Bayesian Clustered Ensemble Prediction for Multivariate Time Series[1]

Shonosuke Sugasawa

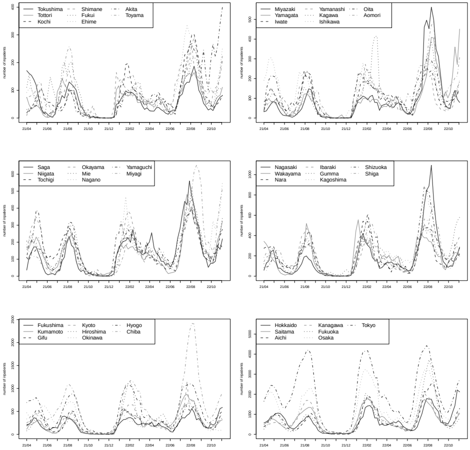
Faculty of Economics, Keio University

BIBC 2025

[1] Joint work with Genya Kobayashi, Yuki Kawakubo, Dongu Han, Taeryon Choi

Motivating Data

Weekly number of inpatients for COVID-19 by prefecture in Japan
(from April 2021 to November 2022)



47 prefectures are grouped by the means of the series.

Motivating Data

Characteristics of the data

- Multiple time series count data for 47 prefectures
- Cross section and temporal correlation

Purpose of data analysis

- Precise future prediction on the number of inpatients
- Uncertainty quantification of the prediction

Strategy

- Ensemble prediction

  In stead of relaying on a single model, combining multiple models is known to improve predictive performance.

Bayesian Predictive Synthesis (BPS)

BPS for univariate time series, y_t $(t = 1, \ldots, T)$

- J models (agents) with the predictive density, $h_{tj}(f_{tj})$ $(j = 1, \ldots, J)$
- General form of BPS (e.g. West, 1992; McAlinn and West, 2019)

$$p(y_t|\boldsymbol{\theta}_t, 1:t) = \int \alpha(y_t|\boldsymbol{f}_t, \boldsymbol{\theta}_t) \left[\prod_{j=1}^{J} h_{tj}(f_{tj})\right] d\boldsymbol{f}_t$$

  - $\boldsymbol{\theta}_t = (\theta_{t0}, \theta_{t1}, \ldots, \theta_{tJ})$: weights for J predictive models
  - $\boldsymbol{f}_t = (f_{t1}, \ldots, f_{tJ})^\top$: draw from the predictive densities, regarded as latent factors
  - $\alpha(y_t|\boldsymbol{f}_t, \boldsymbol{\theta}_t)$: synthesis function that controls how to combine J predictions

- BPS includes existing ensemble methods (e.g. Bayesian model averaging) by appropriately specifying the synthesis function.

Bayesian Predictive Synthesis

Specification of the synthesis function

- Dynamic linear model with latent factor (McAlinn and West, 2019)

$$\alpha(y_t|\boldsymbol{f}_t, \boldsymbol{\theta}_t) = \phi\Big(y_t; \theta_{t0} + \sum_{j=1}^{J} \theta_{tj} f_{tj}, \sigma_t^2\Big), \quad \boldsymbol{\theta}_t \sim \text{random walk}$$

$\phi(x; a, b)$: normal density with mean a and variance b

- BPS for multivariate time series, y_{it} $(i = 1, \ldots, n; t = 1, \ldots, T)$ (McAlinn et al., 2020)
  - Use multivariate dynamic linear models as the synthesis function.
  - Computationally intensive when n is large ($n = 47$ in our example).

- Methodological motivation: Any scalable approach for multivariate time series (of count)?

Proposal: Mixture of BPS

Idea: (soft) clustering multiple time series in terms of model importance

- J predictive models with the predictive densities $h_{itj}(f_{itj})$ $(i = 1, \ldots, n; j = 1, \ldots, J)$.
- **Mixture of BPS (MBPS)**

$$\alpha(y_{it}|\boldsymbol{f}_{it}, \boldsymbol{\theta}_{t,1:K}, \pi_{1:K}) = \sum_{k=1}^{K} \pi_k \alpha_k(y_{it}|\boldsymbol{f}_{it}, \boldsymbol{\theta}_{tk}), \quad \sum_{k=1}^{K} \pi_k = 1$$

  - $\alpha_k(y_{it}|\boldsymbol{f}_{it}, \boldsymbol{\theta}_{tk})$: kth component of the synthesis function
  - $\boldsymbol{\theta}_{tk}$: Parameters in the kth synthesis function

- MBPS reduces the number of parameters in multivariate BPS while sharing cross sectional information

Proposal: Mixture of BPS

Structure of MBPS

- For two time series i and i' in the same cluster k, MBPS places the same weight θ_{tjk} on jth model whose forecasts are generally different between i and i'.

