

全国家計構造調査 「年平均推定値」の推定方法について

理化学研究所革新知能統合研究センター 星野崇宏*
慶應義塾大学大学院経済学研究科 慶野有輝

問題意識

全国消費実態調査（全国家計構造調査）のメリット／デメリット

サンプルサイズが大きい→詳細な区分での分析

2か月(前回3か月) 5年に一度

家計調査のメリット／デメリット

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

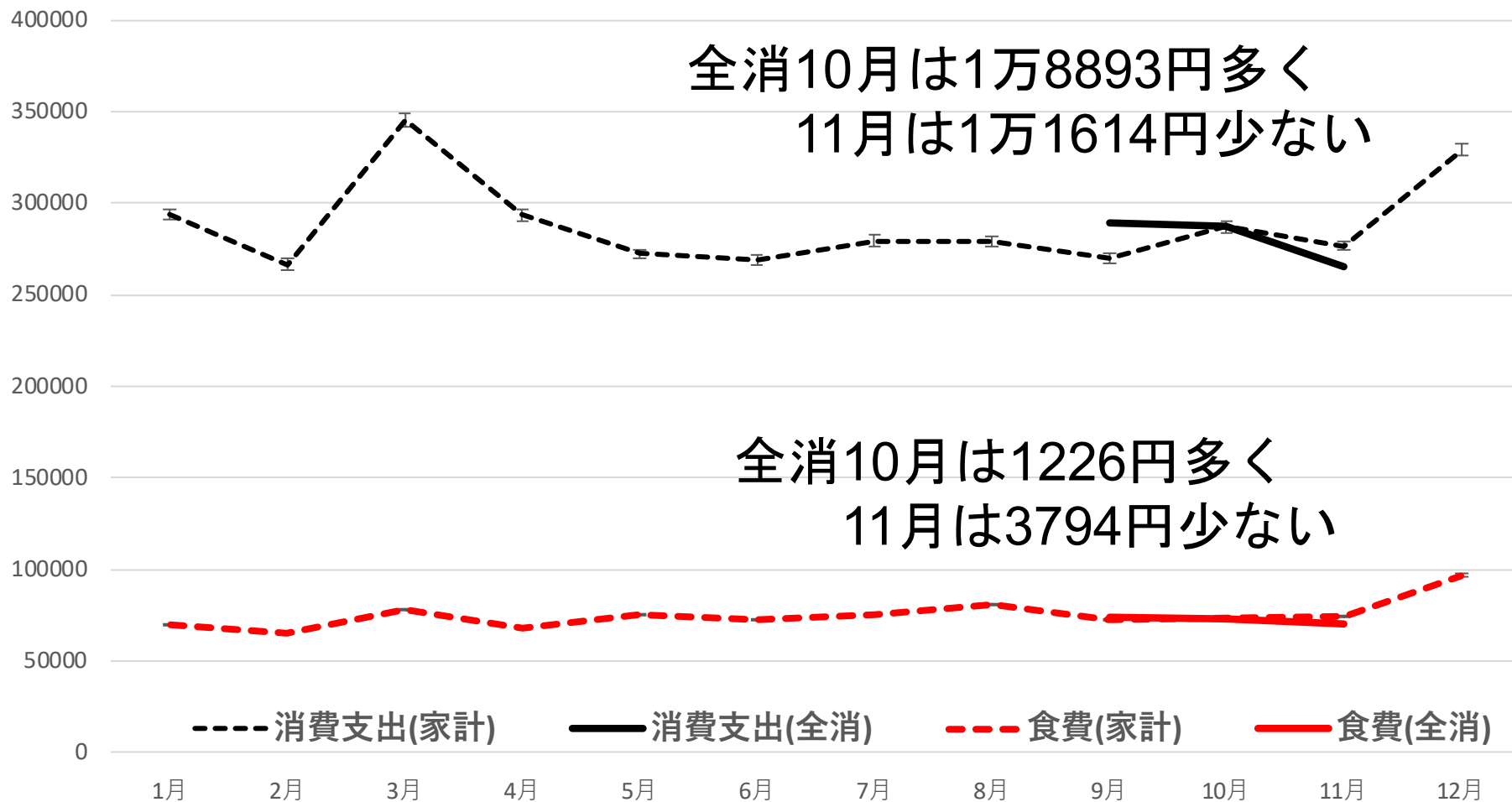
食費など月次の変動が大きいものについて両者を融合させて

通年での推測をしたい

○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



* 過小記入バイアスや儉約化(過小記入バイアスDeaton&Irish,1984;牧,2007; 儉約化・調査疲れStephens&Unayama,2011)の影響かもしれない。
特に全消では3か月の間に減少(家計調査では1~6か月目が混在して消去?)

家計調査と家計構造(旧全消)の違い

“選択バイアス” = 回答集団の違い

“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
回答集団の違い

家計調査回答者

全国消費実態調査回答者

家計調査の調査票	家計調査の結果	欠測
全消の調査票	欠測	全消の結果
補助変数・共変量	回答集団間の違いが生じる属性 (性年代・職種・収入等)	

