



総消費動向指数（CTIマクロ）作成における 民間データ等の利活用に係る取組の紹介

令和 8 年 7 月 27 日

（独）統計センター 技術研究開発課

1. 主な研究課題
2. 民間データを単独で用いたサービス産業売上高計の予測
3. ベイジアンモデル平均化法（BMA）の概略
4. BMAによるサービス産業売上高計の予測と重要度が高い変数に特殊な変動があった場合のシミュレーション
5. まとめと課題

1. 主な研究課題

名目値

◆世帯消費動向指数（CTIミクロ）（総務省統計局）

- ・消費支出の指数値（総世帯，原数値，名目）

◆商業動態統計（経済産業省）

- ・販売額（小売業，原数値）

◆サービス産業動態統計調査※（総務省統計局）

- ・売上高（サービス産業計）

実質値

◆世帯消費動向指数（CTIミクロ）（総務省統計局）

- ・消費支出の指数値（総世帯，原数値，実質）

◆第3次産業活動指数（経済産業省）販売額

- ・広義対個人サービス（原指数）

◆鉱工業生産指数（経済産業省）

- ・消費財（原指数）

CTIマクロの課題

- ◆CTIマクロ推定時に、説明変数に利用するサービス産業動態統計調査が、公表日の都合で最新月の結果を利用できない。
- ◆現状では、未公表の説明変数は状態空間モデルとカルマンフィルタを用い、過去の公表値から時系列回帰モデルを用いて一期先予測する形で算出している。
- ◆そのため、コロナ等の経済ショックを直近の指数に十分に反映できない恐れがある。
- ◆早期に入手できる民間データ等を使い、公表が遅い公的統計を予測できないか検討。

CTI推定月のデータ状況イメージ図
(数値は架空)

	サービス 産業動態	商業動 態	民間 データ等
n-3月	340	500	672
n-2月	352	520	701
n-1月	382	541	720
n月	未公表	495	650

- ◆未公表分のサービス産業動態統計調査を、民間データ等を使い予測。

本研究の目的

- ◆総務省「サービス産業動態統計調査」の「サービス産業売上高計」の系列（以下、「サービス産業売上高計」という）を、民間データ等を用いて予測.
- ◆民間データ等に特殊な変動（ポイント増加キャンペーンによる一時的な売上増など）が生じた場合でも、個々の説明変数の変動を受けにくい予測方法として「モデル平均化法」の利用を検討.

