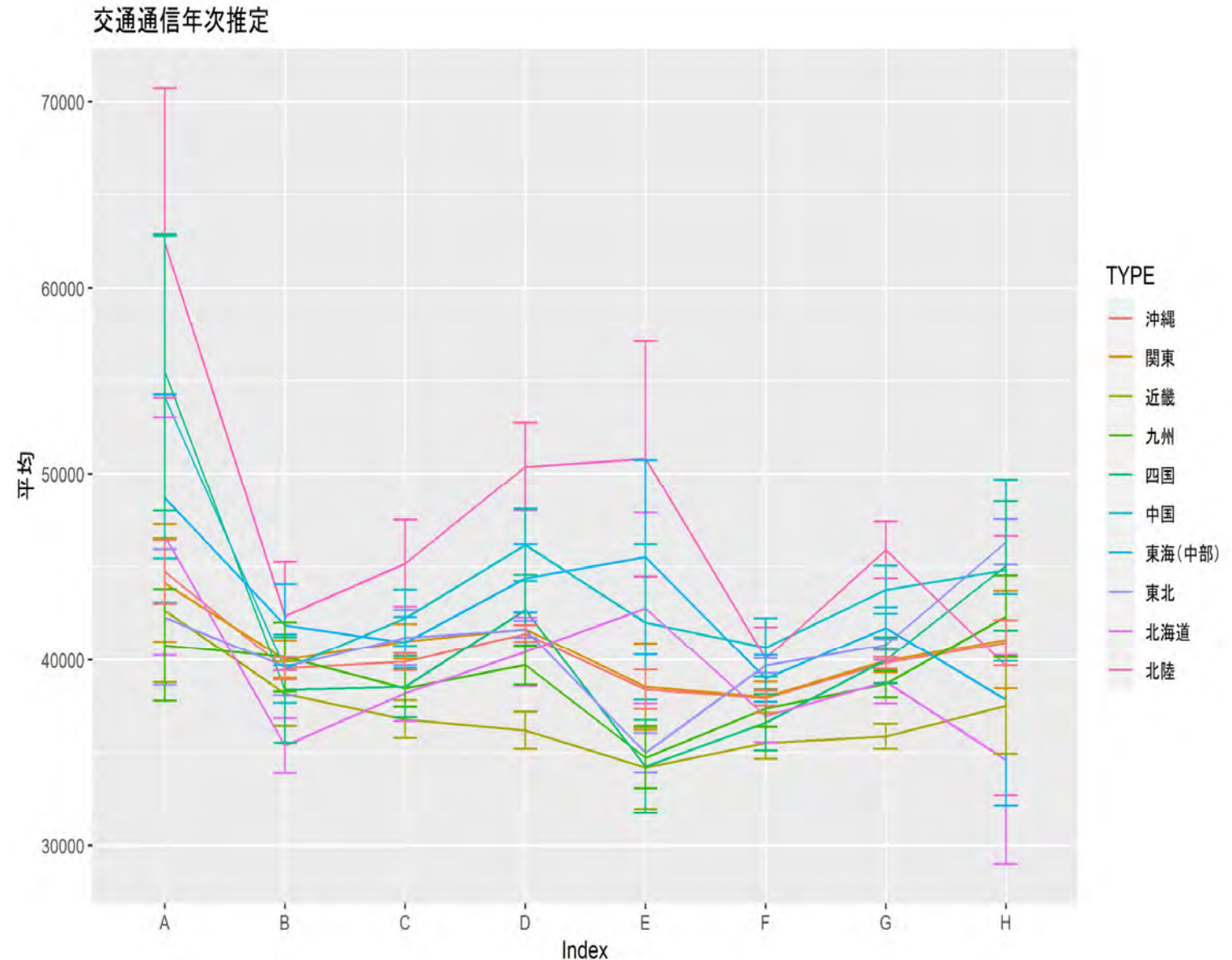
(G)9,10 月平均 (全消)

(H)年平均 (合計 = 家計
と全消混合)

(I)年平均
(家計 + 全消の効果

=

$$E + (G - F)$$



2019年の推定・標準誤差(重回帰)

教養娯楽

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

(G)9,10 月平均 (全消)