  $\Rightarrow$ Clustering n series in terms of the importance of the kth model

- Within the same cluster, the contribution of the models to the BPS forecast is the same.
- The weights θ_{tk} are estimated from the set of time series belonging to the same component.

Model Specification

- Consider a synthesis function based on the Poisson distribution as y_{it} is count in our example.
- Cluster assignment indicator, $z_i \in \{1, \dots, K\}$ for each time series
- Hierarchical model of MBPS for count response

$$
\begin{aligned}
&y_{it} | (z_i = k, \boldsymbol{\theta}_{tk}, \boldsymbol{f}_{it}) \sim \mathrm{Po}(\exp(\boldsymbol{\theta}_{tk}^\top \boldsymbol{F}_{it})), \quad \boldsymbol{F}_{it} = (1, \boldsymbol{f}_{it}^\top)^\top \\
&\Pr(z_i = k) = \pi_k, \quad k = 1, \dots, K, \qquad (\pi_1, \dots, \pi_K) \sim \mathrm{Dir}(a_0) \\
&f_{itj} \sim N(m_{itj}, s_{itj}^2), \quad \boldsymbol{\theta}_{tk} = \boldsymbol{\theta}_{t-1,k} + \boldsymbol{e}_{tk}, \quad \boldsymbol{e}_{tk} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{tk})
\end{aligned}
$$

- (m_{itj}, s_{itj}^2): fixed values (prediction and its variance of log-intensity)
- $a_0 = (1/K, \dots, 1/K)$: it tends to produce empty clusters.

Posterior Computation (MCMC)

- Use the negative binomial approximation of the Poisson distribution with a large dispersion parameter (Hamura et al., 2025) and apply the Pólya-gamma (PG) augmentation (Polson et al., 2013).

$$
\begin{aligned}
\alpha_k(y_{it}|\boldsymbol{f}_{it}, \boldsymbol{\theta}_{tk}) \approx \tilde{\alpha}_k(y_{it}|\boldsymbol{f}_{it}, \boldsymbol{\theta}_{tk}, r) &= \frac{\Gamma(y_{it}+r)}{\Gamma(r) y_{it}!} \frac{(e^{\psi_{itk}})^{y_{it}}}{(1+e^{\psi_{itk}})^{y_{it+r}}} \\
&= 2^{-b_{it}} \exp\left\{\kappa_{it}\psi_{itk}\right\} \int_0^\infty \exp\left\{-\frac{\omega_{itk}\psi_{itk}^2}{2}\right\} p(\omega_{itk}|b_{it}, 0) d\omega_{itk}
\end{aligned}
$$

  - $b_{it} = y_{it} + r$, $\kappa_{it} = (y_{it} - r)/2$, $\psi_{itk} = \boldsymbol{\theta}_{tk}^\top \boldsymbol{F}_{it} - \log r$
  - ω_{itk} follows the PG distribution.

- The resulting joint distribution can be seen as a Gaussian dynamic linear model.

  $\Rightarrow$ Forward filtering and backward sampling can be applied.

Extension: MBPS with Heterogeneous Intercept (MBPSH)

- Intercept in BPS: adjusting the level of inadequacy of ensemble prediction
- MBPS assumes the same intercept within a cluster, which may be restrictive in practice.
- MBPS with heterogeneous intercept (MBPSH)

$$y_{it}|(z_i = k, \boldsymbol{\theta}_{tk}, \boldsymbol{f}_{it}, u_{it}) \sim \mathrm{Po}(\exp(\boldsymbol{\theta}_{tk}^\top \boldsymbol{F}_{it} + u_{it})), \quad u_{it}|(z_i = k) \sim N(0, \tau_{tk}^2),$$
$$\tau_{tk}^2 = \frac{\beta_\tau}{\gamma_t}\tau_{t-1,k}^2, \quad \gamma_t \sim \mathrm{Beta}\left(\frac{\beta_\tau n_{t-1}}{2}, \frac{(1-\beta_\tau)n_{t-1}}{2}\right).$$

It can be regarded as an intermediate model between the univariate BPS and MBPS.

Analysis of COVID-19 Hospitalization in Japan

- Total number of inpatients including severe conditions, for $n = 47$ prefectureds in Japan
- Total data period: 2020/05/07 – 2022/11/23 (134 weeks)
- $J = 4$ (number of prediction models), $K = 47$ (maximum number of clusters)
- **Prediction steps**
  - The four agent models are estimated using the first 50 weeks up to 2021/04/14.
  - MBPS is first run using the data from 2021/04/21 to 2022/03/30 (50 weeks) to produce one step (week) ahead forecast.
  - By expanding the window of past data, the one step forecasts from 2022/04/06 to 2022/11/23 (34 weeks).

