Bayesian forecast combination using time-varying features

Feng Li
feng.li@cufe.edu.cn
http://feng.li/

School of Statistics and Mathematics
Central University of Finance and Economics

Authors

Li Li
(PhD in Beihang, Postdoc in USTB)

Yanfei Kang
(Beihang University)

Feng Li
(Central Uni. of Finance & Economics)

- Appeared in *International Journal of Forecasting* 39(3): 1187-1302, doi: `10.1016/j.ijforecast.2022.06.002`.
- Feng Li and Yanfei's research are supported by the National Natural Science Foundation of China.

Forecast combination is ***GREAT****!*

but ...

- Forecast combinations with pure machine learning approach usually lack interpretability.
- Thus making them difficult to connect with decision-making.
- Sometimes, those proposed methods are not robust against real-world data. Inputs from experts are typically required,
- ... and it is tricky.

The Bayesian forecast evaluation tool

- The **joint predictive probability** for h-ahead forecast is

$$p_k(y_{(T+1):(T+h)}|y^o_{1:T}, M) = \int \prod_{i\in(T+1):(T+h)} p(y_i|\beta)p(\beta|y^o_{1:T})\mathrm{d}(\beta). \tag{1}$$

  which is calculated by averaging the predictive likelihood $\prod_{i\in(T+1):(T+h)} p(y_i|\beta)$ over the posterior $p(\beta|y^o_{1:T})$.
- The one-step-ahead **log predictive score** (LPS) function (Geweke, 2001) of **a single prediction model** M

$$LS\left(y_{T+1}|y^o_{1:T}, M\right) = \sum_{t=1}^{T} \log p\left(y_{t+1}|y^o_{1:t}, M\right). \tag{2}$$

  is ***the unquestionable model evaluation tool for decision makers*** (Geweke, 2001; Geweke & Amisano, 2010).
- The LPS has three main advantages:
  - LPS is based on out-of-sample **probability forecasting**.
  - LPS is easy to compute based on Monte Carlo simulations;
  - LPS is not sensitive to the choice of the priors compared with the marginal likelihood based criterions (Kass, 1993; Richardson & Green, 1997).

Forecasting combination: A Bayesian setup

- We extend LPS to evaluate **a pool of models** $\mathcal{M} = \{M_1, M_2, \cdots, M_m\}$ for a collection of target series $Y^o_S = \{y^o_1, y^o_2, \cdots, y^o_s\}$ (o for "observed").
- The predictive densities of one single series y_{T+1} are written as

$$p\left(y_{T+1} \middle| Y^o_T, \mathcal{M}\right) = \sum_{i=1}^{m} w_i p\left(y_{T+1} | y^o_{1:T}, M_i\right),$$

- The log scoring (LS) rule could evaluate combinations of probability densities (Geweke & Amisano, 2011):

$$LS\left(Y^o_S\right) = \sum_{s=1}^{S} \log\left[\sum_{i=1}^{m} w_i p\left(y^o_{T+1} | y^o_{1:T}, M_i\right)\right]. \tag{3}$$

- Limitations in Geweke & Amisano (2011)
  - *LS* are **constant over time**, only related to prediction models.
  - The number of series is limited.
  - **Not a full probabilistic framework**, an optimization algorithm is used to estimate the optimal weights.

FEBAMA: feature-based Bayesian forecast model averaging

- We **learn** the relationship between LS and weights based on historical data.
- We use time series features to construct **time-varying weights** in the forecasting combination by the following softmax function:

$$w_{t,i} = \frac{\exp\{X_t\beta_i\}}{1 + \sum_{i=1}^{n-1} \exp\{X_t\beta_i\}}, i = 1, 2, \cdots, n-1 \tag{4}$$

- The observed log score $LS(Y_T^o | X_{T-1}, \beta)$ is essentially a log likelihood function of feature-based log-predictive scores.
- **A full Bayesian framework**, namely *feature-based Bayesian forecast model averaging* (**FEBAMA**), is formulated

$$\log p(\beta|X) = LS(\beta, X) + \log p(\beta) + constant \tag{5}$$

where X is the time series feature matrix and LS is the corresponding log likelihood function (log score).