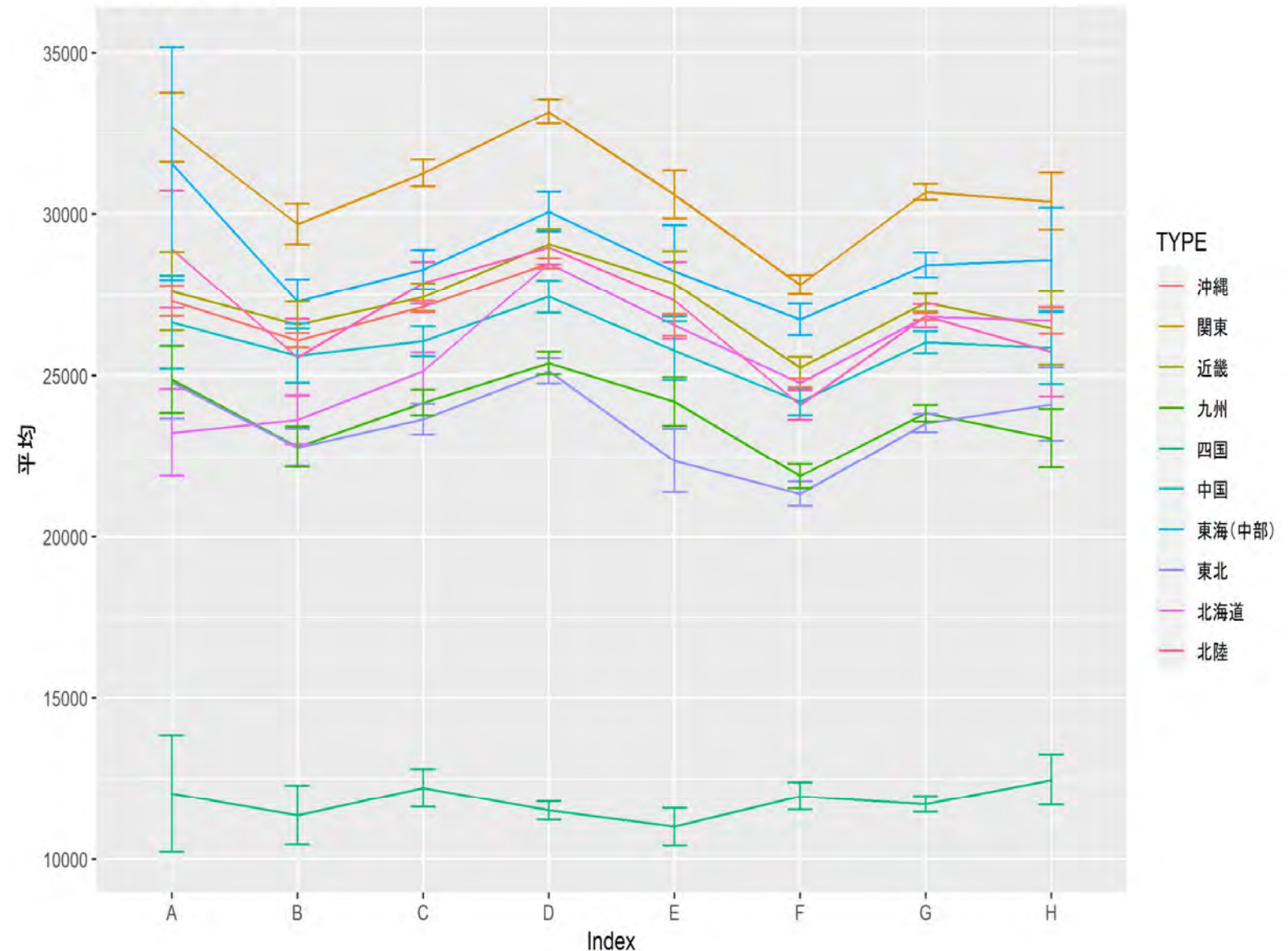
(H)年平均 (合計 = 家計
と全消混合)

(I)年平均

(家計 + 全消の効果

=
 $E + (G - F)$)

教養娯楽年次推定



2019年の推定・標準誤差(重回帰)

その他の消費支出

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

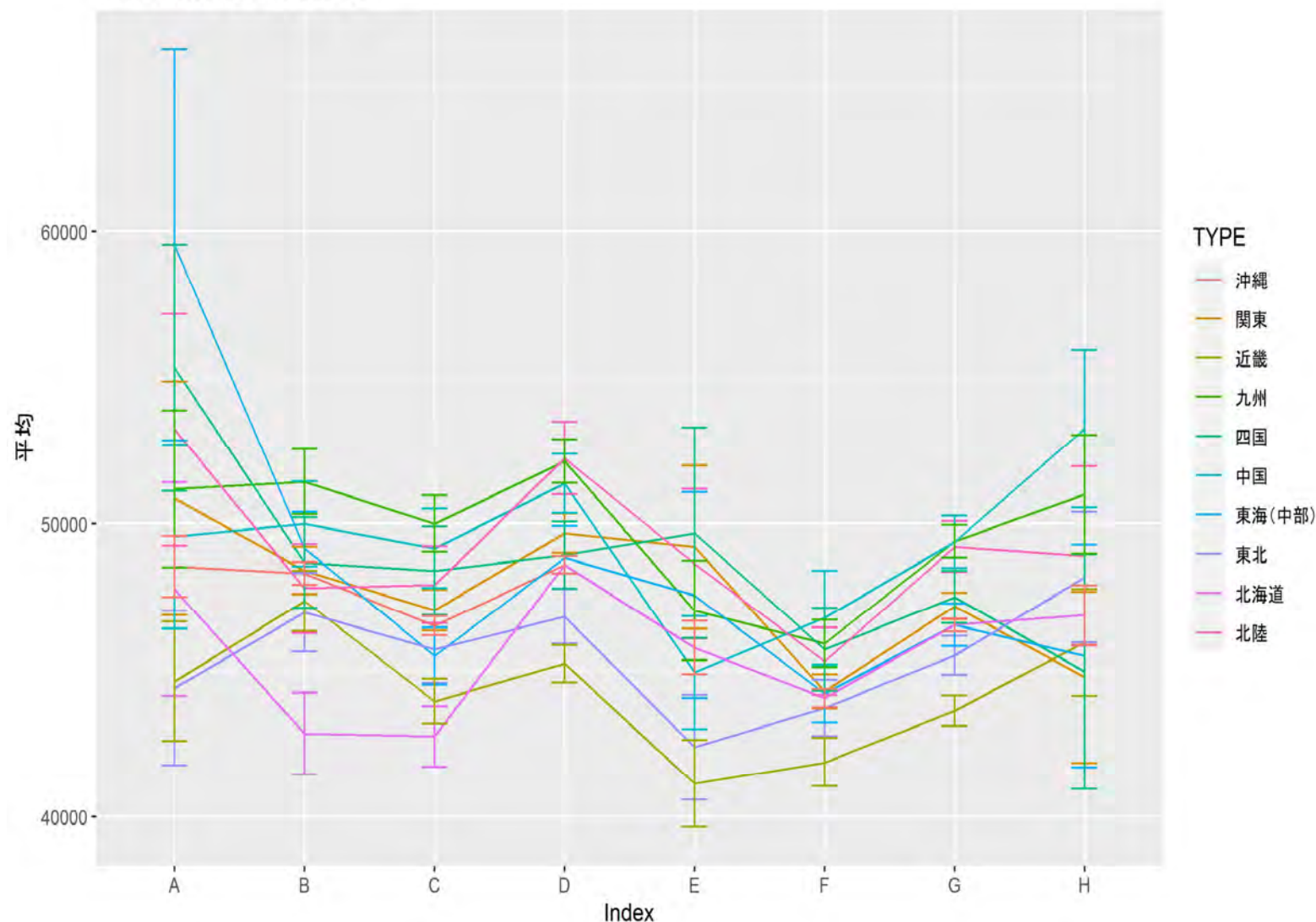
(G)9,10 月平均 (全消)

