CMAF FTT

On performance of temporal aggregation in time series forecasting

Bahman Rostami-Tabar (@Bahman_R_T)

Website www.bahmanrt.com

Outline

- Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?

- How does temporal aggregation approaches perfrom on M4 competition data?

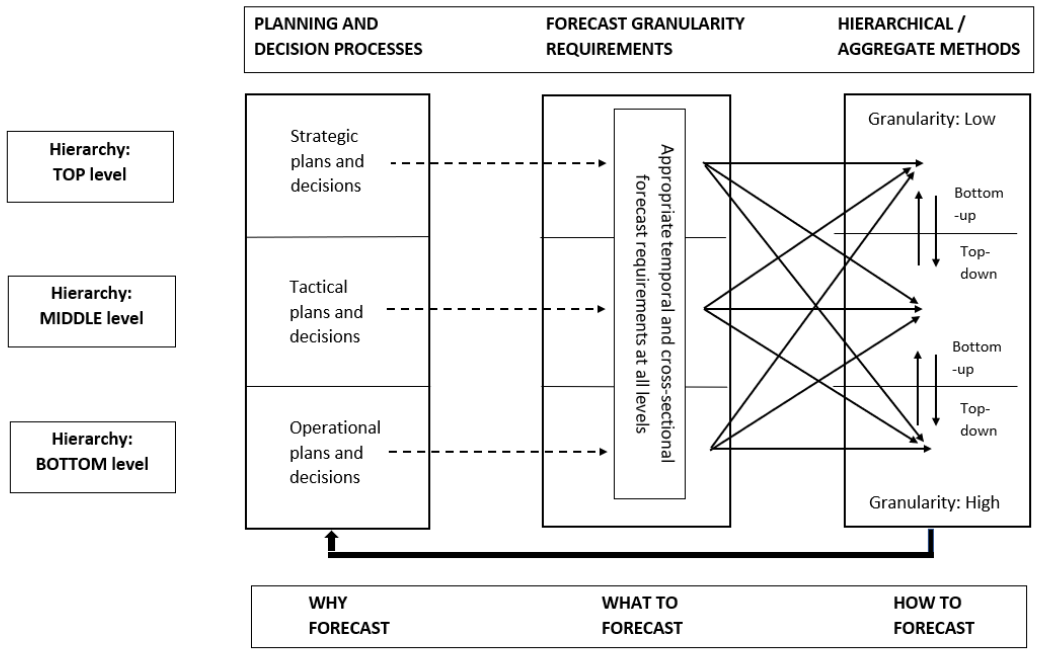
- Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (**Working paper 1**)?

- How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF (**Working paper 2**)?

Outline

- Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?
- How does temporal aggregation approaches perfrom on M4 competition data?
- Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (**Working paper 1**)?
- How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF (**Working paper 2**)?

Using time series forecasting to inform decisions



Babai, M. Zied, John E. Boylan, and Bahman Rostami-Tabar. "Demand forecasting in supply chains: a review of aggregation and hierarchical approaches." International Journal of Production Research (2021): 1-25.

Data and forecast time granularity

- Forecasting time granularity level and its horizon are determined by decisions made in the light of forecast.
- One common assumption is that time series granularity matches forecast requirement, i.e. to produce daily forecasts, we use daily time series.
- However, the level of time series granularity does not necessarily match the level of forecast granularity.
- The level of temporal granularity in the forecast might be lower than the existing time series granularity. For instance, while a forecast might be required at the annual level, a monthly time series is available. With advances in IT, data is often recorded at the finest temporal granularity (e.g. arrival time)

Time series forecasting problem

- We consider a time series forecasting problem where an original time series has a higher temporal granularity (e.g. monthly) than the required forecast (e.g. annual).
- We aim to generate a forecast of the total value over a number of time periods ahead, forecast horizon aggregation or forecast over the leadtime period.

1 Mohammadipour, Maryam, and John E. Boylan. "Forecast horizon aggregation in integer autoregressive moving average (INARMA) models." Omega 40.6 (2012): 703-712.

A key question then to be answered is:

should the original series be used to generated the forecast for the required horizon and then sum them up to obtain the forecast horizon aggregation (lead-time), i.e. Aggregate Forecast (AF) or should we first aggregate time series to match the forecast requirement granularity and then extrapolate directly at that level, i.e. Aggregate Data (AD).

I will illustrate these approaches usign a simple example.