調査モードの違い

目的と方法

【目的】 月次の変動が大きい消費について両調査を融合させて通年での推測／年平均の計算

【方法論】 個票データで欠測が存在していることを仮定した

欠測データ解析

具体的にはEMアルゴリズムを用いた平均と相関構造の推定

今回は多変量 t 分布を仮定したEMアルゴリズム

【注意点】

1) 集計時系列解析での季節性を考慮する方法とは異なる

* 集計での季節性考慮には家計構造の時点数が足りない

2) 家計調査と全国消費実態調査の回答者の標本誤差や過小記載の構造・項目（以後“調査モード”と呼称）などが同じとする仮定で実施

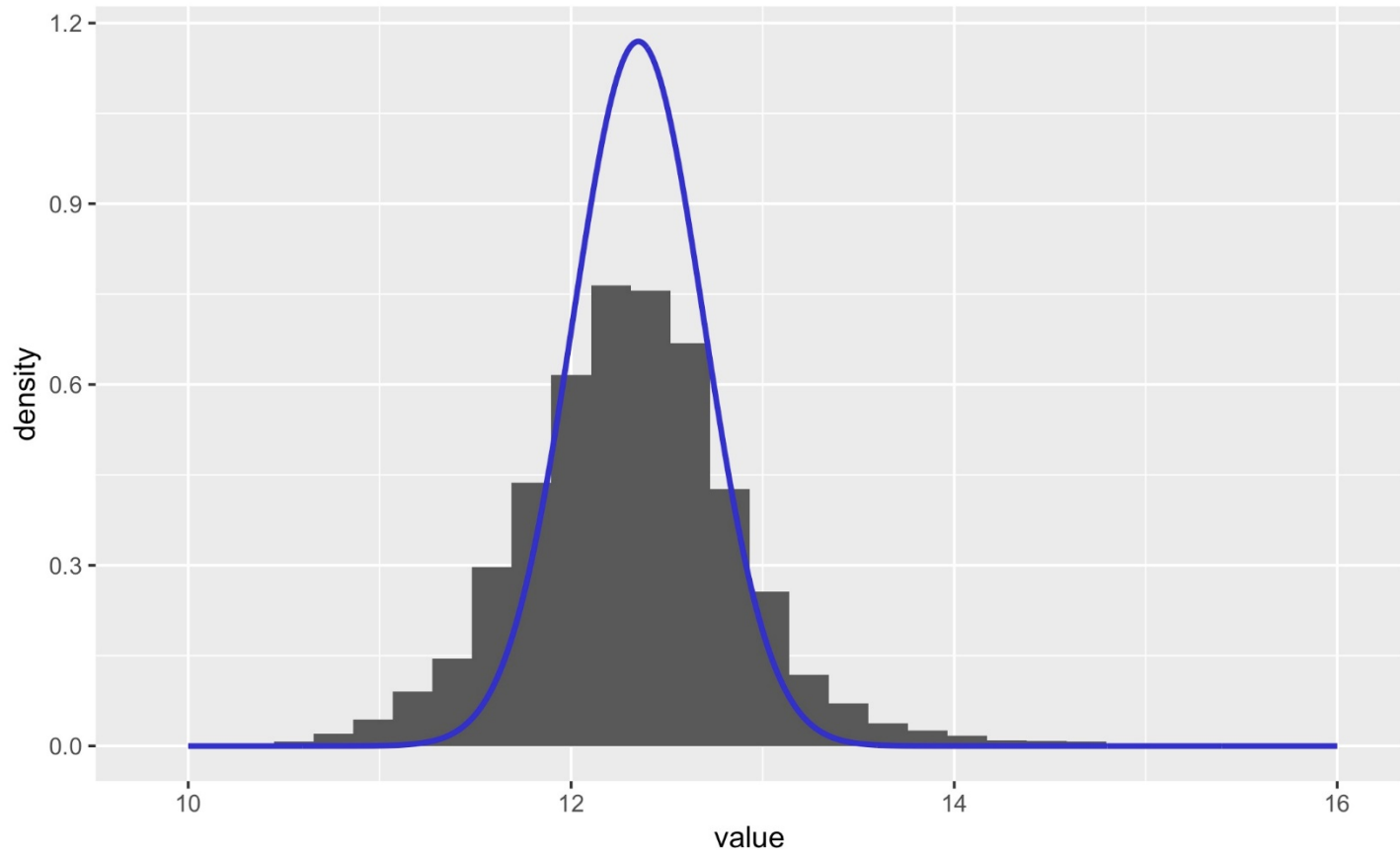
ヒストグラムと対数正規分布の当てはまり

実際の消費支出と推定結果を用いた確率密度関数の比較

ヒストグラム：消費支出

青線：EMアルゴリズムの推定量を使用した正規分布

※図は2014年9月のものを使用



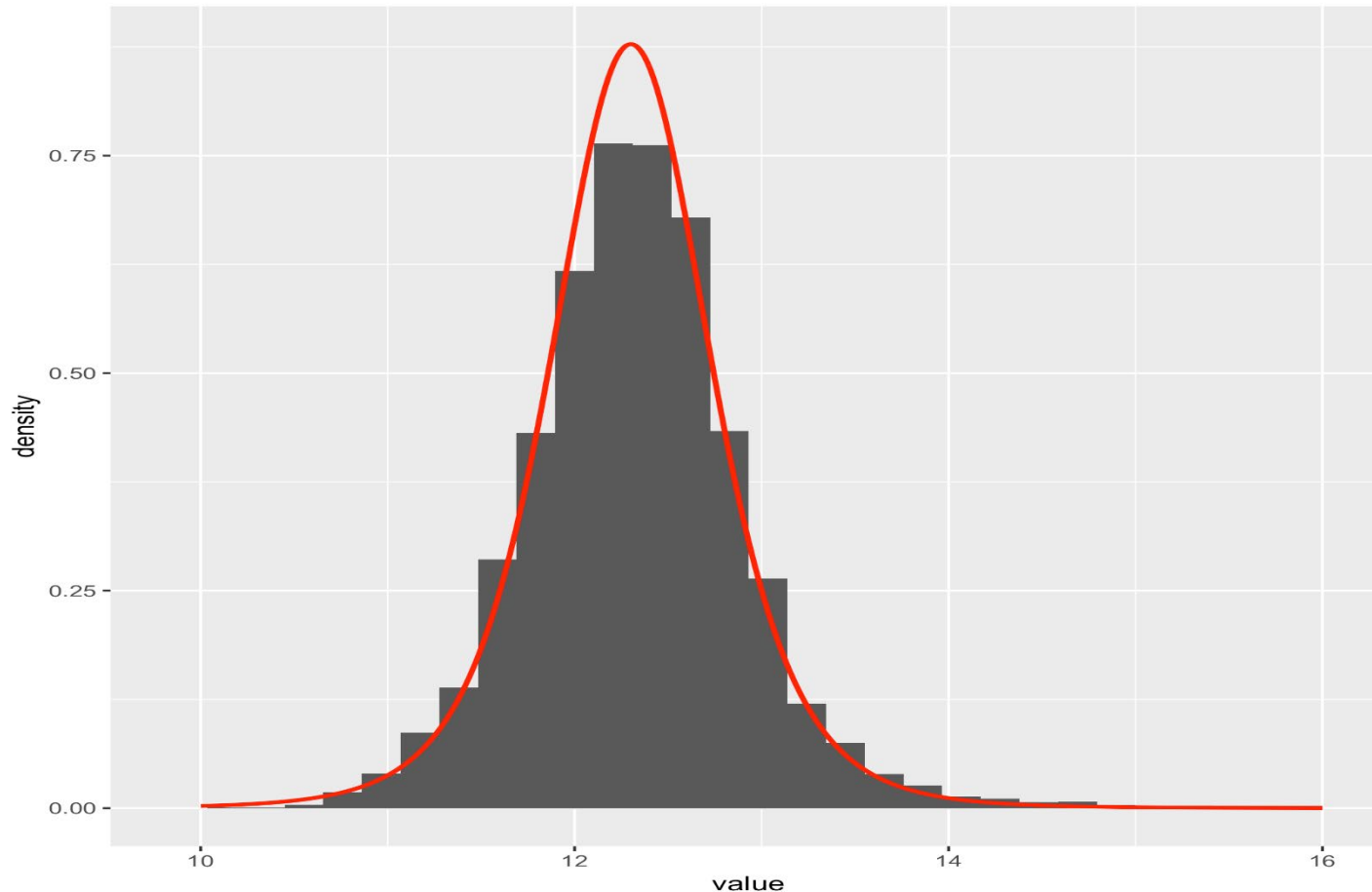
ヒストグラムと対数 t の当てはまり

実際の消費支出と推定結果を用いた確率密度関数の比較

ヒストグラム：消費支出

※図は2014年9月のものを使用

赤線：多変量 t 分布を仮定したもの(経済学における適用例としてBarrosoら2019)