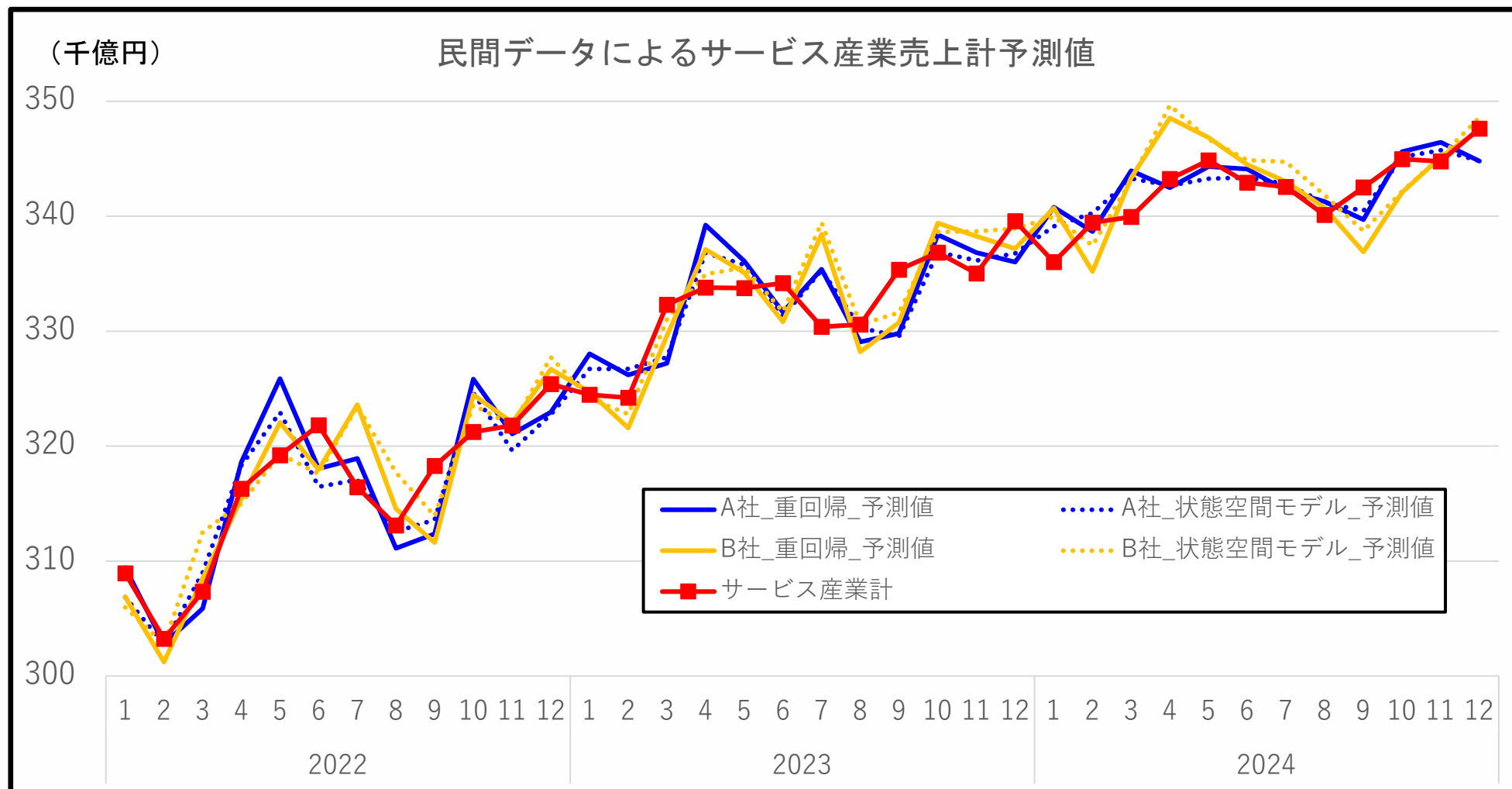
サービス産業動態統計調査の概要

- ◆我が国におけるサービス産業の事業活動を明らかにする月次調査.
- ◆2025年から、それまでの「サービス産業動向調査（総務省統計局）」と「特定サービス産業動態統計調査（経済産業省）」を統合する形で調査を実施.