(H)年平均 (合計 = 家計
と全消混合)

(I)年平均
(家計 + 全消の効果

=
 $E + (G - F)$

その他の消費支出年次推定



2019年の推定・標準誤差(重回帰)

教養娯楽

バーは95%信頼区間

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

(G)9,10 月平均 (全消)

(H)年平均 (合計 = 家計
と全消混合)

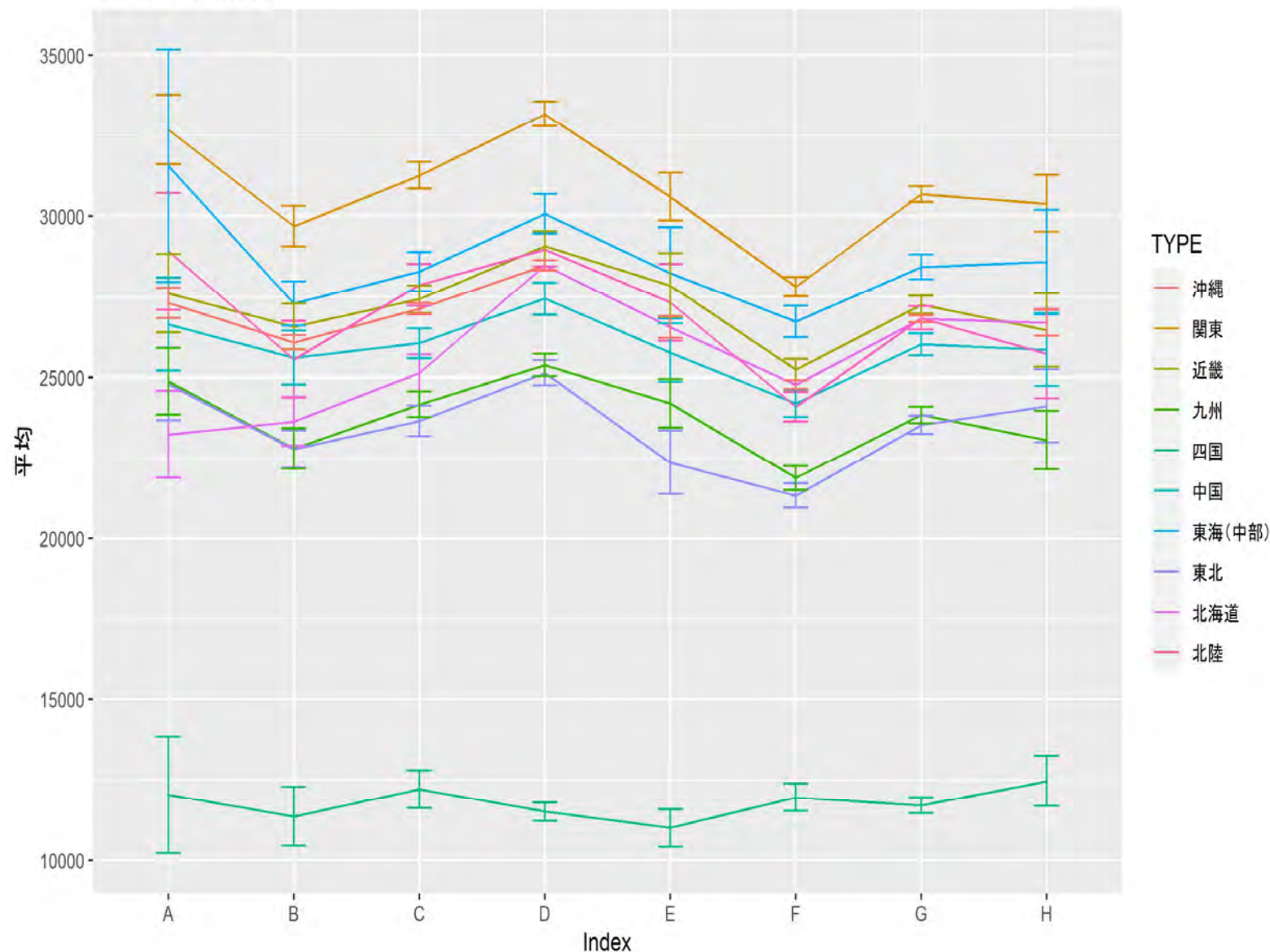
(I)年平均

(家計 + 全消の効果

=

$$E + (G - F)$$

教養娯楽年次推定