There is no disaggregation to the original time granularity

Terminilogy

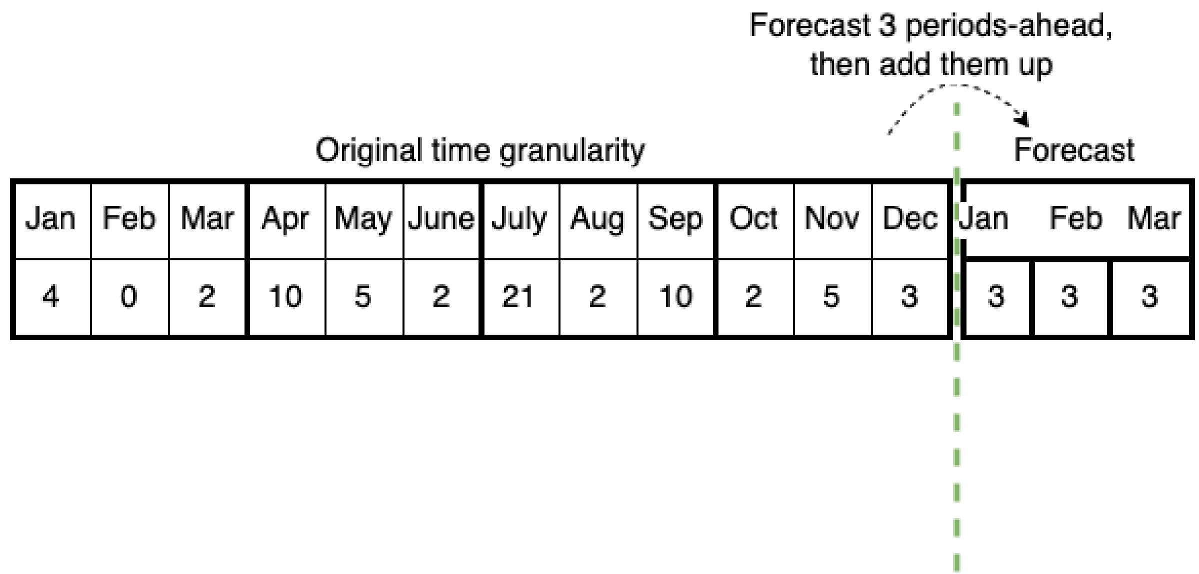
One time series

- Data time granularity (e.g. daily, monthly, annual)
- Forecast time granularity (e.g. daily, monthly, annual)
- Forecast horizon (e.g. 12 months ahead)
- Forecast horizon aggregation /leadtime (e.g. 1 week, 1 quarter, 1 year)
- Temporal aggregation
    - Aggregate Forecast (or Bottom-Up)
    - Aggregate Data
        - Non-overlapping temporal aggregation (NOA)
        - Overlapping temporal aggregation (OA)

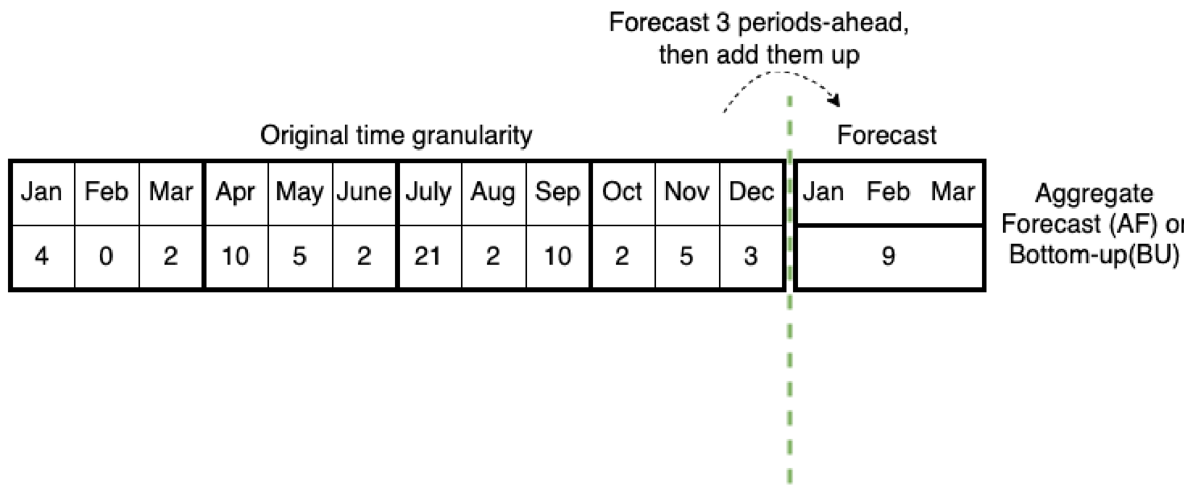
Forecast horizon aggregation: an example

Original time granularity												Forecast		
Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
4	0	2	10	5	2	21	2	10	2	5	3	?		

