Feature-based Bayesian Forecast Model Averaging

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

School of Statistics and Mathematics
Central University of Finance and Economics

Authors

Li Li
(Beihang University)

Yanfei Kang
(Beihang University)

Feng Li
(Central Uni. of Finance & Economics)

- 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 of 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

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

  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

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

  is ***the unquestionable model evaluation tool for decision makers*** (Geweke, 2001; Geweke and 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 and Green, 1997).

Forecasting combination: A Bayesian setup

- We extend LPS to evaluate **a pool of models** $A = \{A_1, A_2, \cdots, A_m\}$ for a collection of target series $Y_S^o = \{y_1^o, y_2^o, \cdots, y_s^o\}$ (o for "observed").
- The predictive densities of one single series y_{T+1} are written as

$$p\left(y_{T+1} | Y_T^o, A\right) = \sum_{i=1}^{m} w_i p\left(y_{T+1} | y_{1:T}^o, A_i\right),$$

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

$$LS\left(Y_S^o\right) = \sum_{s=1}^{S} \log \left[\sum_{i=1}^{m} w_i p\left(y_{T+1}^o | y_{1:T}^o, A_i\right) \right]. \tag{1}$$

- Note that in Geweke and Amisano (2011),
  - *LS* are **constant over time**, only related to prediction models.
  - The number of series is limited.
  - An optimization algorithm is used to estimated 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:

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

- 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

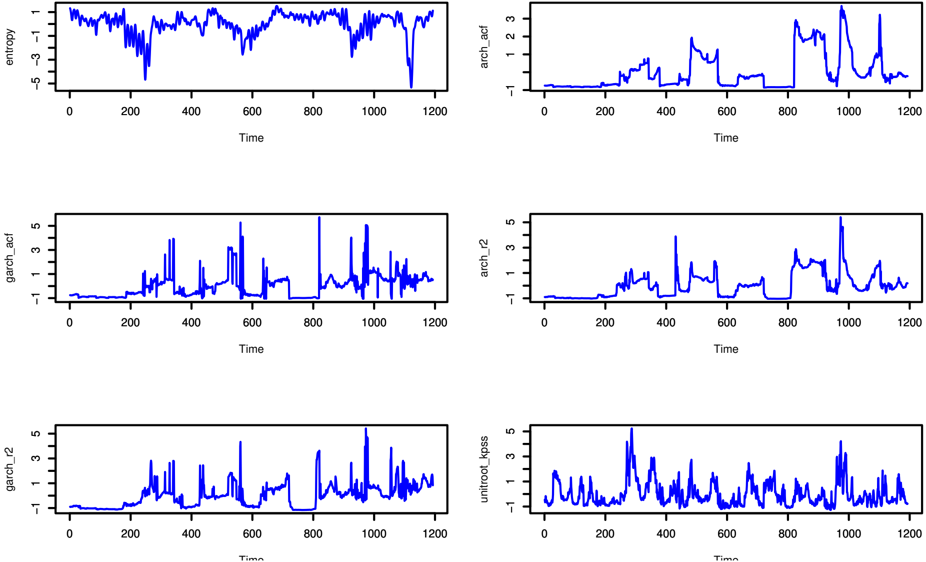
$$\log p(\beta | X) = LS(\beta, X) + \log p(\beta) + constant$$

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

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



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 optimal pool (OptPool) (Geweke and 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$'$".

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	-	-	-	✓

The candidate models and features for stock data

- The features are

Features	Description	Values
F_1: entropy	spectral entropy	$(0, 1)$
F_2: arch_acf	ARCH ACF statistic (sum of squares of the first 12 autocorrelations of x^2)	$(0, \infty)$
F_3: garch_acf	GARCH ACF statistic (sum of squares of the first 12 autocorrelations of r^2)	$(0, \infty)$
F_4: arch_r2	ARCH R^2 statistic (R^2 value of an AR model applied to x^2)	$[0, 1]$
F_5: garch_r2	GARCH R^2 statistic (R^2 value of an AR model applied to r^2)	$[0, 1]$
F_6: unitroot_kpss	test statistic based on KPSS test	$(0, \infty)$

- We consider the following forecasting models:

Component	A_1: GARCH	A_2: EGARCH	A_3: SV
	-1.3395	-1.2850	-1.2911
Combinations	SA	OptPool	FEBAMA
C12	-1.3122	-1.2850	**-1.2818**
C13	-1.3153	-1.2911	**-1.2868**
C23	-1.2881	-1.2821	**-1.2770**
C123	-1.3052	-1.2812	**-1.2740**
Total	-1.3052	-1.2849	**-1.2799**

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 and 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

Component	LS (density forecasts)				MASE (point forecasts)			
	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**

Future work

- FEBAMA with machine learning features in e.g. Li et al. (2020)
- FEBAMA at scale with stochastic Markov chain Monte Carlo methods

References

Geweke, J. (2001), "Bayesian econometrics and forecasting," *Journal of Econometrics*, 100, 11–15.

Geweke, J. and Amisano, G. (2010), "Comparing and evaluating Bayesian predictive distributions of asset returns," *International Journal of Forecasting*, 26, 216–230.

— (2011), "Optimal Prediction Pools," *Journal of Econometrics*, 164, 130–141.

Hyndman, R. J. and Khandakar, Y. (2008), "Automatic time series forecasting: the forecast package for R," *Journal of Statistical Software*, 26, 1–22.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., and Grose, S. (2002), "A state space framework for automatic forecasting using exponential smoothing methods," *International Journal of Forecasting*, 18, 439–454.

Kass, R. (1993), "Bayes factors in practice," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 42, 551–560.

Li, X., Kang, Y., and Li, F. (2020), "Forecasting with time series imaging," *Expert Systems with Applications*, 160, 113680.

Richardson, S. and Green, P. (1997), "On Bayesian analysis of mixtures with an unknown number of components (with discussion)," 59, 731–792.

Thank you!

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