2014年・2019年平均の推定（総支出・全国）

総支出

バーは95%信頼区間

(A)年推定（調査継続効果平均）①

(B)年推定（調査継続効果最小）②

(C)調査継続効果1,2か月目での

年推定(回帰利用)③

(D)調査継続効果1,2か月目での

年推定(平均利用)④

(E)年平均（家計）

(F)9,10 月平均（家計）

(G)9,10 月平均（全消）

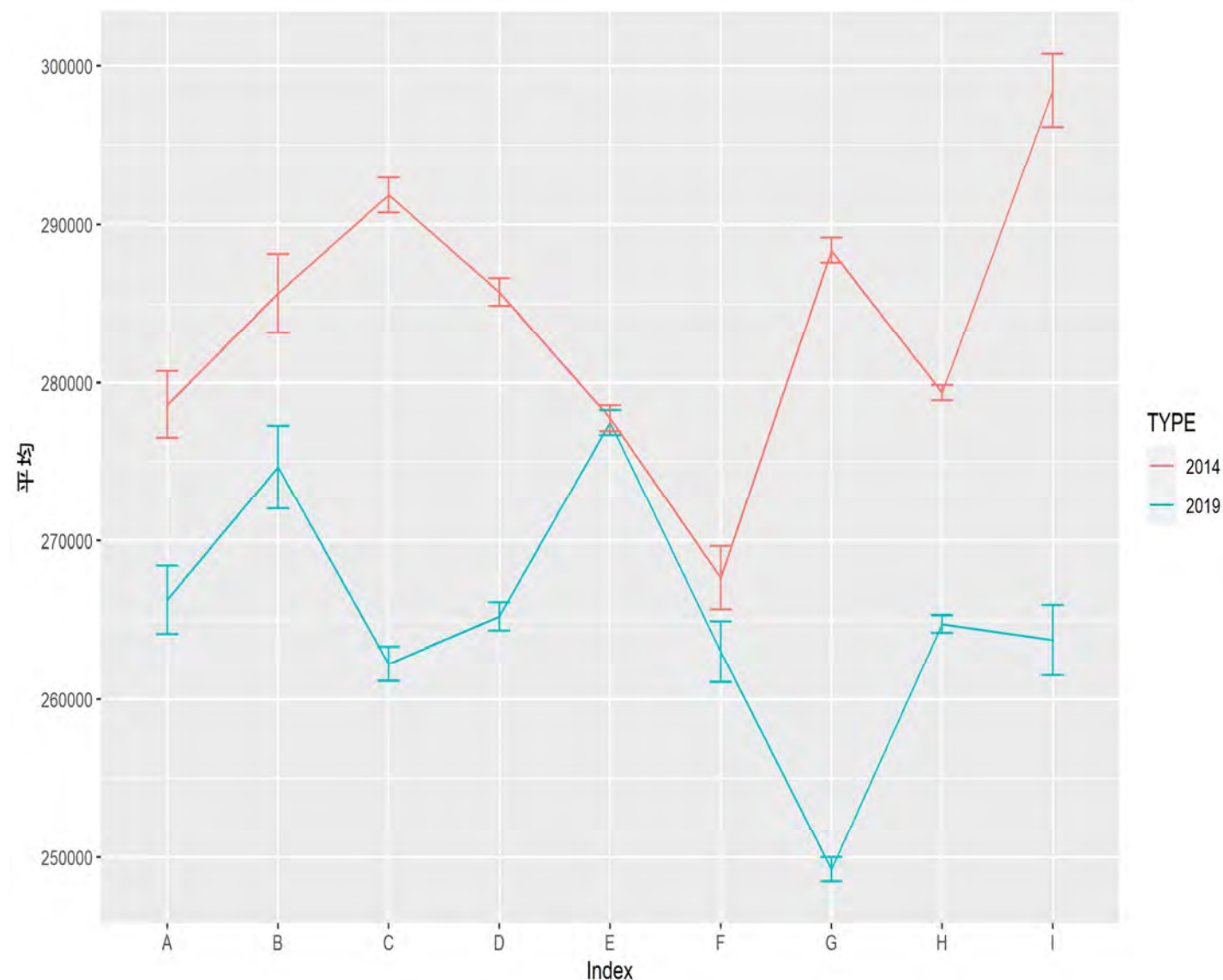
(H)年平均（合計＝家計と全消混合）

(I)年平均

（家計＋全消の効果

=

$$E + (G - F)$$



年平均の推定・標準誤差(2014年・総支出)