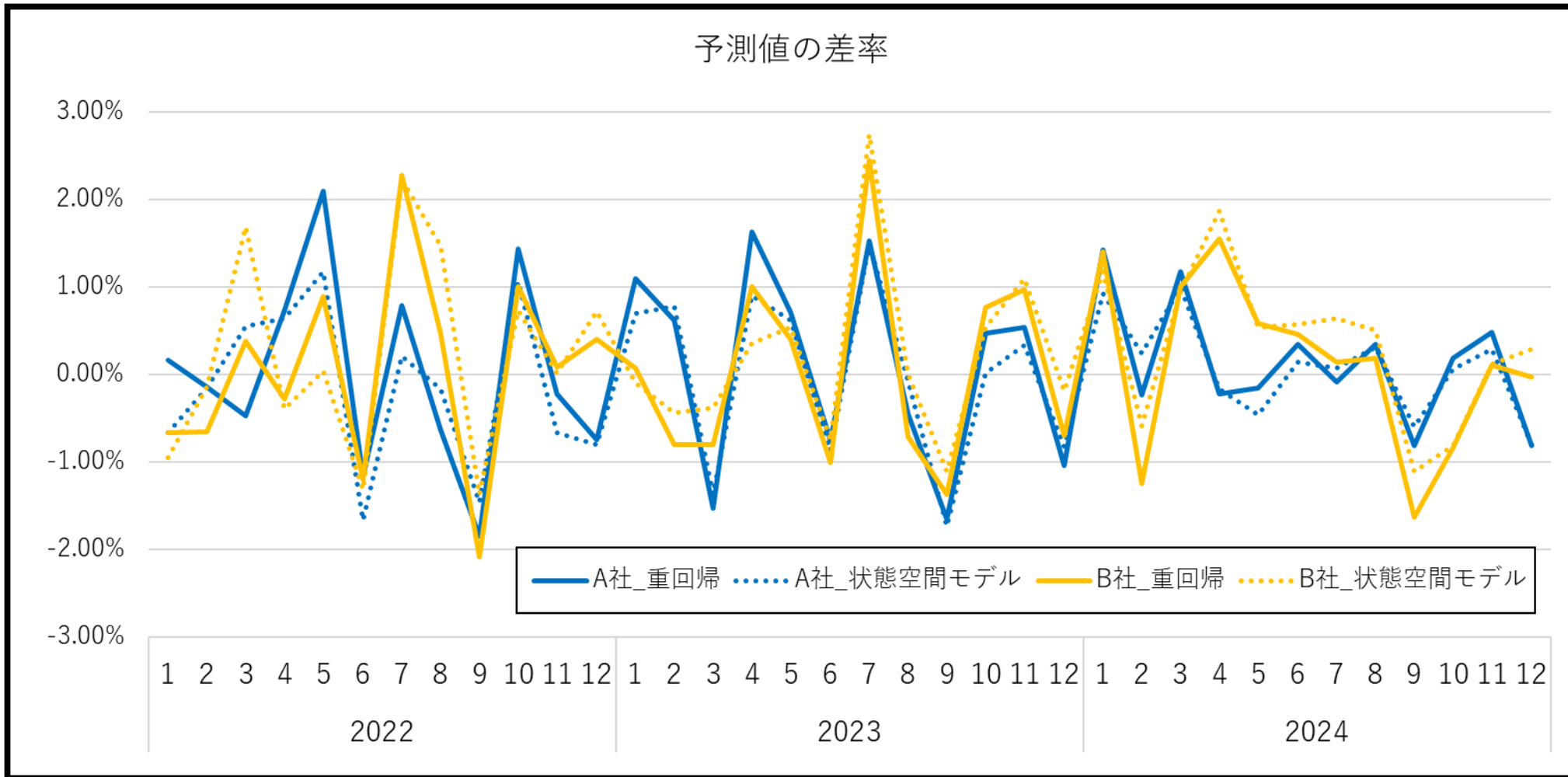
2. 民間データを単独で用いた サービス産業売上高計の予測

単独の民間データを用いた予測

- 民間データ（A社・B社）を単独で使い、重回帰分析や状態空間モデル等を利用しサービス産業売上計を予測.



単独の民間データを用いた予測結果_（差率）



- それぞれのデータを単独で用いても、一定の精度での予測は可能.
- 複数の民間データ等を用いた予測方法の研究に着手.

3. ベイジアンモデル平均化法 (BMA) の概略

- 例として消費（C）を，所得（Y），保有資産（S），世帯人員（Z）の3つの変数から推定することを考える．

一般的な推定方法

- ① 経済モデル等で利用されている関数（例：ケインズ型消費関数）を使って推定する方法．

$$\hat{C} = \hat{\beta}_0 + \hat{\beta}_1 Y$$

- ② 複数の説明変数を用いて重回帰分析を行う方法．

$$\hat{C} = \hat{\beta}_0 + \hat{\beta}_1 Y + \hat{\beta}_2 S + \hat{\beta}_3 Z$$

さらにAIC等を用いて1つのモデルを選択し，そのモデルに基づいて予測を行う方法．

$$\hat{C} = \hat{\beta}_0 + \hat{\beta}_1 Y + \hat{\beta}_2 S$$

モデル平均化法

- 複数の候補モデルを用意し，各モデルから得られる予測値にウェイトを付け，それらを加重平均して最終的な予測値を求める方法．

$$\hat{C}_1 = f_1(Y), \quad \hat{C}_2 = f_2(S), \quad \hat{C}_3 = f_3(Z),$$

$$\hat{C}_4 = f_4(Y, S), \quad \hat{C}_5 = f_5(Y, Z),$$

$$\hat{C}_6 = f_6(S, Z), \quad \hat{C}_7 = f_7(Y, S, Z),$$

$$\hat{C}_8 = f_8 \text{ (説明変数を含まないモデルの予測値)}$$

各モデルのウェイトを w_i とすると，モデル平均加法による予測値は

$$\hat{C}_M = \sum_{i=1}^8 w_i \hat{C}_i, \quad \sum_{i=1}^8 w_i = 1$$

で与えられる．なお，必要に応じてウェイトの大きい，いくつかの上位のモデルのみを用いる場合もある．

複数のデータ提供元の利用

- A社およびB社のデータを1つの系列に統合することなく、個別の説明変数として候補モデルに取り入れ、両社のデータを含むモデルと一方のみを含むモデルを合わせて評価することが可能。

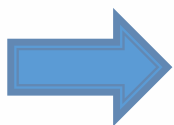
モデルの不確実性への対応

- 1つのモデルに予測を依存させることなく、複数の候補モデルの不確実性を予測に反映することが可能。

特殊な変動が生じた説明変数への対応

- 特殊な変動が生じた説明変数を候補から除外し、残された説明変数からなる候補モデルについて再推定することで、予測を継続することが可能。

世帯人員(Z)に特殊な
変動が生じた場合



Zを説明変数の候補から除外し、残る4つの候補モデルについてウェイトを再計算して予測する

$$\hat{C}_1 = f_1(Y), \hat{C}_2 = f_2(S), \hat{C}_4 = f_4(Y, S), \hat{C}_8 = f_8,$$

$$\hat{C}_M = w_1 \hat{C}_1 + w_2 \hat{C}_2 + w_4 \hat{C}_4 + w_8 \hat{C}_8,$$

$$w_1 + w_2 + w_4 + w_8 = 1$$

複数の候補モデルを予測に反映できるほか、特定の説明変数を除外した場合にも予測を再構成できる。

- ◆ 説明変数の候補が k 個あるとき，説明変数の組合せに応じて候補モデルを考える．切片のみのモデルを含めると，候補モデルの総数は 2^k 個となる．

$$M_1 : Y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$$

$$M_2 : Y_i = \beta_0 + \beta_2 x_{i2} + \varepsilon_i$$

$$M_3 : Y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \varepsilon_i$$

$$\vdots$$

$$M_{full} : Y_i = \beta_0 + \sum_{\ell=1}^k \beta_{\ell} x_{i\ell} + \varepsilon_i, \quad i = 1, \dots, n$$