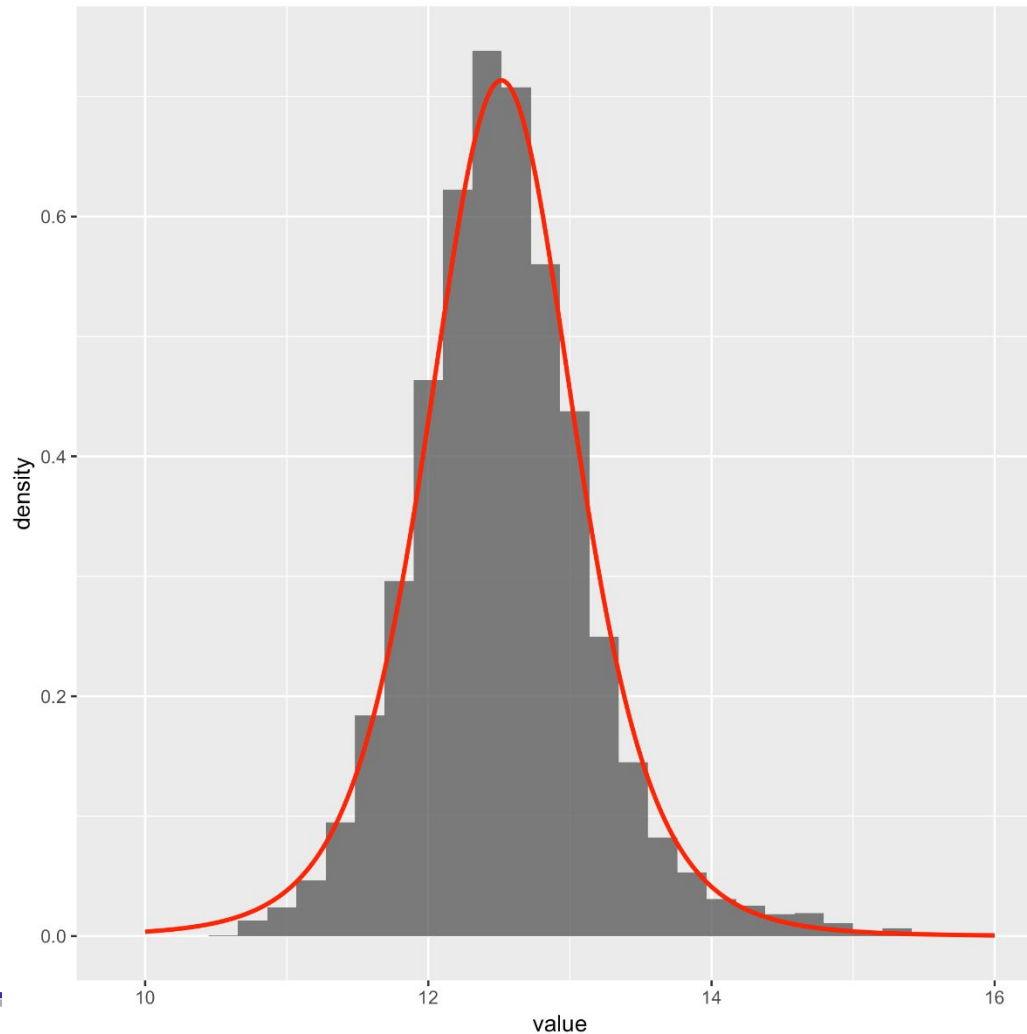
ヒストグラムと対数 t の当てはまり

実際の消費支出と推定結果を用いた確率密度関数の比較

ヒストグラム：消費支出

※図は2014年3月のものを使用

赤線：多変量 t 分布を仮定したもの



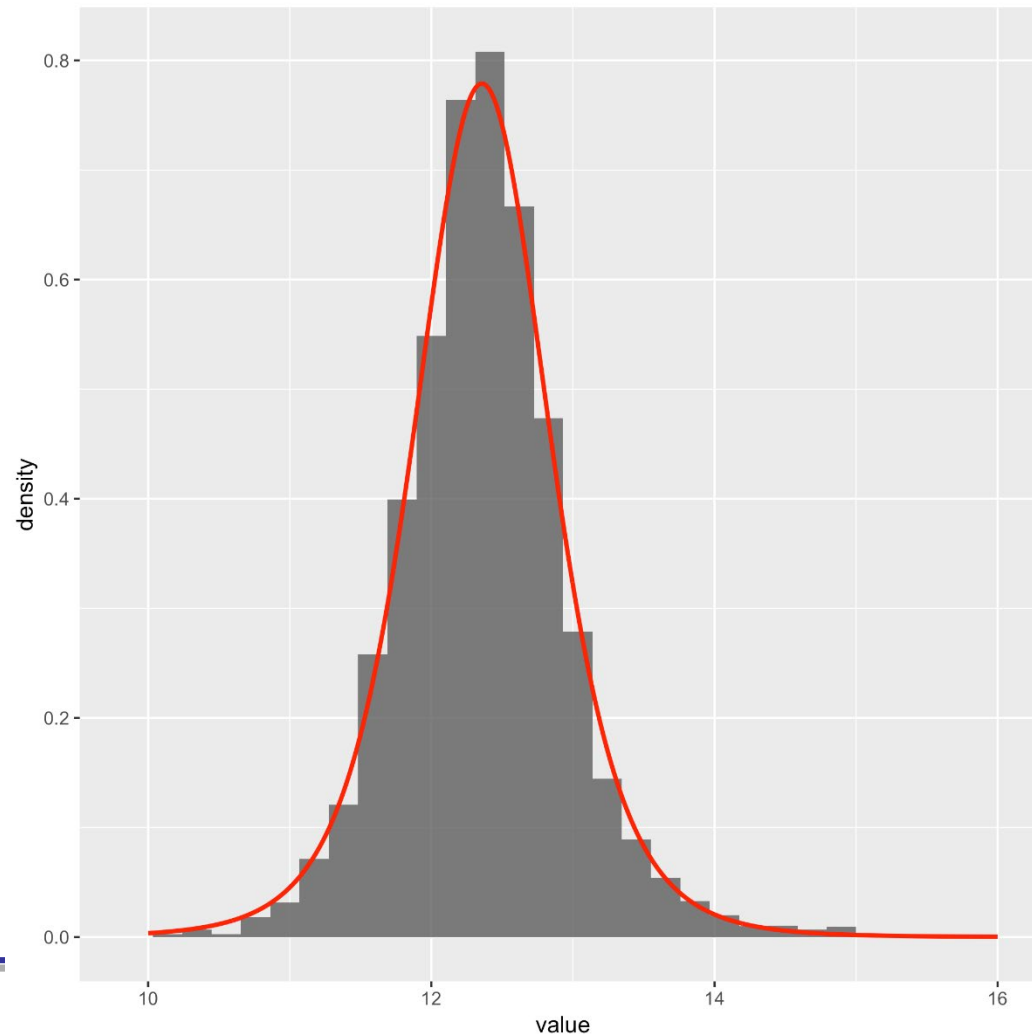
ヒストグラムと対数 t の当てはまり

実際の消費支出と推定結果を用いた確率密度関数の比較

ヒストグラム：消費支出

※図は2014年10月のものを使用

赤線：多変量 t 分布を仮定したもの



初期の計画としては

家計調査では年間で集計値が存在

家計構造では10・11月

⇒「10・11月」と

「それ以外」の2変数での
分析であれば

①家計調査で回帰モデルを
作成し予測値代入

②傾向スコアによる補正
などが考えられる

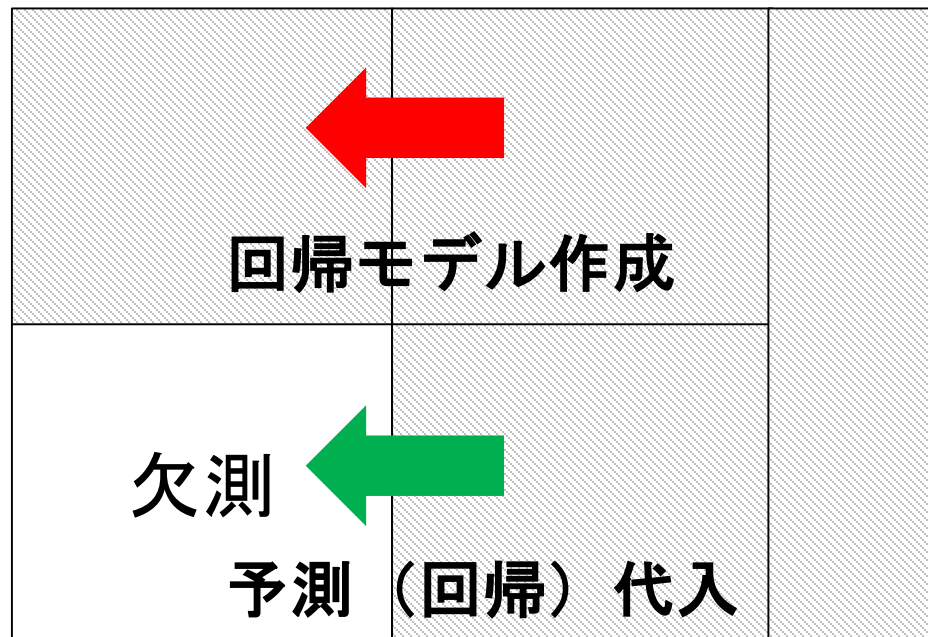
しかしデータの構造が単純ではないので実際にはその方法はとれない
(理由)(1)家計調査がローテーションパネル(2)月ごとの欠測が存在

家計調査
対象者

家計構造
対象者
(今回全消)

10・11月以外 10・11月

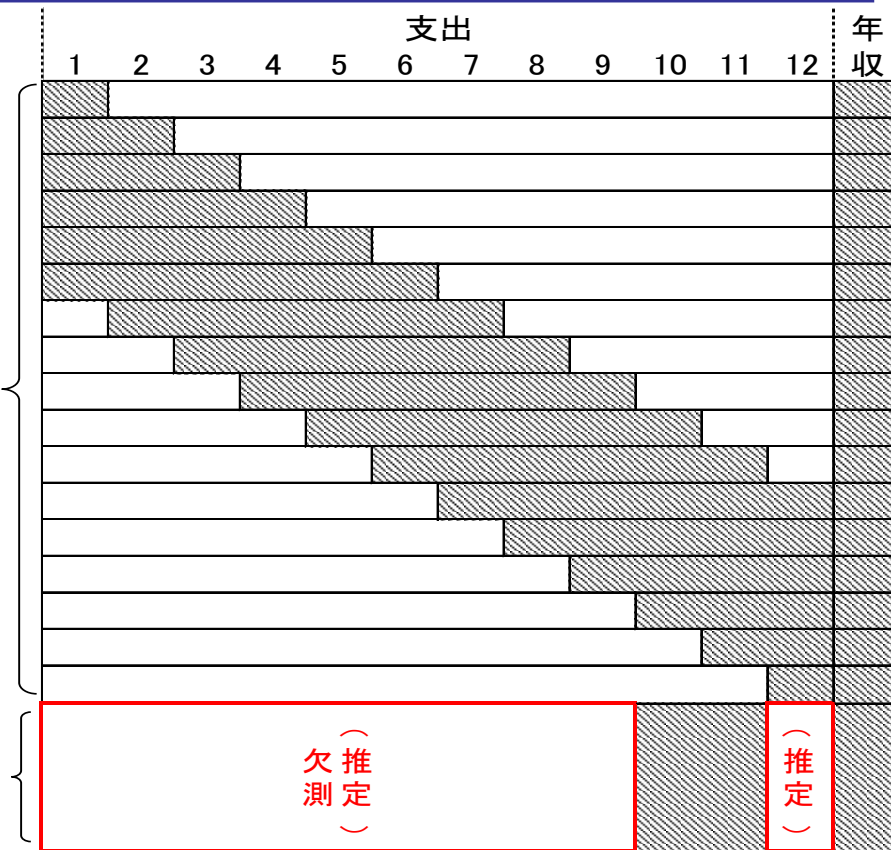
補助
変数



提案手法

ローテーションパネルである
家計調査と全国消費実態調査の
関係を右図のように整理し、
欠測データの構造であること
を考慮し両者を併合させた分析
仮定A)欠測が「ランダムな欠測」
仮定B) 両者は共に代表性が
あるとして月次の平均は共通