41

全国

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

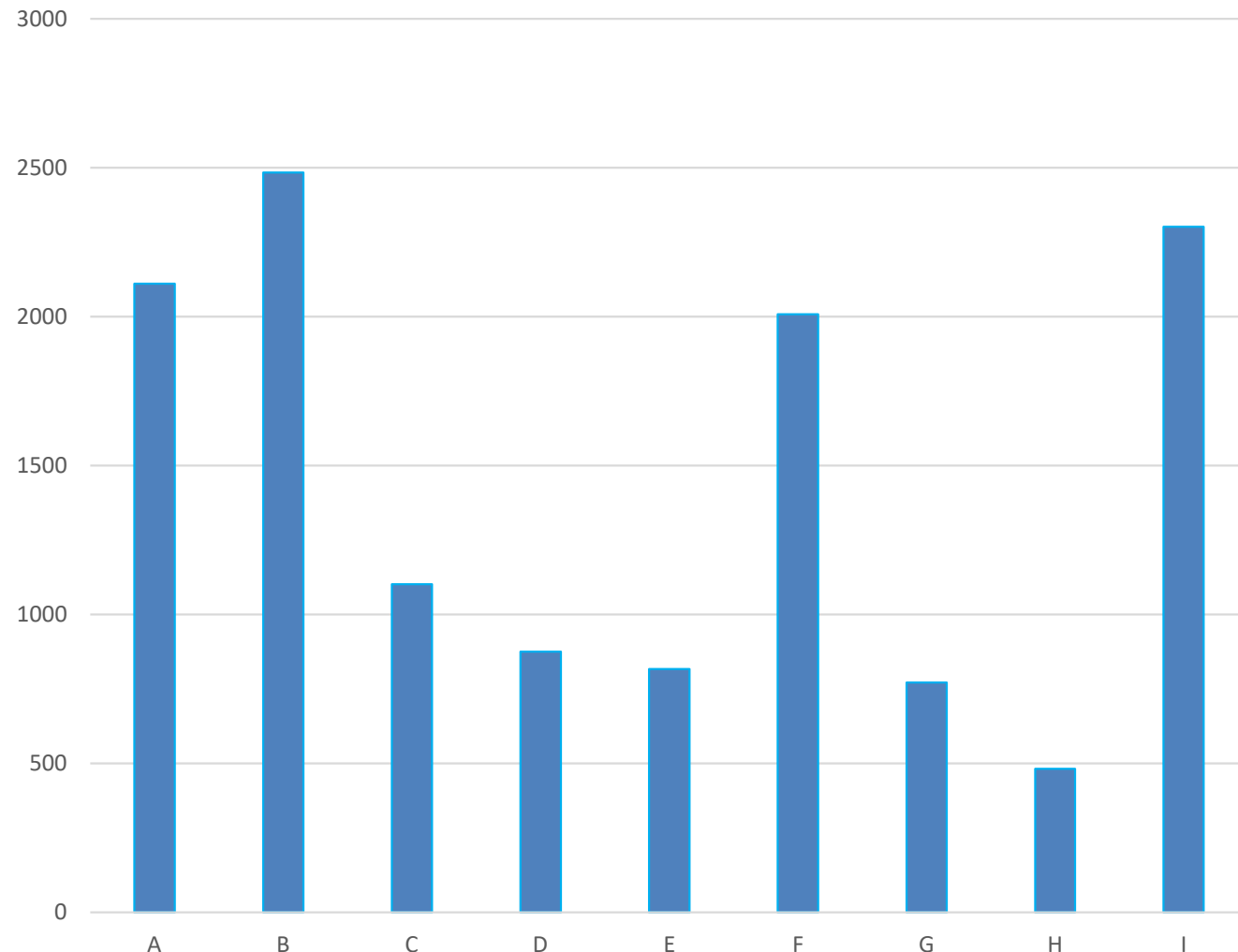
(G)9,10 月平均 (全消)

(H)年平均 (合計 = 家計
と全消混合)

(I)年平均
(家計 + 全消の効果

=
 $E + (G - F)$

全国 総支出 標準誤差



年平均の推定・標準誤差(2019年・総支出)

42

全国

(A)年推定 (調査継続効果平均)①

(B)年推定 (調査継続効果最小)②

(C)調査継続効果1,2か月目での
年推定(回帰利用)③

(D)調査継続効果1,2か月目での
年推定(平均利用)④

(E)年平均 (家計)

(F)9,10 月平均 (家計)

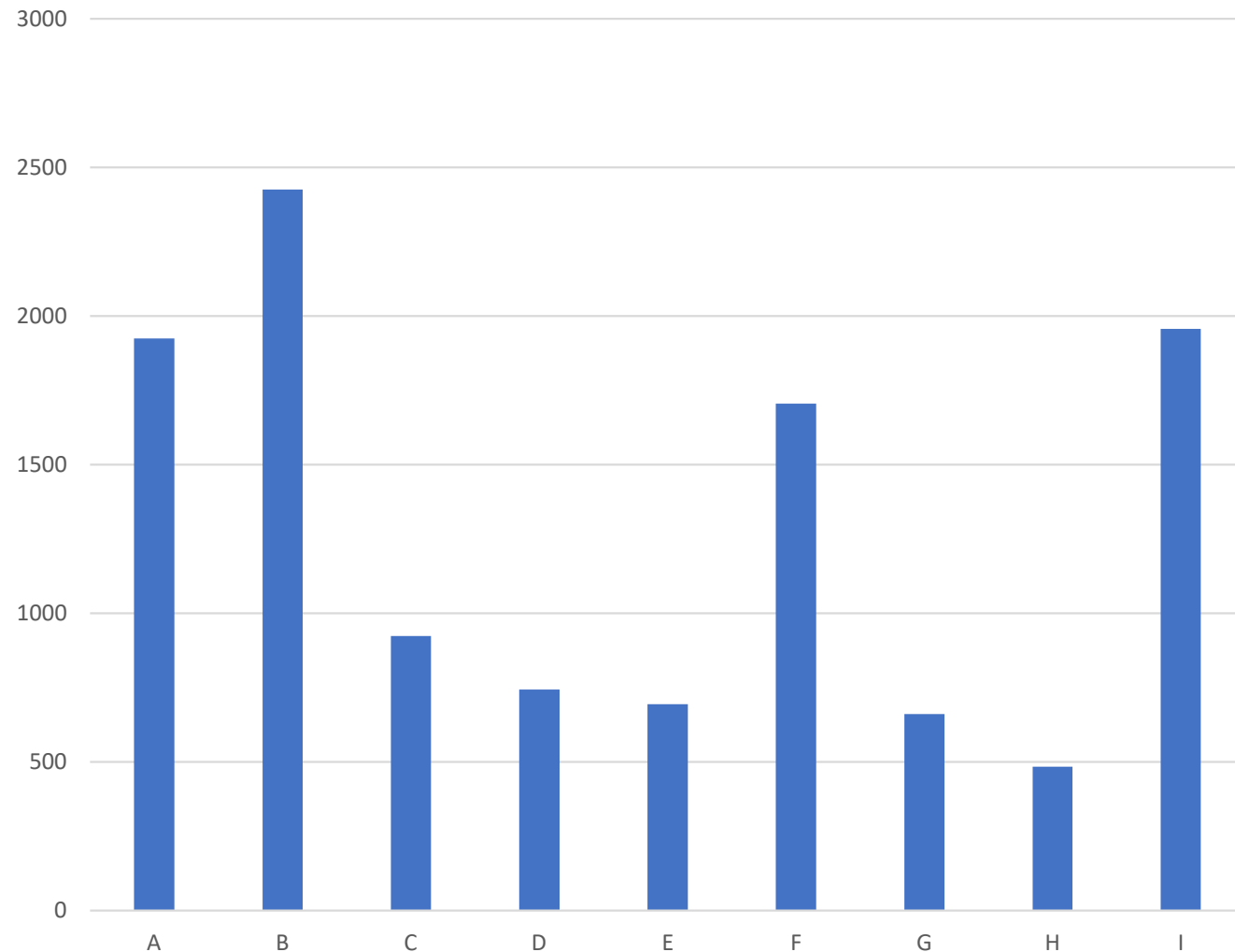
(G)9,10 月平均 (構造)