ただし，各式の β_{ℓ} はそれぞれのモデルにおける回帰係数を表す．また，

$\mathcal{M} = \{M_1, \dots, M_{2^k}\}$ ：説明変数の組合せから構成される候補モデルの集合

$\mathbf{y} = (y_1, \dots, y_n)'$ ：目的変数の観測値ベクトル（「'」は転置を表す）

- ◆ BMAでは，各モデルのパラメータに対する事前分布と，各モデルに対する事前確率を設定する．
- ◆ 本報告では，回帰係数の事前分布にZellnerのg事前分布を，各モデルの事前確率 $P(M_i)$ にベータ二項分布を用いる．
- ◆ MCMCを用いてモデル空間を探索し，事後モデル確率を近似する．

○ PMP (posterior model probability)

観測データのもとで、各候補モデルがどの程度支持されるかを表す指標。候補モデル M_j におけるPMPは次式で定義され、BMAにおける各候補モデルのウェイト w_j とする。

$$\text{PMP}_j = P(M_j|\mathbf{y}) = \frac{p(\mathbf{y}|M_j)P(M_j)}{\sum_{m=1}^{2^k} p(\mathbf{y}|M_m)P(M_m)}$$

ただし、 $p(\mathbf{y}|M_j)$ はモデル M_j の周辺尤度、 $P(M_i)$ はモデル M_j の事前確率。

○ PIP (posterior inclusion probabilities)

各説明変数がモデルに含まれる事後確率を表す。説明変数 x_ℓ におけるPIPは次式で定義される。

$$\text{PIP}(x_\ell) = \sum_{j=1}^{2^k} \delta_{\ell j} \text{PMP}_j, \quad \delta_{\ell j} = \begin{cases} 1 & (x_\ell \text{がモデル } M_j \text{に含まれる}) \\ 0 & (x_\ell \text{がモデル } M_j \text{に含まれない}) \end{cases}$$

○ BMAによる予測値

候補モデル M_j による予測値を $\hat{y}^{(j)}$ とすると、BMAによる予測値 $\hat{y}$ は次式で求めることができる。

$$\hat{y} = \sum_{j=1}^{2^k} w_j \hat{y}^{(j)} = \sum_{j=1}^{2^k} \text{PMP}_j \hat{y}^{(j)}$$

○ GDP の予測

- Fernández, Ley and Steel (2001)
 - 説明変数が全て揃う72か国のデータを用い、41の説明変数を使って成長回帰分析をBMAを用いて行った。最も事後確率が高いモデルでもその確率は1.24%で、多くのモデルに事後確率が分散されていた。
- Ley and Steel(2009)
 - 上記Fernández, et al.(2001)等のデータを用い、BMAで推定する際のモデルの事前確率を変え同じく成長回帰分析を行っている。重要とされる説明変数が使う事前分布によって変わり得ることが示された。
- Bencivelli, Marcellino and Moretti (2017)
 - BMAブリッジモデルを用いてユーロ圏、ドイツ、フランス、イタリアのGDPを予測。ドイツ、フランス、イタリアの予測において、BMAブリッジモデルに基づく予測は、標準的なブリッジモデルよりも予測誤差が小さいとしている。

○ インフレ率の予測

- Koop and Korobilis (2012)
 - Dynamic Model Averaging (DMA)という手法を使い、米国のインフレ率を予測。TVPモデル等よりも精度の高い予測結果を得られたと報告。

4. BMAによるサービス産業売上高計の予測と 重要度が高い変数に特殊な変動があった場合の シミュレーション

- ◆ 説明変数を変えた以下のパターン①, ②について, BMAによるサービス産業売上高計の予測をそれぞれ行う.
- ◆ 比較のため, 説明変数の数を増やしたパターン②については, 重回帰分析 (AICを使用) による予測も行う.

パターン① (5変数)

- 日経平均月末終値 (1期ラグ)
- 新車登録台数※¹ (普通車 + 小型車)
- サービス産業売上高計の1期ラグ
- A社総額データ
- B社総額データ

パターン② (11変数)

- 日経平均月末終値 (1期ラグ)
- 新車登録台数※¹ (普通車 + 小型車)
- サービス産業売上高計の1期ラグ
- A社分類別データ (7分類) ※²
- B社総額データ

※¹: データ出所: 国土交通省「自動車保有車両数」

※²: サービス産業売上高計の内訳の代替案として「宿泊業」, 「飲食業」, 「旅行・レジャー業」, 「交通」, 「医療・介護」, 「通信」, 「その他」の7つを選定.