家計調査

全国消費
実態調査

上図) 2人以上世帯の場合

* 仮定Bは無くても識別可能

【利点】

- ・ 家計調査単体より精度が向上
- ・ 全消の諸変数と相関算出可能

実際には

家計調査は2人以上世帯は6か月

継続して回答：17パターン

全消を合わせて欠測18パターン

* 家計完全回答は37.2%

家計調査

(17中10パターンは必ず5か月以下)

家計調査で欠測が生じている！

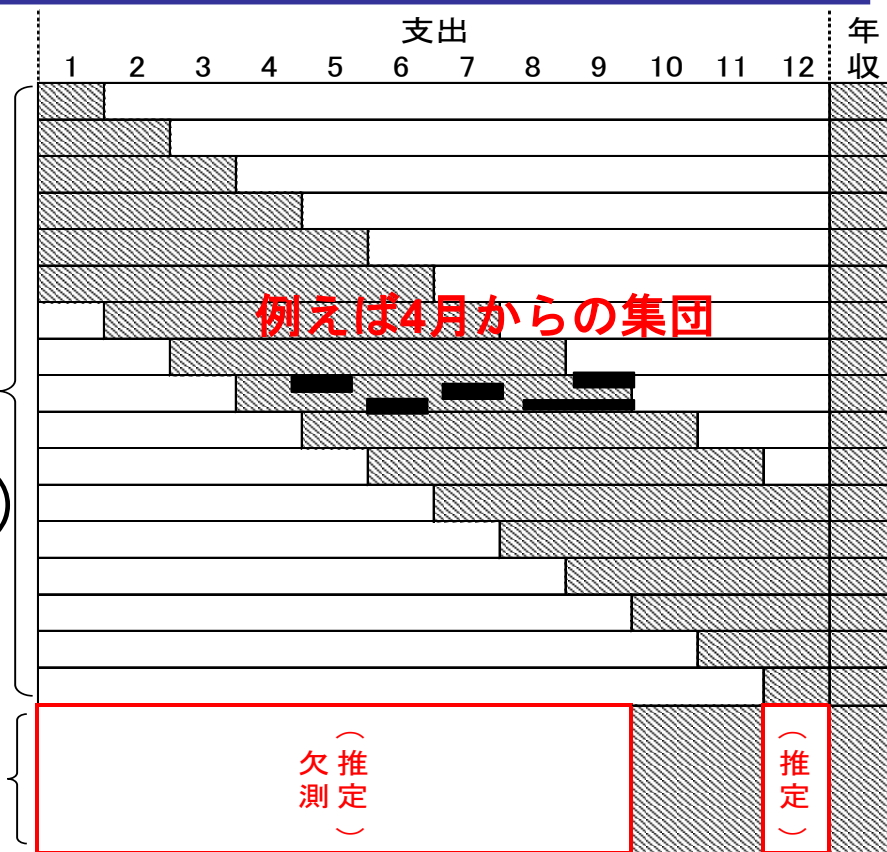
これを考慮してすべてのデータを

全国消費
実態調査

利用すると26パターンの

欠測を考慮した分析を背後で

行っている



上図) 2人以上世帯の場合

消費支出と食費についての情報

今回は以下の情報を利用

【家計調査】 ※家計簿の種類は2種類有、同時記入

①家計簿A（二人以上の世帯用）

I 口座自動振替による支払

II 口座への入金(給与・年金等)[世帯主,世帯主の配偶者,その他(間柄を記入)] ※世帯収入・消費は構成員の収入の合計

III 現金収入又は現金支出

IV クレジット・電子マネーなど現金以外による購入

②家計簿B（二人以上の世帯用）

- ・ 口座自動振替による支払、現金収入又は現金支出
 - ・ クレジットカード,掛買い,月賦による購入又は現物
-

消費支出と食費についての情報

今回は以下の情報を利用

【全国消費実態調査】 ※家計簿の種類は2種類有

①家計簿A（9月、10月分）

②家計簿B（11月分）

I 口座振替による自動支払

II 現物(現物支給, 貰い物・もてなし, 自家産, 自分の店の商品)