(H)年平均 (合計 = 家計
と構造混合)

(I)年平均
(家計 + 構造の効果
=

$$E + (G - F)$$

全国 総支出 標準誤差



議論

【まとめ】家計構造調査の年次集計に家計調査を利用する場合、過去の家計調査と家計構造調査の乖離の原因を理解する必要

“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

本研究では①～③を考慮する方法を提案

■ 2014～2019で調査継続効果は各費目比較的安定

■ 2 調査の融合により標準誤差を小さくすることができた

【今後の方向性】

2019年でのウェブ回答世帯に限った分析

前々から議論になっている“何が本来の年平均か？”の再議論

5. 資料

- 漸近論
- 2標本調査の具体的な違い（図示）
- 推定値および変数選択の基準等

参考：漸近論

- $N \rightarrow \infty$ のときの推定量の性質を述べる.
- 簡単化のため, ローターショナルパネルの欠測は無視する.
- 家計 i に対する確率変数のベクトルを Z_i で表す.
- 真のデータ生成過程がタイプII Tobitであり, その誤差項が Z_i と独立かつ, 次の多変量正規分布にi.i.d.に従うと仮定する.

$$\begin{pmatrix} \epsilon_i \\ \varepsilon_i \end{pmatrix} \equiv \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{i12} \\ \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{i12} \end{pmatrix} \sim N(O, \Sigma) = N\left(O, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right),$$

参考：漸近論

- ただし, Σ は次を満たすように定義された行列である.

$$\mathbb{E}[\epsilon_{t1}\epsilon_{s1}] = \begin{cases} 1 & \text{if } t = s \\ \sigma_{\text{cross}}^{(1)} & \text{if } t \neq s \end{cases},$$

$$\mathbb{E}[\epsilon_{it}\epsilon_{is}] = \begin{cases} \sigma^{(2)2} & \text{if } t = s \\ \sigma_{\text{cross}}^{(2)} & \text{if } t \neq s \end{cases},$$

$$\mathbb{E}[\epsilon_{it}\epsilon_{is}] = \begin{cases} \sigma^{(12)} & \text{if } t = s \\ \sigma_{\text{cross}}^{(12)} & \text{if } t \neq s \end{cases}$$

- 誤差の月間の相関を捉えるために, 次ページの12変量プロビットモデルと非線形モデルを考える.

参考：漸近論

- **1st step.** 12変量プロビットモデル (パラメータ π_1)

$$P(w_i | \pi_1) = \int_{A_{i12}} \dots \int_{A_{i1}} \phi_{12}(u | 0, \Sigma_{11}) du$$

ただし, $w_{it} = \mathbf{1}\{y_{it}^* > 0\}$, $w_i = (w_{i1}, w_{i2}, \dots, w_{i12})^T$, $\phi_{12}(u | 0, \Sigma_{11})$

は平均0, 分散 Σ_{11} の 12 変量正規分布の確率密度関数,

A_{it} は区間

$$A_{it} = \begin{cases} (-\infty, \lambda^T x_i + \phi) & \text{if } w_{it} = 1 \\ [\lambda^T x_i + \phi, \infty) & \text{if } w_{it} = 0 \end{cases}$$

- **2nd step.** 非線形モデル (パラメータ π_1, π_2)

$$y_i = H(Z_i, \pi_1, \pi_2) + \sigma^{(12)} h(x_i, \pi_1) + U_i$$

ただし,

$$H(Z_i, \pi_1, \pi_2) = \begin{pmatrix} \phi(\lambda^T x_i + \phi) \left(\gamma^T x_i + \sum_{m=1}^{12} \alpha_m D_{i1} + \sum_{k=2}^6 \beta_k E_{i1k} \right) \\ \phi(\lambda^T x_i + \phi) \left(\gamma^T x_i + \sum_{m=1}^{12} \alpha_m D_{i2} + \sum_{k=2}^6 \beta_k E_{i2k} \right) \\ \vdots \\ \phi(\lambda^T x_i + \phi) \left(\gamma^T x_i + \sum_{m=1}^{12} \alpha_m D_{i12} + \sum_{k=2}^6 \beta_k E_{i12k} \right) \end{pmatrix}, h(x_i, \pi_1) = \begin{pmatrix} \phi(\lambda^T x_i + \phi) \\ \phi(\lambda^T x_i + \phi) \\ \vdots \\ \phi(\lambda^T x_i + \phi) \end{pmatrix}.$$

参考：漸近論

- 1st step目の推定方程式は，対数尤度を偏微分して

$$m_1(Z_i, \pi_1) = \frac{\partial}{\partial \pi_1} \log P(w_i | \pi_1)$$

- 2nd step目の推定方程式は，非線形最小二乗法の一階の条件より

$$m_2(Z_i, \pi_1, \pi_2) = \frac{\partial}{\partial (\pi_1^T, \pi_2^T)^T} \left\{ y_i - H(Z_i, \pi_1, \pi_2) - \sigma^{(12)} h(x_i, \pi_1) \right\}^T \left\{ y_i - H(Z_i, \pi_1, \pi_2) - \sigma^{(12)} h(x_i, \pi_1) \right\}$$