FEBAMA: Variable selection

- With the above setup, a Bayesian variable selection method can be directly applied to select important time series features, which we denote as "FEBAMA+VS".
- Let $\mathcal{I}_i$ be the variable selection indicator vector for the model M_i, and $\beta_{\mathcal{I}_i}$ is the corresponding coefficient vector of features. So the expression for the combination weights is changed to

$$w_{i,t} = \frac{\exp\{x_t' \beta_{\mathcal{I}_i}\}}{1 + \sum_{i=1}^{m-1} \exp\{x_t' \beta_{\mathcal{I}_i}\}}, i = 1, 2, \cdots, m-1, \tag{6}$$

- The joint posterior for both the coefficients and variable selection indicators is

$$p(\beta, \mathcal{I} | Y_T, X_T, \mathcal{M}) \propto \prod_{t=1}^{T} p(y_t | Y_{t-1}, X_t, \beta, \mathcal{I}, \mathcal{M}) p(\beta, \mathcal{I}), \tag{7}$$

for a target series $Y_T = \{y_1, y_2, \cdots, y_T\}$ in forecasting model pool $\mathcal{M}$.

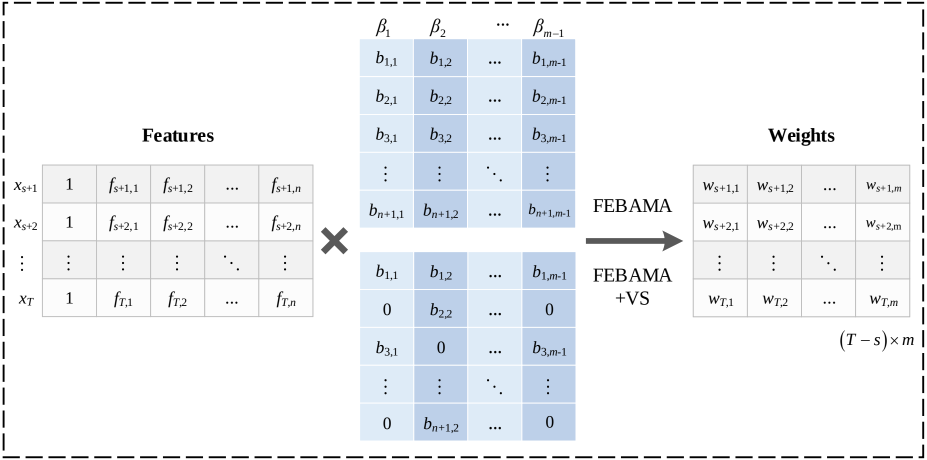
Comparison between different methods

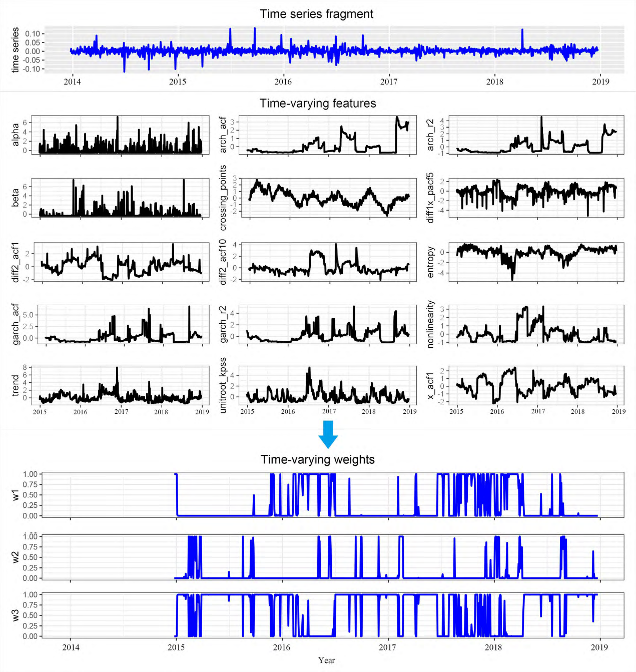
	SA	OptPool	FEBAMA	FEBAMA+VS
Whether weights are optimized	-	✓	✓	✓
Whether weights are time-verying	-	-	✓	✓
Whether including features into weights	-	-	✓	✓
Whether features can be penalized/selected	-	-	-	✓

Why FEBAMA?

- It **improves the interpretability** of forecasting combination by defining a log score with time series features.
- The Bayesian framework assesses both
  - **the model uncertainty**, and
  - **the forecast combination uncertainty**.
- Expert's information for forecast combination is easy to elaborate through the prior.
- Other forecast evaluation metrics, such as MASE, is equally well applied in our framework.