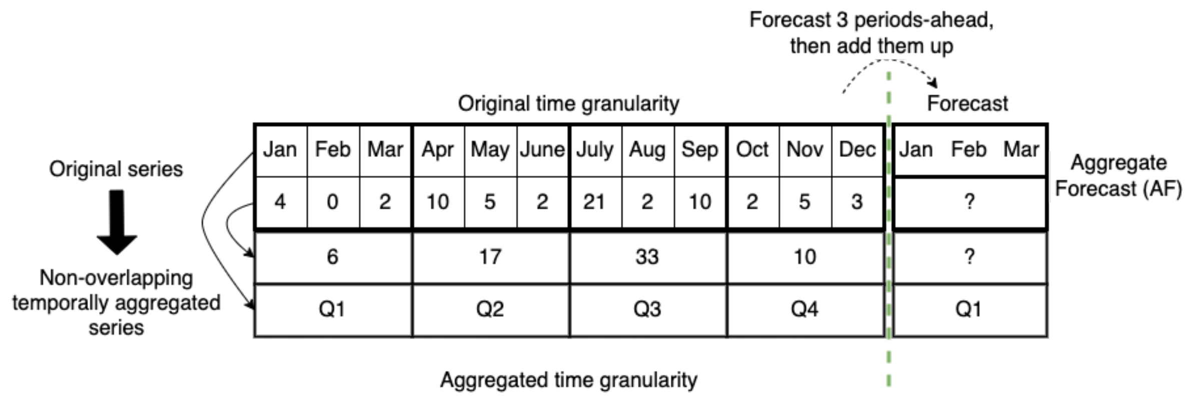
Temporal aggregation: aggregate forecast



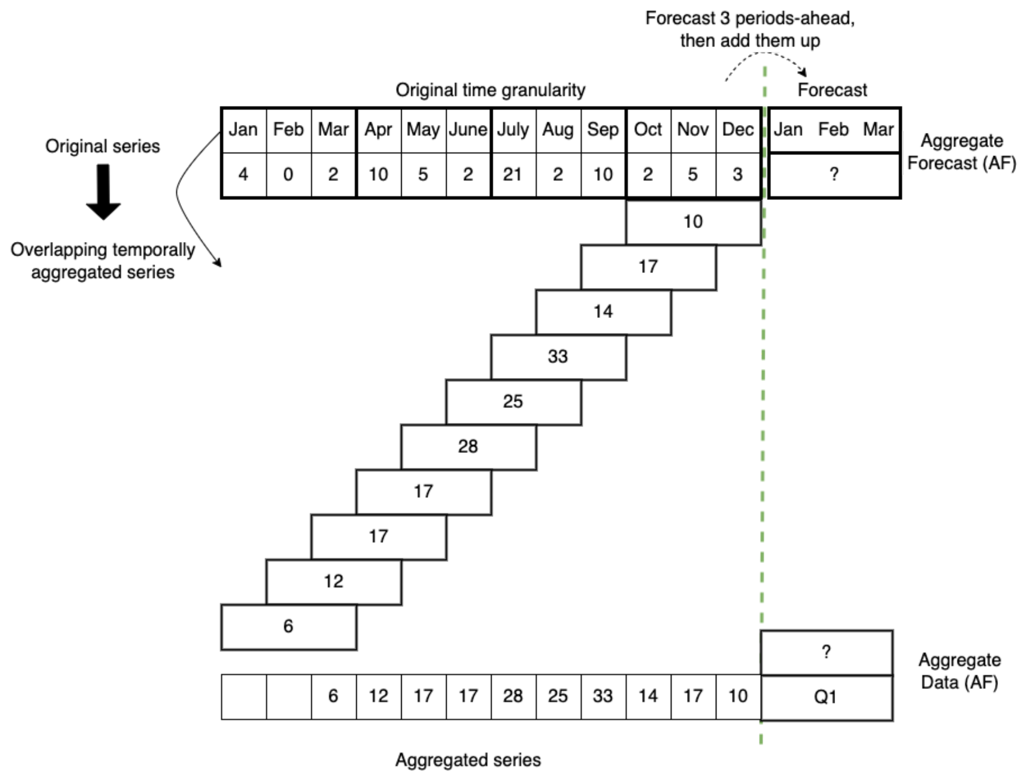
Temporal aggregation: aggregate forecast