- $\pi = (\pi_1^T, \pi_2^T)^T$ とおく．
- スタックした推定方程式

$$m(Z_i, \pi) = \begin{bmatrix} m_1(Z_i, \pi_1) \\ m_2(Z_i, \pi_1, \pi_2) \end{bmatrix}$$

をモーメント条件として扱う．

- 正則な条件のもとで一致性と漸近正規性を得る．

参考：漸近論

- 特に, 2nd step目のモデルの推定量は次の漸近正規性を持つ.

$$\sqrt{N}(\hat{\pi}_2 - \pi_{20}) \xrightarrow{d} N(0, V)$$

ただし,

$$V = M_{22}^{-1} \mathbb{E} \left[\{m_2(Z_i, \pi_{10}, \pi_{20}) + M_{21}\psi(Z_i)\} \{m_2(Z_i, \pi_{10}, \pi_{20}) + M_{21}\psi(Z_i)\}^T \right] (M_{22}^{-1})^T$$

かつ,

$$M_{22} = \mathbb{E} \left[\frac{\partial}{\partial \pi_2^T} m_2(Z_i, \pi_{10}, \pi_{20}) \right], M_{21} = \mathbb{E} \left[\frac{\partial}{\partial \pi_1^T} m_2(Z_i, \pi_{10}, \pi_{20}) \right],$$

$$M_{11} = \mathbb{E} \left[\frac{\partial}{\partial \pi_1^T} m_1(Z_i, \pi_{10}) \right], \psi(Z_i) = -M_{11}^{-1} m_1(Z_i, \pi_{10})$$

- 証明はNewey and McFadden (1994) を参照されたい.

分析で出した指標

「全国消費実態調査・家計構造調査（添字 zi ）を1年やった場合」の結果

A. 1年間「家計調査での平均的な“調査継続による効果”」の元での年平均

$$\mu_{avg} = E \left[\gamma^T x_{zi} + \sum_{m=1}^{12} \alpha_m D_{zit} + \frac{1}{5} \sum_{k=2}^6 \beta_k + \sigma \lambda_{mills, zi} \mid y_{zi} > 0 \right] P(y_{zi} > 0)$$

の推定、具体的には

$$\widehat{\mu_{avg}} = \frac{1}{\#(y_{zi} > 0)} \sum_{i: y_{zi} > 0} \left[\hat{\gamma}^T x_{zi} + \hat{\alpha}_m D_{zit} + \frac{1}{5} \sum_{k=2}^6 \hat{\beta}_k + \hat{\sigma} \hat{\lambda}_{mills, zi} \right] \frac{\#(y_{zi} > 0)}{N_Z}$$

B. なるべく調査継続による効果が一番少ない場合の下での年平均

$$\mu_{min} = E \left[\gamma^T x_{zi} + \sum_{m=1}^{12} \alpha_m D_{zit} + \min(\beta_k) + \sigma \lambda_{mills} \mid y_{zi} > 0 \right] P(y_{zi} > 0)$$

の推定、具体的には

$$\widehat{\mu_{min}} = \frac{1}{\#(y_{zi} > 0)} \sum_{i: y_{zi} > 0} \left[\hat{\gamma}^T x_{zi} + \hat{\alpha}_m D_{zit} + \min(\hat{\beta}_k) + \hat{\sigma} \hat{\lambda}_{mills, zi} \right] \frac{\#(y_{zi} > 0)}{N_Z}$$

を計算する。但し $\hat{\gamma}, \hat{\alpha}_m, \hat{\beta}_k, \hat{\sigma}$ と逆ミルズ比 λ_{mills} のパラメータは家計調査から推定。

分析で出した指標

C. 「全国消費実態調査・家計構造調査を1年やった場合」の結果

但し1,2か月目の調査継続による効果にしたvar.1(10,11月は構造利用)

$$\frac{N_{Z10}\bar{y}_{Z10}}{N} + \frac{N_{Z11}\bar{y}_{Z11}}{N} + \sum_{m \neq 10,11}^{12} \frac{N_{Am}}{N} \frac{1}{\#(y_{Zi} > 0)} \sum_{i: y_{Zi} > 0} \left[\hat{\alpha}_m + \frac{\hat{\beta}_1 + \hat{\beta}_2}{2} + \hat{\gamma}^T x_{Zi} + \hat{\sigma} \hat{\lambda}_{mills,Zi} \right] \frac{\#(y_{Zi} > 0)}{N_Z}$$

* N_{Zm}, N_{Am} は全国消費実態調査・家計構造調査と家計調査のm月の人数,

* N_Z は全国消費実態調査・家計構造調査の人数,

* N_A は家計調査の10,11月以外の月の人数,

* $N = N_Z + N_A$,

* $\bar{y}_{Zm}$ は全国消費実態調査・家計構造調査のm月平均. ここで,

$$\frac{1}{\#(y_{Zi} > 0)} \sum_{i: y_{Zi} > 0} \left[\hat{\alpha}_m + \frac{\hat{\beta}_1 + \hat{\beta}_2}{2} + \hat{\gamma}^T x_{Zi} + \hat{\sigma} \hat{\lambda}_{mills,Zi} \right] \frac{\#(y_{Zi} > 0)}{N_Z}$$

は $E \left[\alpha_m + \frac{\beta_1 + \beta_2}{2} + \gamma^T x_{Zi} + \sigma \lambda_{mills} | y_{Zi} > 0 \right] P(y_{Zi} > 0) + 0 * P(y_{Zi} = 0)$ の推定値

分析で出した指標

D. 「全国消費実態調査・家計構造調査を1年やった場合」の結果、

但し1, 2か月目の調査継続による効果にしたvar. 2 (10, 11月は構造調査利用)