Agent Models

- Model 1: Poisson dynamic generalized linear model (DGLM):

$$y_{it} \sim \mathrm{Po}(\lambda_{it}), \quad \log \lambda_{it} = \mathbf{x}_{it}^{\top} \boldsymbol{\beta}_{it}, \quad \boldsymbol{\beta}_{it} \sim N(\boldsymbol{\beta}_{i,t-1}, \boldsymbol{V}_{it})$$

  where $\mathbf{x}_{it} = (1, \tilde{I}_{it}, \tilde{I}_{it}^2)$ and $\tilde{I}_{it}$ is the log of the 7 days lag of the 14 days moving average of the number of infected.

- Model 2: Poisson generalized additive model (GAM):

$$y_{it} \sim \mathrm{Po}(\lambda_{it}), \quad \log \lambda_{it} = \mu + s(\tilde{I}_{it}) + s(t),$$

  where s denotes the smoothing splines. The model is fitted by `gam`.

Agent Models

- Model 3: Poisson integer autoregressive model (INAR):

$$y_{it} \sim \mathrm{Po}(\lambda_{it}), \quad \log \lambda_{it} = \gamma_i y_{i,t-1} + \mathbf{x}_{it}^\top \boldsymbol{\beta}_i$$

where $\mathbf{x}$ is the same as DGLM. The model is fitted by `tscount`.

- Model 4: Power-weighted Poisson-SIHR model:

$$p(y_{i,1:t}|\lambda_T) = \prod_{s=0}^{T} \left[\frac{\lambda_{i,T-s}^{y_{i,T-s}} e^{-\lambda_{i,T-s}}}{y_{i,T-s}!} \right]^{\rho_s}$$

$\lambda_{i,T-s}$ is the solution of the SIHR (susceptible-infectious-hospitalized-recovered) model:

$$dS/dt = -\alpha I \cdot S/n, \quad dI/dt = \alpha I \cdot S/n - (\beta + \delta_I)I$$
$$dH/dt = \beta I - \delta_H H, \quad dR/dt = \delta_I I + \delta_H H$$

Comparative Methods

- **MBPS, MBPSH** (proposed methods)
- **BPS**: Univariate BPS (with count response) separately in each prefecture, with the same four models as MBPS and MBPSH.
- **FMPR**: Finite mixture of Poisson regression

$$y_{it}|(z_i = k) \sim \mathrm{Po}(\exp(\mathbf{x}_{it}^\top \boldsymbol{\beta}_k)), \quad \mathbf{x}_{it} = (1, \tilde{I}_{it}, \tilde{I}_{it}^2, y_{i,t-1})^\top$$
$$\Pr(z_i = k) = \pi_k, \quad k = 1, \ldots, K$$

Performance Measures

- Cumulative absolute prediction errors (CAPE):

$$\mathrm{CAPE}_t = \sum_{i=1}^{n} \sum_{t^*=T}^{t} |y_{i,t^*+k} - \hat{y}_{i,t^*+k}|$$

- Log predictive density ratios (LPDR):

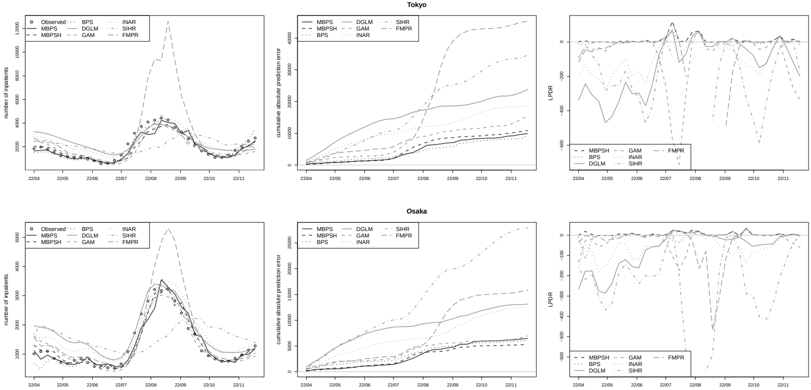
$$\mathrm{LPDR}_t = \log \frac{p_j(y_{t+k}|y_{1:t})}{p_{MBPS}(y_{t+k}|y_{1:t})}$$