III 現金収入又は現金支出

IV クレジットカード, 掛買い, 月賦, 電子マネーによる購入

消費支出：各項目の合計

食費：各項目のうち食料に分類されているものの合計

過去報告時との変更点

過去報告時には

□ 多変量正規分布を仮定した分析を紹介(2019年3月)

⇒家計調査 + 全消の動きに連動し自然な結果を与えた

一方分布の当てはまりに課題

□ 多変量 t 分布を仮定したEMアルゴリズムの結果を報告(19年8月)

⇒分布の当てはまりは良いが挙動がおかしい

□ 多変量 t 分布を仮定した分析の問題点を探索し解決

原因：多変量 t で使ったデータセットが比較手法と異なっていた

(改善点)

- ・ 標準誤差の推定が正しく行えるようになった

- ・ 加えて全国家計構造と同じ10・11月に全消を変更(過去は9・10)

多変量 t 分布の仮定

- 家計調査のデータと全国消費実態調査のデータ(10/11月)を融合したことの効果としてのメディアンの変化を比較

(家計調査単独)

(EMアルゴリズムを用いた全消と家計調査との融合)

以上について、各月のメディアンを、仮定する分布ごとの結果比較

※次回の全国家計構造調査は10/11月のみ実施

- ここでのモデルは家計調査も全国消費実態調査も同じ多変量 t 分布に従うと仮定(自由度 $v: v \rightarrow \infty$ で多変量正規分布)

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_{12} \end{bmatrix} \sim t_{12} \left(\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{12} \end{bmatrix}, \Phi = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1,12} \\ \vdots & \ddots & \vdots \\ \sigma_{1,12} & \cdots & \sigma_{12}^2 \end{bmatrix}, v \right)$$

多変量 t 分布とEMアルゴリズム

Liu and Rubin(1995)よりこの多変量 t 分布に従うYは

$$Y \sim N_{12}(\mu, \Phi/\tau)$$

とし τ がガンマ分布

$$\tau \sim \text{Gamma}(v/2, v/2)$$

に従うと考えるもよい。ここで τ を観測できない潜在変数と考えてEMアルゴリズムを実行すれば

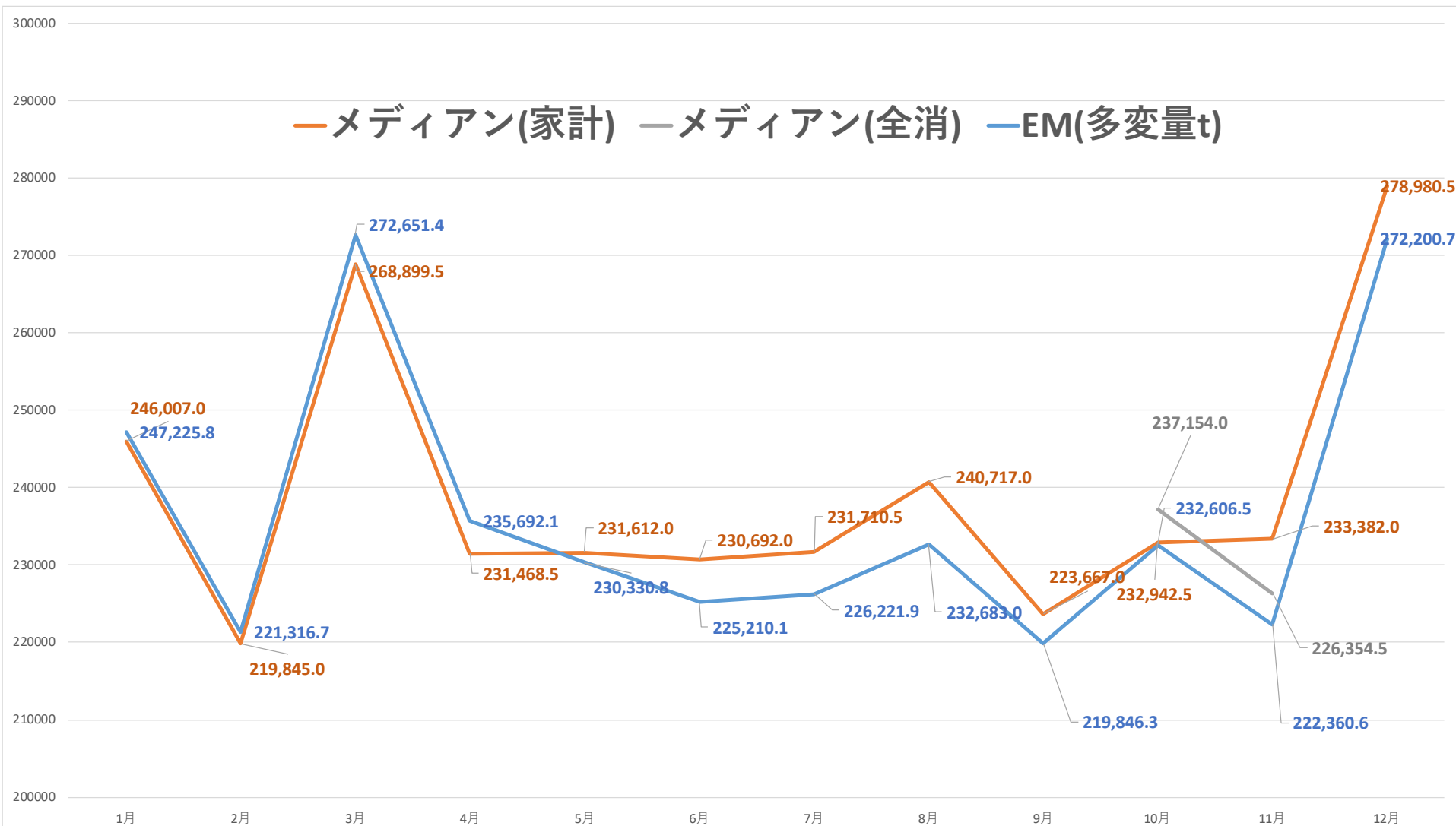
- ① τ が分かっているとした場合にYの中の観測値部分（全消なら9~11月）が得られているときの欠測値部分（9~11以外）の期待値を計算
- ② この期待値を代入して μ と Φ を推定
- ③ τ の期待値を代入して v を推定