* 共変量調整を利用せずに平均値をもとに推定する場合

= 1~9、12月は「1, 2か月目の調査継続による効果にした家計調査」

$$\frac{N_{Z10}\bar{y}_{Z10}}{N} + \frac{N_{Z11}\bar{y}_{Z11}}{N} + \frac{1}{N} \sum_{m \neq 10,11}^{12} N_{Am} \left[\bar{y}_{Am} + \left(\frac{\hat{\beta}_1 + \hat{\beta}_2}{2} - \frac{1}{5} \sum_{k=2}^6 \hat{\beta}_k \right) \frac{\#(y_{Aim} > 0)}{N_{Am}} \right]$$

* N_{Zm}, N_{Am} は全国消費実態調査・家計構造調査と家計調査のm月の人数,

* N_Z は全国消費実態調査・家計構造調査の人数

* N_A は家計調査の10,11月以外の月の人数,

* $N = N_Z + N_A$,

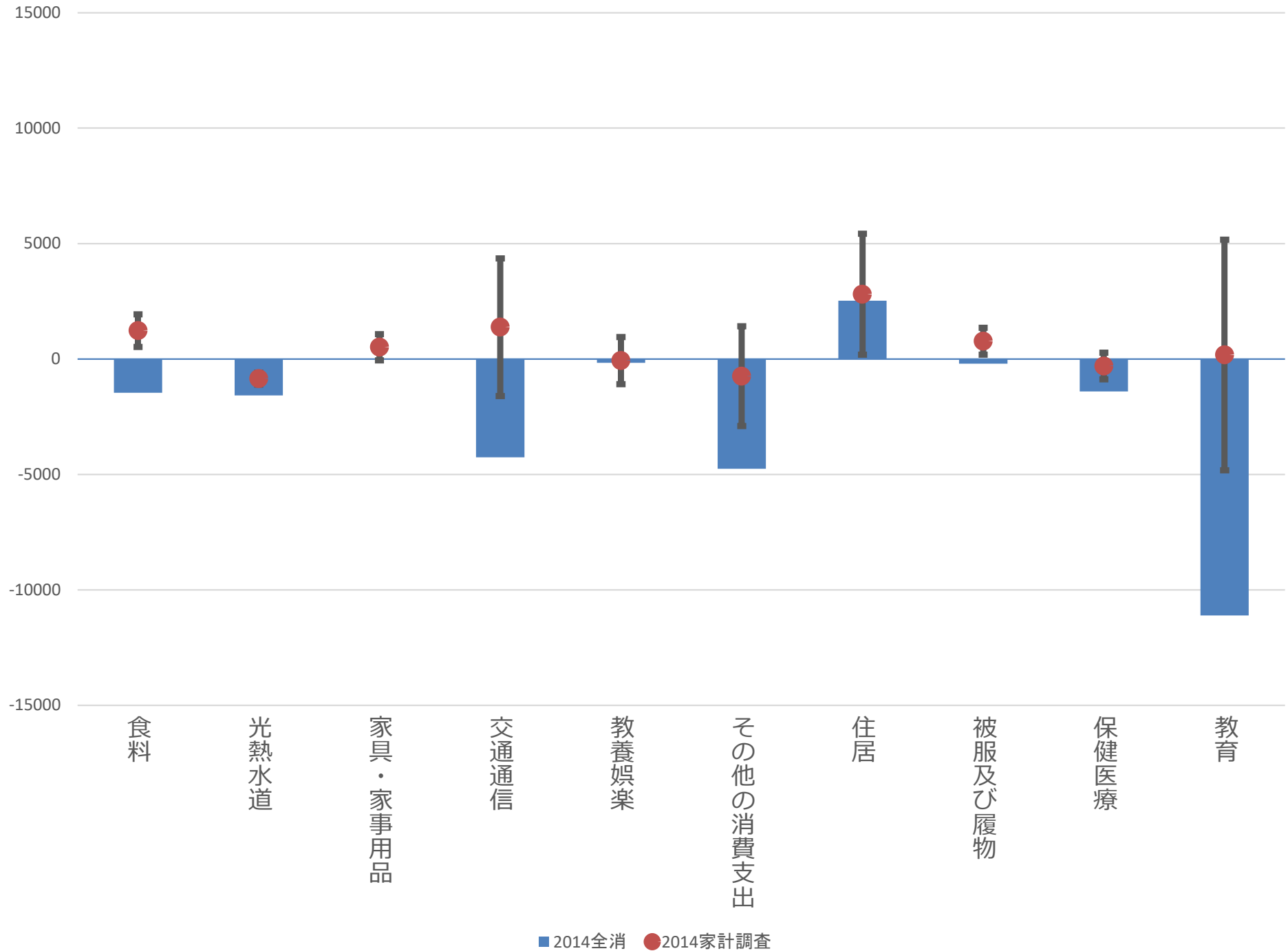
* $\bar{y}_{Zm}$ は全国消費実態調査・家計構造調査のm月平均

推定値の結果

- 先述したA~Dの4指標に家計構造調査や家計調査の平均をとった4つの指標を加え、8つの指標を図示
- 2014年と2019年それぞれで「重率なし」「重率あり」「重みづけあり、重率算出に使用された変数を回帰で無使用」の3パターンを計算
 - ◆ 重率（ウエイト補正）は、家計調査や家計構造調査等で世帯人員や地域等で行われている補正
 - ◆ 2014年の全国消費実態調査では、地域区分、世帯人員
 - ◆ 2019年の家計構造調査では、世帯主の年齢、性別、世帯人員、世帯主の就業状況、都道府県
- Cが最も安定しているとみられるが、議論中
 - ◆ 2標本の平均であるE~Hが「本当の平均」とは限らず、難しい

家計調査と全消における2ヶ月目効果比較(2014年)

54



家計調査と家計構造における2ヶ月目効果比較(2019年)

55