Time-varying features

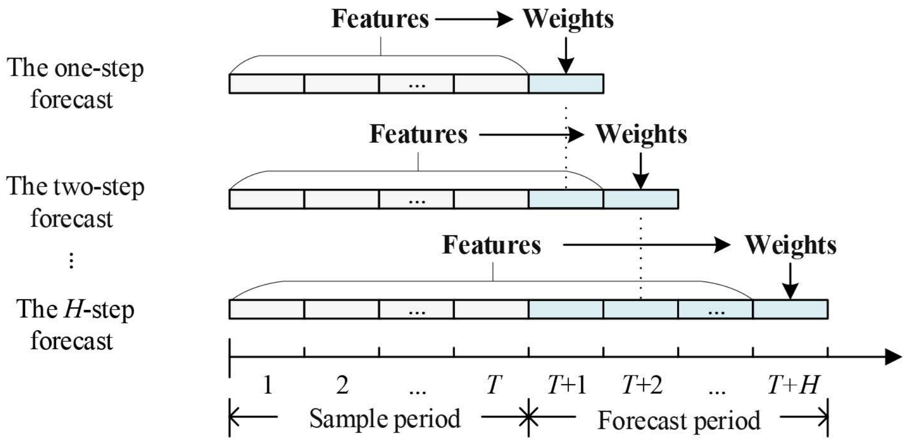



Time series fragment
time series
Time-varying features
alpha
arch_acf
arch_r2
beta
crossing_points
diff1x_pacf5
diff2_acf1
diff2_acf10
entropy
garch_acf
garch_r2
nonlinearity
trend
unitroot_kpss
x_acf1
Time-varying weights
w1
w2
w3
Year

Parameter settings in the priors

Parameter	Description	Prior
β_i	Coefficient vectors of features for the model M_i.	Each element follows an independent normal distribution $N(0, \sigma^2)$.
$\mathcal{I}_i$.	Indicator vectors of features for the model M_i.	Each element follows an independent $Bernoulli(p)$ distribution where $p \sim Beta(1, 1)$.
$\beta_{\mathcal{I}_i}$.	Nonzero coefficient vectors of features on the condition of $\mathcal{I}_i$.	Each element follows a conditional normal distribution.

Scheme of one-step and multi-steop forecast



Application to financial data and M3 data

- We illustrate the effectiveness and superiority of our framework through two experiments based on stock market data and M3 competition data.
- The simple averaging (SA) and an improved version of the optimal pool (OptPool) (Geweke & Amisano, 2011) are used as benchmark methods in the comparison.
- The original OptPool method only generates one-step predictions. We extend to multiple step predictions, which we label it as OptPool$'$.

The candidate models and features for stock data

Feature	Description	Value range
F_1: alpha	ETS(A,A,N) $\widehat{\alpha}$	$[0, 1]$
F_2: arch_acf	ARCH ACF statistic	$(0, \infty)$
F_3: arch_r2	ARCH R^2 statistic	$[0, 1]$
F_4: beta	ETS(A,A,N) $\widehat{\beta}$	$[0, 1]$
F_5: crossing_points	Number of times the time series crosses the median	$\{1, 2, 3, \dots\}$
F_6: diff1x_pacf5	Sum of squares of first 5 PACF values of differenced series	$(0, \infty)$
F_7: diff2_acf1	First ACF value of the twice-differenced series	$(-1, 1)$
F_8: diff2_acf10	Sum of squares of first 10 ACF values of twice-differenced series	$(0, \infty)$
F_9: entropy	Spectral entropy	$(0, 1)$
F_{10}: garch_acf	GARCH ACF statistic	$(0, \infty)$
F_{11}: garch_r2	GARCH R^2 statistic	$[0, 1]$
F_{12}: nonlinearity	Nonlinearity coefficient	$[0, \infty)$
F_{13}: trend	Strength of trend	$[0, 1)$
F_{14}: unitroot_kpss	Test statistic based on KPSS test	$(0, \infty)$
F_{15}: x_acf1	First ACF value of the original series	$(-1, 1)$

Table: In-sample log score comparison of individual models and different forecast combination methods. We consider all the possible combinations of M_1, M_2 and M_3 (e.g., $M_{1,2}$ means the combination of M_1 and M_2). "Total" means the average results of the four combinations. For each combination (row), the highest LS is marked in bold. The last column refers to the p-values of the two-sided DM tests against the best method of the three benchmarks among SA, OP, and GP. The p-values less than 0.05 are marked in bold.