のEMアルゴリズムが利用可能

解析結果（消費支出メディアン）

横軸：月，縦軸：消費支出

自由度5.56562

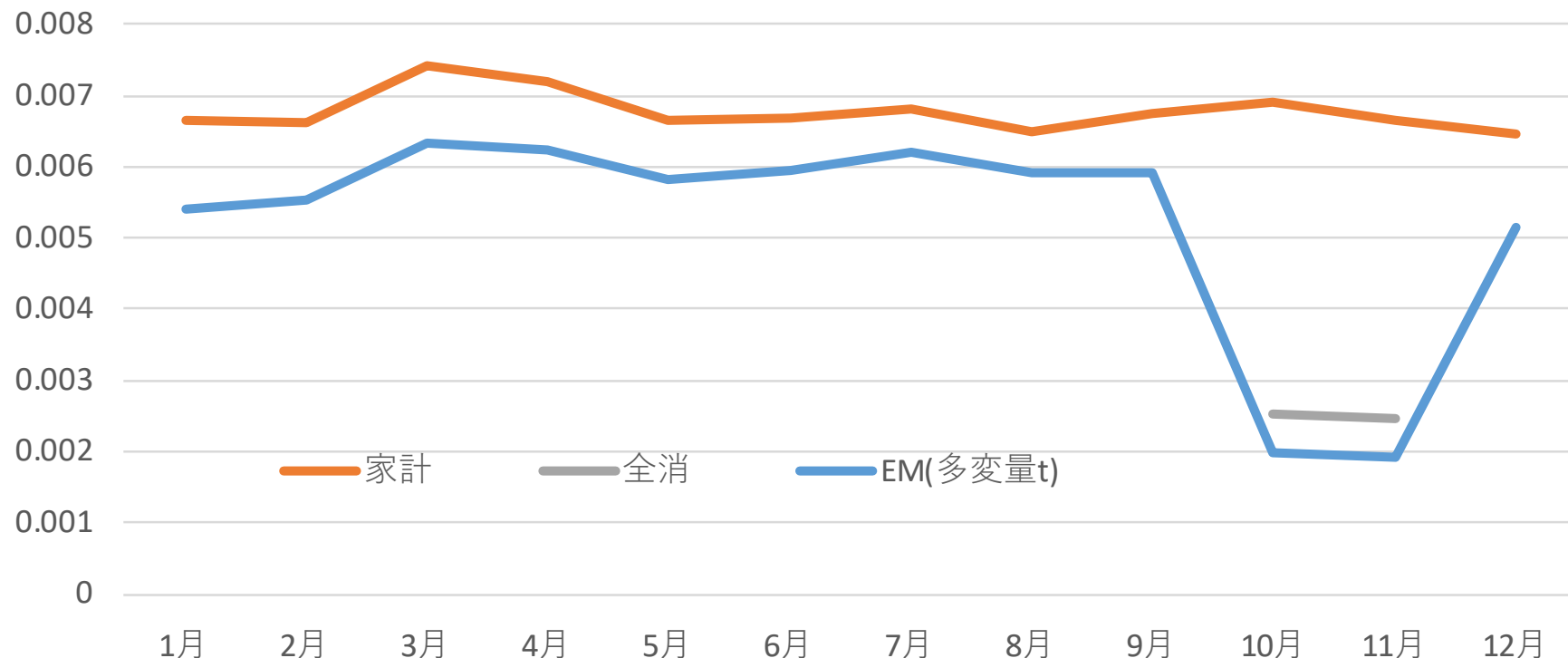


解析結果（消費支出の月次額の標準誤差）

横軸：月，縦軸：標準誤差

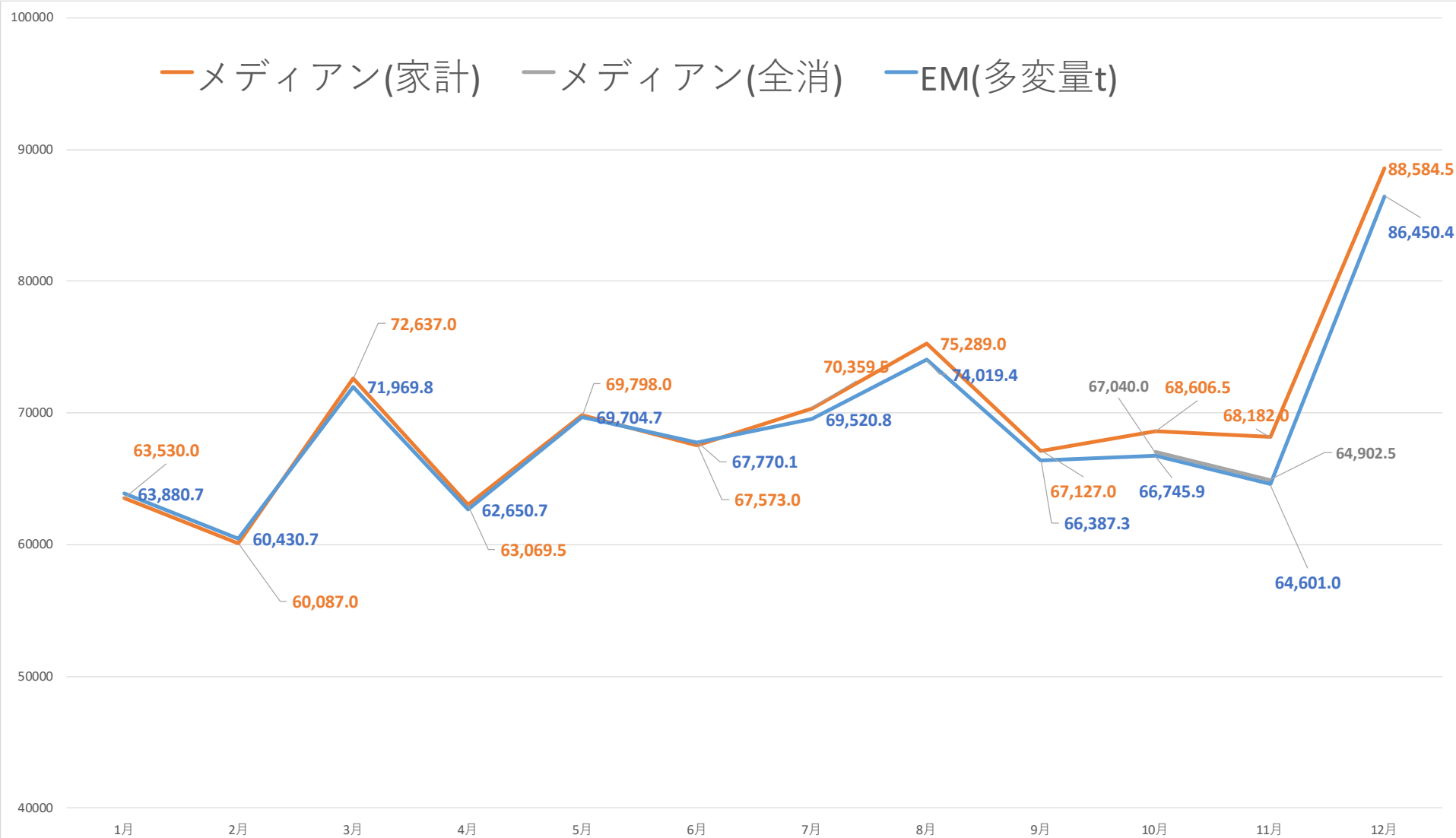
融合により精度は家計調査単体に比べて平均1.55倍（サンプルサイズは2.39倍換算）

10・11月の全消に比べても平均1.26倍(同1.58倍)