データの前処理

- ◆各月次系列にDECOMPを適用し、季節成分とノイズ成分を除去する。
- ◆非定常性による見せかけの回帰を回避するため、全ての変数を対数差分変換する。

推定に用いる候補モデル

$$M_1 : \Delta \log y_t = \beta_0 + \beta_1 \Delta \log x_{t1} + \varepsilon_t$$

$$M_2 : \Delta \log y_t = \beta_0 + \beta_1 \Delta \log x_{t1} + \beta_3 \Delta \log x_{t3} + \varepsilon_t$$

$$\vdots$$

$$M_{full} : \Delta \log y_t = \beta_0 + \sum_{\ell=1}^k \beta_{\ell} \Delta \log x_{t\ell} + \varepsilon_t$$

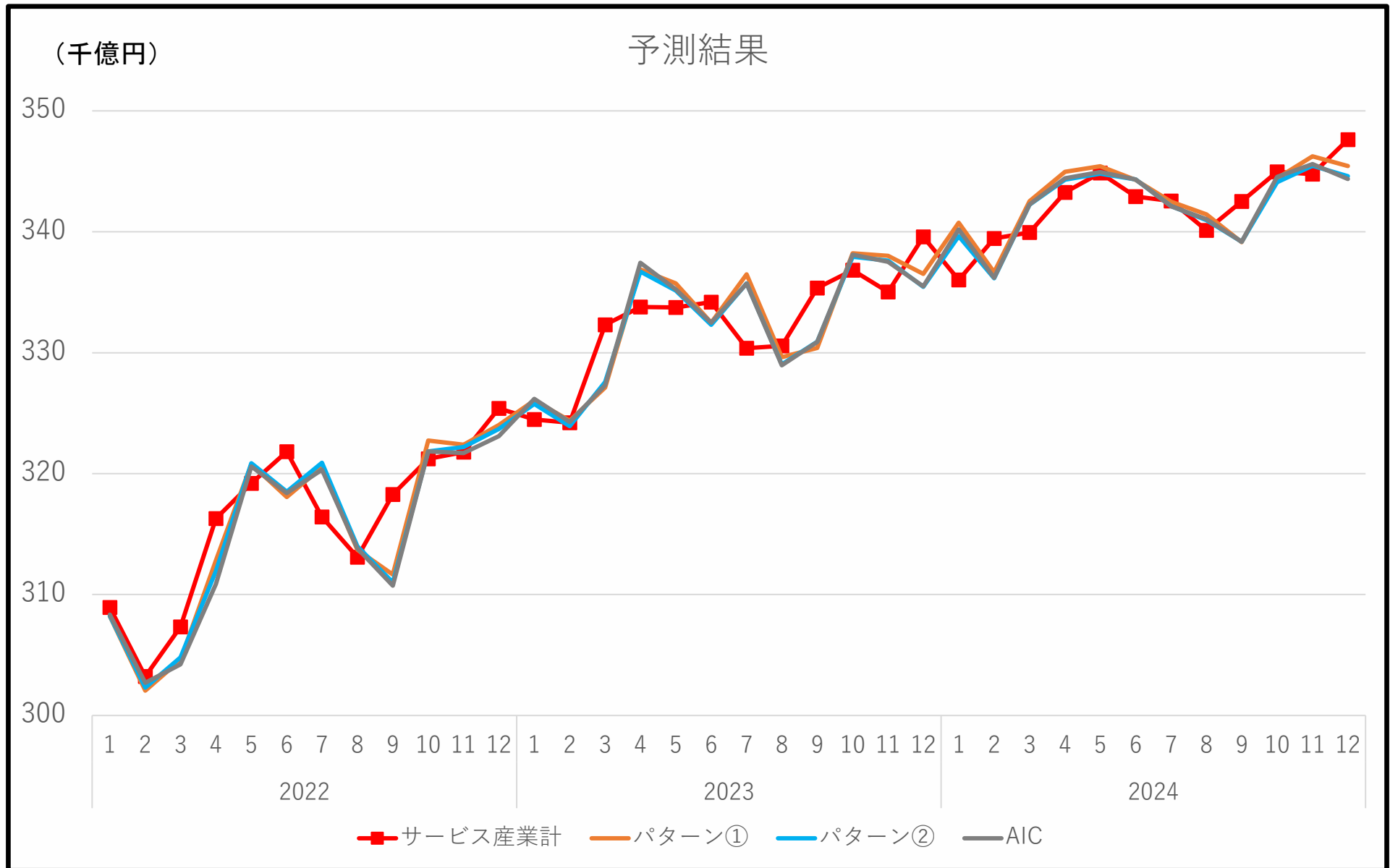
ただし、 $\Delta \log y_t$ は y_t の対数階差とする。

利用する事前分布等

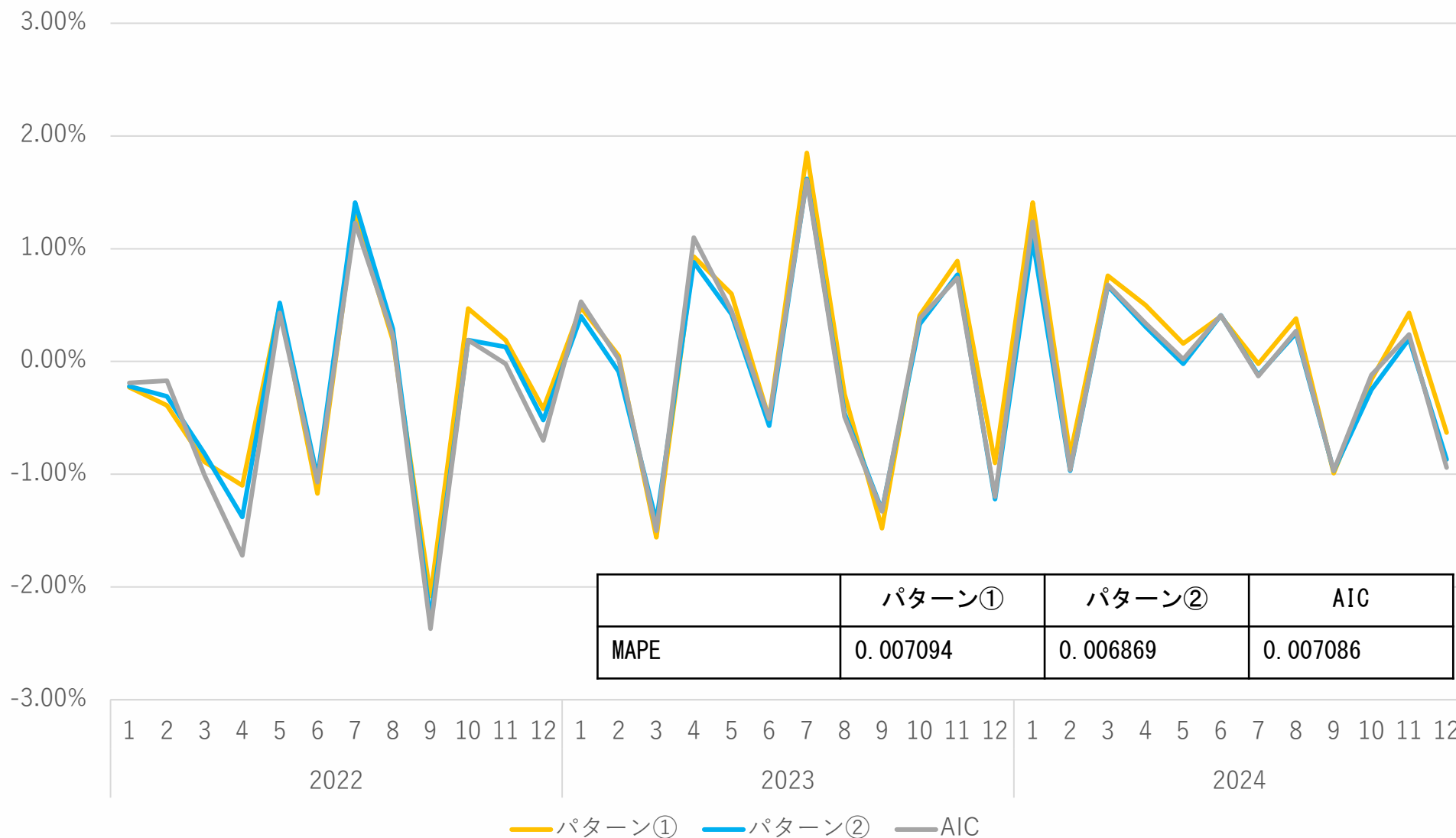
- ◆回帰係数の事前分布：Zellnerのg事前分布
- ◆各モデルの事前確率 $P(M_j)$ ：ベータ二項分布
- ◆MCMCは4000回実行（ただし最初1000回はburn-in分）。

予測値の評価

- ◆差率やMAPE（平均絶対パーセント誤差）を用いて評価。



予測値の差率



予測精度

- ◆ どのパターンも差率でみると、概ね $\pm 2\%$ 程度の誤差で予測が出来ていた。
- ◆ MAPEを計算すると、パターン①：0.007094，パターン②：0.006869，AIC：0.007086であった。