Individual	M_1: GARCH	M_2: RGARCH	M_3: SV		
	-1.3395	-1.2867	-1.2911		
Combination	SA	OP	GP	FEBAMA	DM test p-value
$M_{1,2}$	-1.3138	-1.2885	-1.2771	**-1.2752**	0.1749
$M_{1,3}$	-1.3153	-1.2911	-1.2850	**-1.2783**	**0.0326**
$M_{2,3}$	-1.2890	-1.2832	-1.2770	**-1.2709**	0.0821
$M_{1,2,3}$	-1.3065	-1.2827	-1.2716	**-1.2626**	**0.0075**
Total	-1.3062	-1.2864	-1.2777	**-1.2717**	

Table: Out-of-sample forecasting performance of the individual models and different forecast combination methods with regard to LS. We consider all the possible combinations of M_1, M_2 and M_3 (e.g., $M_{1,2}$ means the combination of M_1 and M_2). "Total" means the average results of the four combinations. For each combination (in rowwise), the highest LS is marked in bold.

Individual	M_1: GARCH	M_2: RGARCH	M_3: SV		
	-1.3310	-1.3065	-1.3033		
Combination	SA	OP	GP	FEBAMA	FEBAMA+VS
$M_{1,2}$	-1.3109	-1.3067	-1.2984	-1.2791	**-1.2756**
$M_{1,3}$	-1.3102	-1.3054	-1.2913	-1.2744	**-1.2726**
$M_{2,3}$	-1.3024	-1.3008	-1.2958	-1.2776	**-1.2775**
$M_{1,2,3}$	-1.3031	-1.3015	-1.2943	-1.2768	**-1.2703**
Total	-1.3067	-1.3036	-1.2950	-1.2770	**-1.2740**

The candidate models and features for the M3 data

- We consider the following forecasting models:

Models	Description	Setting
A_1: `ets`	automated exponential smoothing algorithm (Hyndman et al., 2002)	`model = "AAN"`
A_2: `naive`	naïve	
A_3: `rw_drift`	random walk with drift	`drift = TRUE`
A_4: `auto.arima`	automated ARIMA algorithm (Hyndman & Khandakar, 2008)	

- The features are

Features	Description	Values
F_1: `x_acf1`	first ACF value of the original series	$(-1, 1)$
F_2: `diff1_acf1`	first ACF value of the differenced series	$(-1, 1)$
F_3: `entropy`	spectral entropy	$(0, 1)$
F_4: `alpha`	the smoothing parameter for the level in ETS(A,A,N)	[0,1]
F_5: `beta`	the smoothing parameter for the trend in ETS(A,A,N)	[0,1]
F_6: `unitroot_kpss`	test statistic based on KPSS test	$(0, \infty)$

Forecast performance based on M3 Monthly data

	LS (density forecasts)				MASE (point forecasts)			
Component	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
	-4.368	-4.435	-3.698	-5.126	2.514	2.599	2.233	2.192
Combinations	SA	OptPool′	FEBAMA	FEBAMA+VS	SA	OptPool′	FEBAMA	FEBAMA+VS
C12	**-4.349**	-4.372	-4.359	-4.352	2.520	**2.513**	2.521	2.514
C13	-3.397	-3.283	-3.258	**-3.155**	2.238	2.207	2.223	**2.193**
C14	-3.524	-3.565	-3.353	**-3.320**	2.228	2.218	2.189	**2.179**
C23	-3.480	-3.280	-3.286	**-3.189**	2.322	2.257	2.286	**2.253**
C24	-3.510	-3.527	-3.333	**-3.328**	2.234	2.208	2.195	**2.187**
C34	**-3.406**	-3.842	-3.756	-3.611	2.074	2.135	2.139	**2.112**
C123	-3.669	-3.273	-3.287	**-3.202**	2.325	**2.208**	2.238	2.212
C124	-3.754	-3.517	-3.399	**-3.288**	2.307	2.217	2.208	**2.173**
C134	**-3.158**	-3.423	-3.293	-3.162	2.135	2.172	2.162	**2.124**
C234	-3.178	-3.401	-3.265	**-3.146**	2.147	2.158	2.151	**2.119**
C1234	-3.399	-3.480	-3.258	**-3.157**	2.209	2.172	2.165	**2.129**
Total	-3.529	-3.542	-3.441	**-3.355**	2.249	2.224	2.225	**2.120**

Forecasting performance for point and density forecasts

Table: Average LS and MASE of different methods for density forecasts and point forecasts, respectively. The results for the individual models and all the possible combinations are presented. "Total" means the average results of all the combinations.