解析結果（食費メディアン）

横軸：月，縦軸：食費

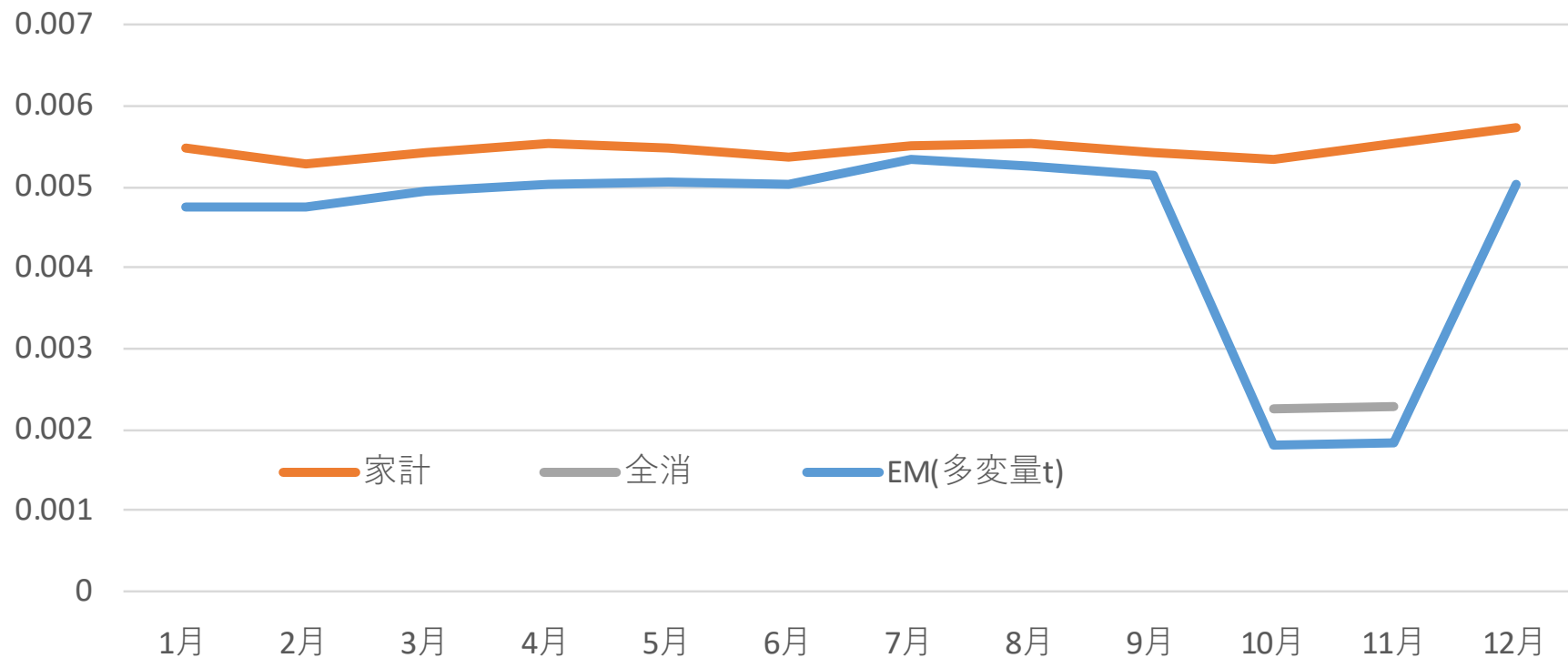


解析結果（食費の月次額の標準誤差）

横軸：月，縦軸：標準誤差

融合により精度は家計調査単体に比べて平均1.41倍（サンプルサイズは1.98倍換算）

10・11月の全消に比べても平均1.28倍(同1.63倍)



年メディアンについて

(A)全消での10~11月のメディアン

(B)家計調査での10~11月のメディアン

(C)家計調査での年メディアン

(D)消費支出に多変量t分布を仮定した場合のEMアルゴリズムを仮定した場合の年メディアン推定

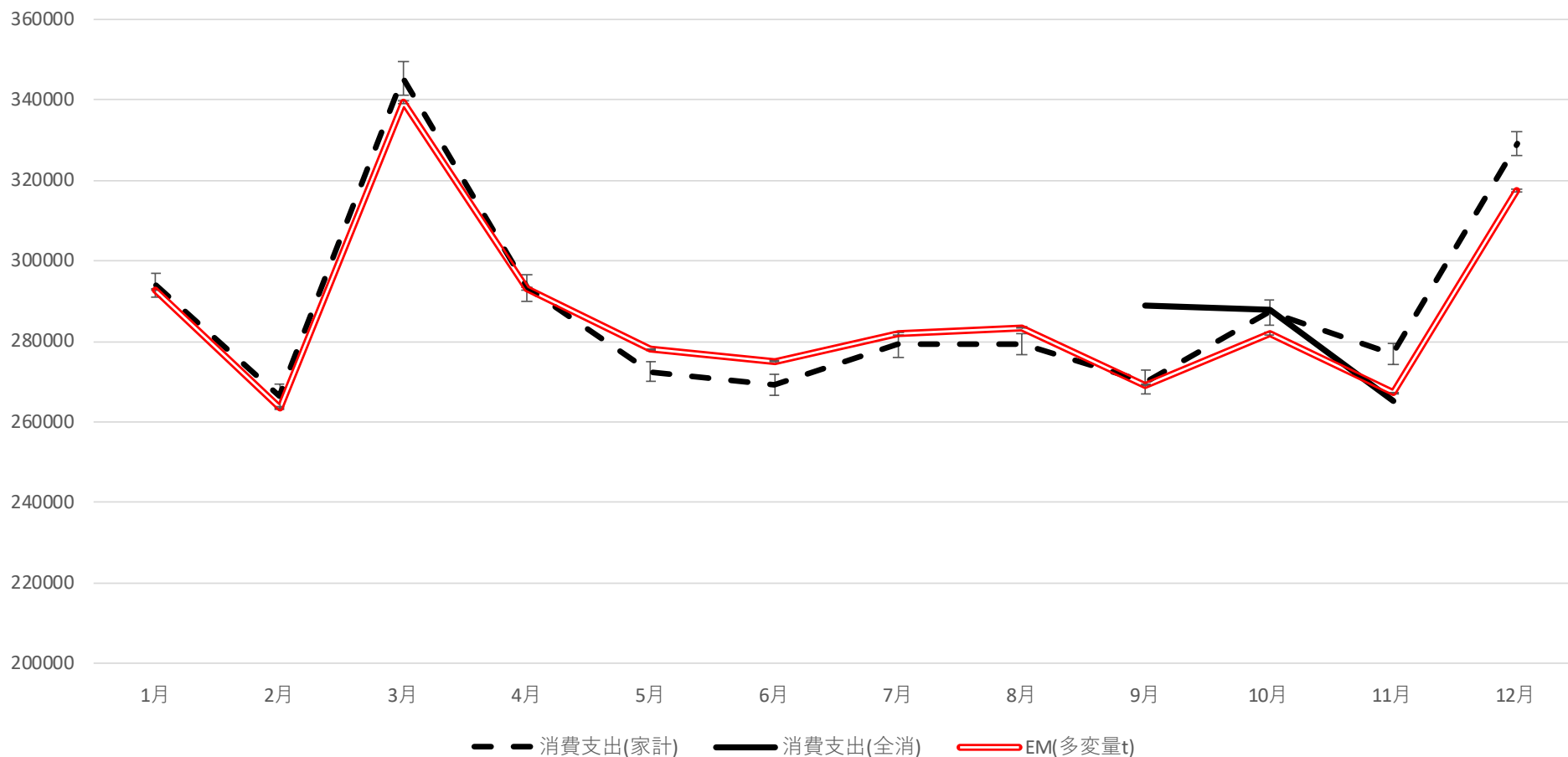
	(A)	(B)	(C)	(D)
消費支出	231754.25	233162.25	239160.2917	236528.8253
食費	65971.25	68394.25	69570.25	68677.61188