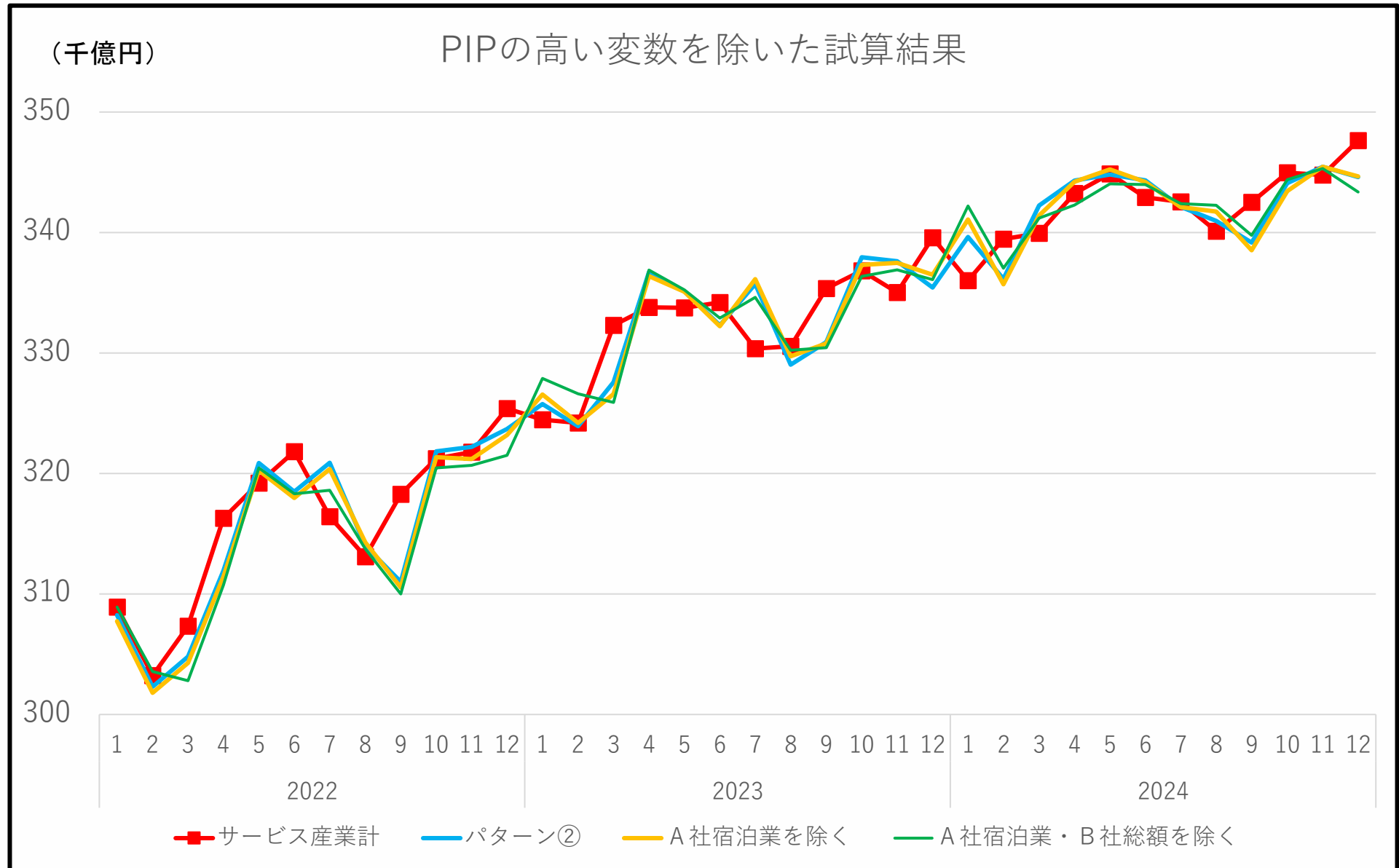
重要度の高い（PIPの高い）変数

- | | |
|---|---|
| <ul style="list-style-type: none">◆ パターン①<ul style="list-style-type: none">• A社総額データ• B社総額データ• サービス産業売上高計の一期ラグ• 日経平均 | <ul style="list-style-type: none">◆ パターン②<ul style="list-style-type: none">• A社宿泊業データ• A社医療介護データ• B社総額データ• サービス産業売上高計の一期ラグ |
|---|---|