- Coverage:

$$\frac{1}{T^*} \sum_t I\{y_{t+k} \in 95\% \text{ prediction interval}\}$$

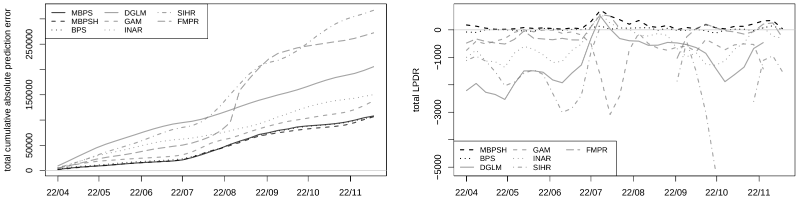
where T^* is the length of prediction periods.

One-step-ahead forecasting (Tokyo, Osaka)



Predictions (left), CAPE (middle), LPDR (right)

One-step-ahead forecasting (Total)

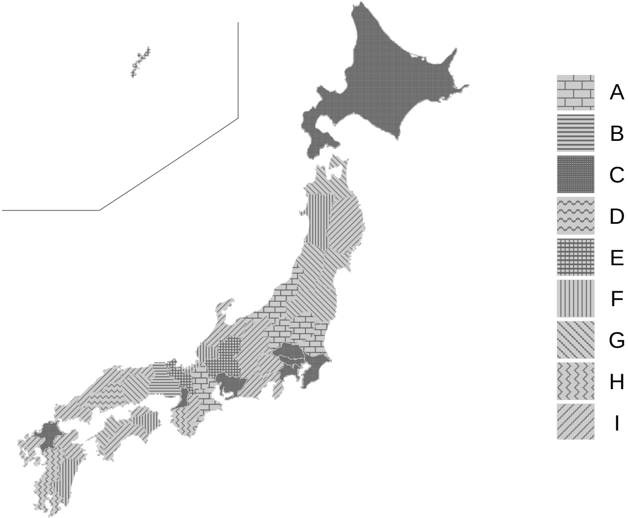


Coverage

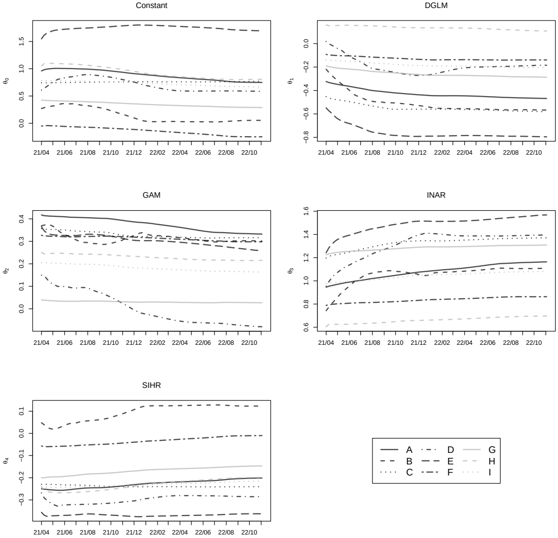
MBPS	MBPSH	BPS	DGLM	GAM	INAR	SIHR	FMPR
0.556	**0.928**	0.583	0.158	0.122	0.236	0.036	0.287

- The proposed methods provide better prediction accuracy.
- MBPSH provides much better coverage performance than MBPS.

Clustering Result on 2022/11/23



Synthesis weights by cluster



Summary

- Mixture of Bayesian predictive synthesis (MBPS)
  - BPS for multivariate time series, sharing the common synthesis weights in the same cluster
  - Introduction of heterogeneity in the intercept improves the coverage.
  - MBPS is useful for large-dimensional multivariate time series prediction.

- For more information:

  Kobayashi, G., Sugasawa, S., Kawakubo, Y., Han, D. and Choi, T. (2024). Predicting COVID-19 hospitalisation using a mixture of Bayesian predictive syntheses. *The Annals of Applied Statistics* 18, 3383-3404.

Supplement: Analysis of COVID-19 Isolation in Korea

- Daily numbers of isolated cases by first division obtained from `data.go.kr`
- Total data period: 2020/08/01 – 2021/11/30 (487 days)
- $K = n = 17$
- Multistep-ahead forecasts

Coverage

s-step	MBPS	MBPSH	BPS	DGLM	GAM	INAR	SIHR	FMPR
$s = 1$	0.918	0.964	0.932	0.605	0.169	0.848	0.028	0.351
$s = 3$	0.724	0.944	0.767	0.590	0.164	0.654	0.022	0.354
$s = 7$	0.637	0.953	0.697	0.575	0.169	0.456	0.028	0.360