解析結果（消費支出平均）

横軸：月，縦軸：消費支出

対数多変量t分布から実数の平均は陽に計算できないため

モンテカルロ平均を算出(100万サンプル)

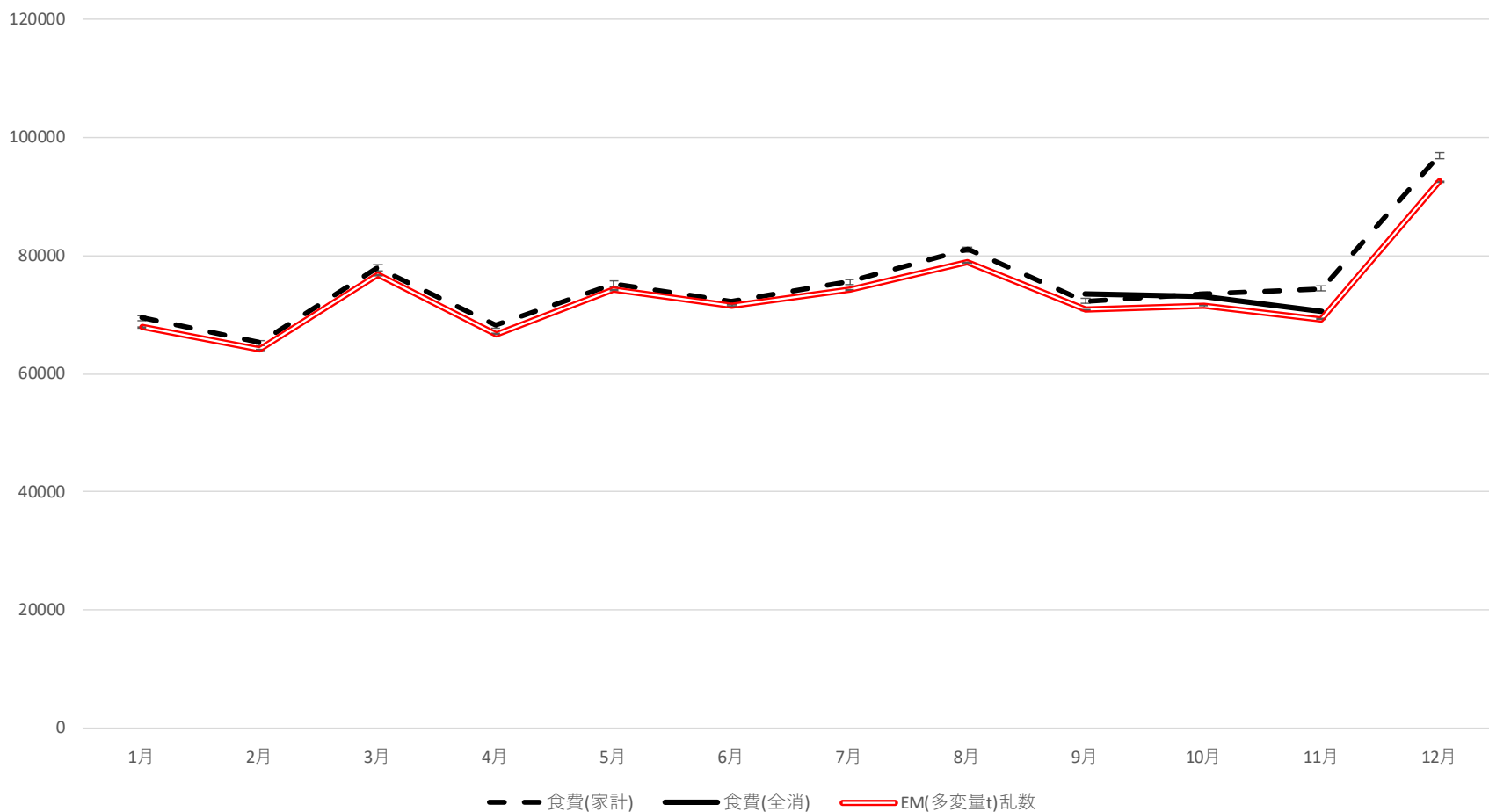


解析結果（食費平均）

横軸：月，縦軸：食費

対数多変量t分布から実数の平均は陽に計算できないため

モンテカルロ平均を算出(100万サンプル)



年平均について

(A)全消での10~11月の平均値

(B)家計調査での10~11月の平均値

(C)家計調査での年平均値

(D)消費支出に多変量t分布を仮定した場合のEMアルゴリズムを仮定した場合の年平均推定

	(A)	(B)	(C)	(D)
消費支出	276502.42	281983.59	288539.2274	286902.6024
食費	71811.522	73932.369	75168.10395	73244.39953

まとめ

【まとめ】

- ・ 全国家計構造調査は10/11月の実施だが消費の季節変動が大
⇒季節変動を適切に考慮せずに年間の消費額や食費とすると様々な政策的決定を誤る危険性
 - ・ 集計時系列上での季節調整が難しいのでミクロレベルでの調整
 - ・ 家計調査では年間で集計値が存在／家計構造では10・11月
但し「10・11月」と「それ以外」の2変数での分析ではできない
⇒月次データに対する多変量の欠測構造を考慮した分析が必要
対数をとっても非正規性が確認されたので多変量t分布によるEMアルゴリズムを用いた推定
 - ・ 融合により集計値の変動(標準誤差)を小さくすることができた
⇒当初の目的を実現するだけでなく融合した家計調査にもメリット
 - ・ 過去うまくいかなかった原因が分かった(データの問題)
-

今後の課題と検討事項

- ・ 単身世帯でのプログラム作成と2人以上での統合

方法論的には容易だがデータハンドリングの問題

⇒ 単身と2人以上は別々に実施して集計で融合することは可能

- ・ 代入の検討をしていたが、データ構造上ほぼ難しいと

考えるに至る

- ・ 調査疲れ等の影響(調査開始月からの低下)を理解した上でこれを

考慮した集計を行うかどうか？が課題では？

平均して徐々に低下していく構造を考慮することは可能

- ・ 十大費目の他費目の実施は容易だがデータの整形が課題

- ・ 共変量を用いた推定精度の向上

⇒ 多変量tのEMアルゴリズムではまだ試験中

参考文献

- Bassoro, F.J.C., García-Pérez, C. and Prieto-Alaiz, M. (2019) “Modelling income distribution using the log Student’s t distribution: New evidence for European Union countries”, *Economic Modelling*, in press.
- Kim, JK. and Shao, J. (2014) Statistical Methods for Handling Incomplete Data. Chapman & Hall.
- Little, R.A.J and Rubin, D.B. (2019) Statistical Analysis with Missing Data, 3rd ed. Wiley.
- Liu, C. and Rubin, D.B. (1995) “ML Estimation Of The t-Distribution Using EM and Its Extensions, ECM and ECME”, *Statistica Sinica*, 5, 19-39.
- Ridder, G. & Moffitt, R. (2007) The Econometrics of Data Combination, in J. J. Heckman & E. E. Leamer, eds, ‘Handbook of Econometrics’, Vol. 6, Elsevier, chapter 75, pp.5469-5547.
- Schafer, J.L. (1997) Analysis of Incomplete Multivariate Data, Chapman & Hall.
- 星野崇宏(2009)「調査観察データの統計科学」岩波書店
- 高井啓二・星野崇宏・野間久史 (2016)「欠測データの統計科学」岩波書店
-