データに特殊な変動が生じた場合のシミュレーション

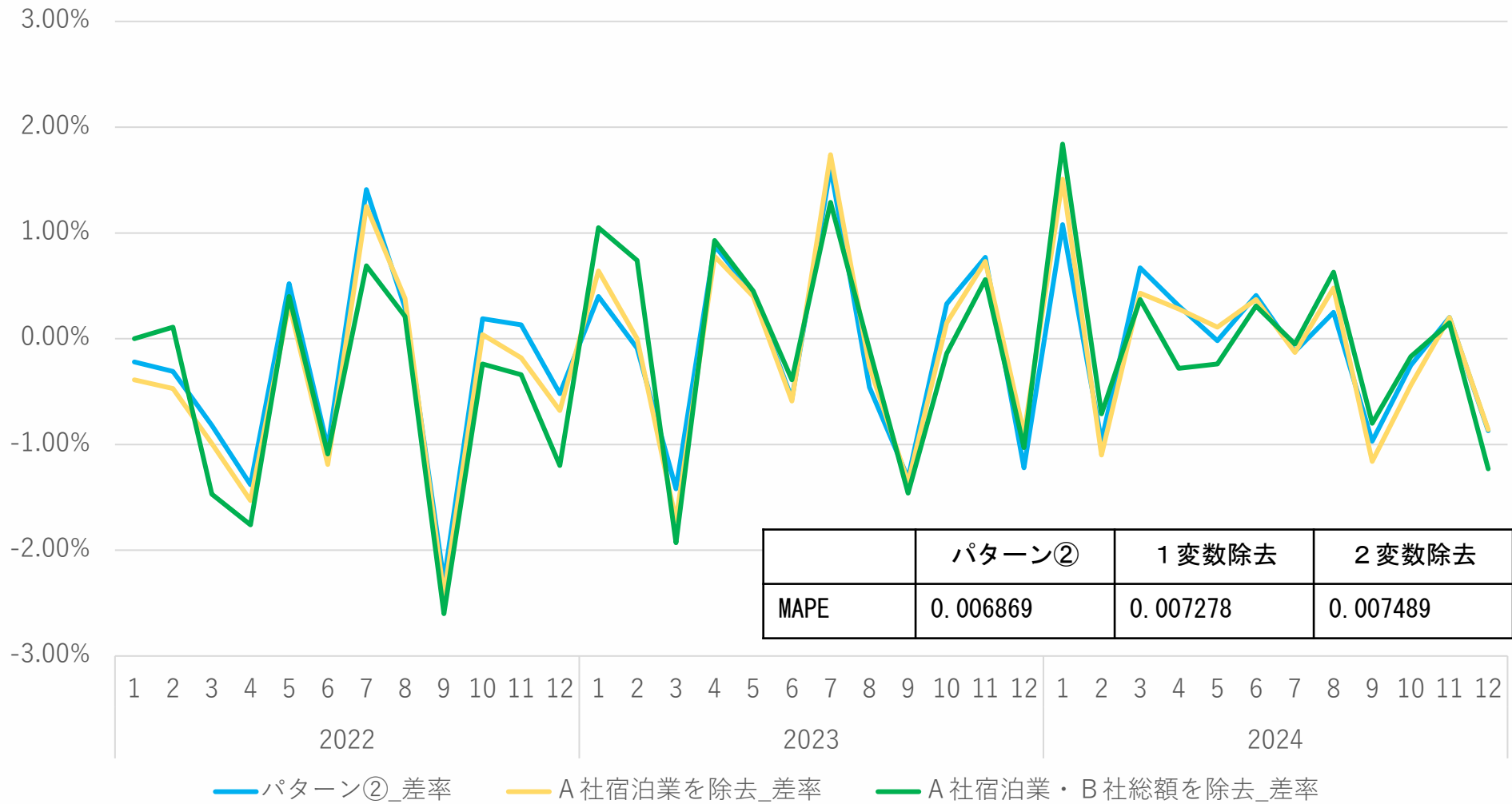
- ◆ パターン②において、PIPの高い変数に特殊な変動が生じたと仮定。
 - A社宿泊業データの1変数を除いたパターンと、A社宿泊業データ・B社総額データの2変数を除いたパターンで予測を行った。

PIPの高い変数を除いた試算結果



PIPの高い変数を除いた試算結果_（差率）

予測値の差率



PIPの高い変数を 1 ～ 2 個除いても、一定の予測精度が確認できた。

5. まとめと課題

まとめ

- ◆民間データ等を用い，サービス産業売上高計の予測を試みた．
- ◆モデル平均化法を利用することで，仮に民間データ等に特殊な変動が生じた際も，一旦そのデータを外して予測を行うという対応が可能．

課題

- ◆ベイジアンモデル平均化法は，（恐らく）日本の公的統計で初めて検討される手法で，本手法に対する理解が不十分．
- ◆他のモデル平均化法との比較．
- ◆CTIマクロへの影響の試算．
- ◆その他，CTIマクロ算出の実装に至るまでの課題の抽出とその検討．

- [1] Bencivelli, L., Marcellino, M., & Moretti, G. (2017). Forecasting economic activity by Bayesian bridge model averaging. *Empirical Economics*, **53**(1), 21-40.
- [2] Fernandez, C., Ley, E., & Steel, M. F. (2001). Model uncertainty in cross - country growth regressions. *Journal of applied Econometrics*, **16**(5), 563-576.
- [3] Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and El George, and a rejoinder by the authors). *Statistical science*, **14**(4), 382-417.
- [4] Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, **53**(3), 867-886.
- [5] Ley, E., & Steel, M. F. (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of applied econometrics*, **24**(4), 651-674.