	LS					MASE					
Individual	M_1	M_2	M_3	M_4		M_1	M_2	M_3	M_4		
	-4.368	-4.435	-3.698	-5.126		2.514	2.599	2.233	2.192		
Combination	SA	OP	GP	FEBAMA	FEBAMA +VS	SA	OP	GP	FFORMA	FEBAMA	FEBAMA +VS
$M_{1,2}$	**-4.349**	-4.372	-4.368	-4.358	-4.352	2.520	2.513	2.514	**2.512**	2.519	2.514
$M_{1,3}$	-3.397	-3.283	-3.265	-3.224	**-3.155**	2.238	2.207	2.212	2.208	2.219	**2.193**
$M_{1,4}$	-3.524	-3.565	-3.438	-3.353	**-3.282**	2.228	2.218	2.220	2.197	2.189	**2.168**
$M_{2,3}$	-3.480	-3.280	-3.257	-3.255	**-3.153**	2.322	2.257	**2.243**	2.262	2.277	2.245
$M_{2,4}$	-3.510	-3.527	-3.432	-3.331	**-3.269**	2.234	2.208	2.218	2.190	2.195	**2.172**
$M_{3,4}$	**-3.406**	-3.842	-3.814	-3.665	-3.611	2.174	2.135	2.146	2.154	2.123	**2.108**
$M_{1,2,3}$	-3.669	-3.273	-3.561	-3.187	**-3.165**	2.325	2.208	2.263	2.243	**2.138**	2.201
$M_{1,2,4}$	-3.754	-3.517	-3.632	-3.384	**-3.252**	2.307	2.217	2.202	2.212	2.208	**2.167**
$M_{1,3,4}$	-3.158	-3.423	-3.318	-3.067	**-3.064**	2.135	2.172	2.197	2.116	2.113	**2.104**
$M_{2,3,4}$	-3.178	-3.401	-3.397	-3.150	**-3.098**	2.147	2.158	2.132	2.130	2.130	**2.118**
$M_{1,2,3,4}$	-3.399	-3.480	-3.523	-3.178	**-3.122**	2.209	2.172	2.221	2.165	2.153	**2.120**
Total	-3.529	-3.542	-3.546	-3.377	**-3.320**	2.249	2.224	2.233	2.217	2.206	**2.192**

R package

```r
devtools::install_github("lily940703/febama")

lpd_features = lpd_features_multi(data = data_example,
                                  model_conf = model_conf_default)
parameters = febama_mcmc(data = lpd_features,
                         model_conf = model_conf_curr)

data_fore = forecast_feature_results_multi(
                      ts = data_example,
                      model_conf = model_conf_curr,
                      data = lpd_features,
                      beta_out = parameters)
```

Future work

- FEBAMA with machine learning features in e.g. Li et al., 2020.
- FEBAMA at scale with a much larger model pool.
- FEBAMA using efficient stochastic Markov chain Monte Carlo methods.

References

- Li, X., Kang, Y. & Li, F. (2020). "Forecasting with Time Series Imaging". *Expert Systems with Applications* 160, p. 113680.
- Geweke, J. & Amisano, G. (2011). "Optimal Prediction Pools". *Journal of Econometrics* 164.(1), pp. 130–141.
- Geweke, J. & Amisano, G. (2010). "Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns". *International Journal of Forecasting* 26.(2), pp. 216–230.
- Hyndman, R. J. & Khandakar, Y. (2008). "Automatic Time Series Forecasting: The Forecast Package for R". *Journal of Statistical Software* 27.(3), pp. 1–22.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D. & Grose, S. (2002). "A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods". *International Journal of Forecasting* 18.(3), pp. 439–454.
- Geweke, J. (2001). "Bayesian Econometrics and Forecasting". *Journal of Econometrics* 100.(1), pp. 11–15.
- Richardson, S. & Green, P. (1997). "On Bayesian Analysis of Mixtures with an Unknown Number of Components (with Discussion)". *Journal of the Royal Statistical Society. Series B. Statistical Methodology* 59.(4), pp. 731–792.
- Kass, R. (1993). "Bayes Factors in Practice". *Journal of the Royal Statistical Society. Series D (The Statistician)* 42.(5), pp. 551–560.

Thank you!

Web: https://feng.li
Lab: https://kllab.org