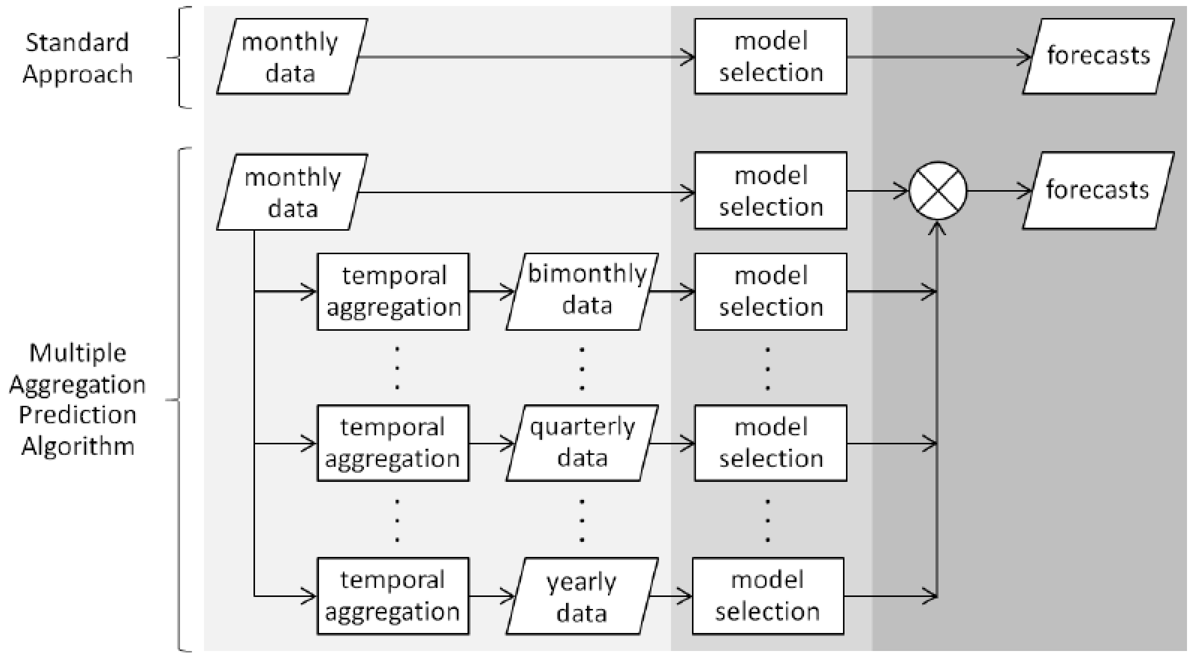
Non-overlapping temporal aggregation: aggregate data



Overlapping temporal aggregation

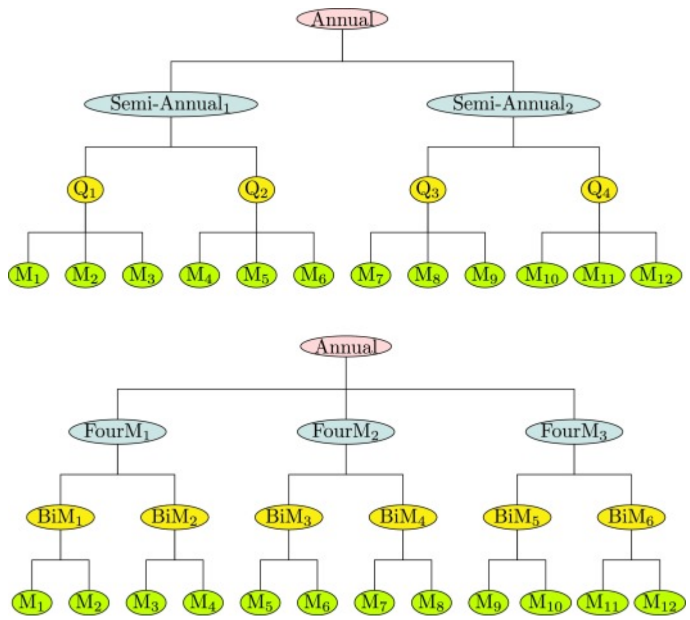


Using information at multiple levels of time granularity instead of a single level - MAPA



Kourentzes, Nikolaos, Fotios Petropoulos, and Juan R. Trapero. "Improving forecasting by estimating time series structural components across multiple frequencies." International Journal of Forecasting 30.2 (2014): 291-302.

Using information at multiple levels of time granularity instead of a single level- temporal hierarchies



Athanasopoulos, George, et al. "Forecasting with temporal hierarchies." European Journal of Operational Research 262.1 (2017): 60-74.

It is often recommended to aggregate data and then forecast when a time series history is recorded at a higher frequency time granularity (e.g. monthly) and forecast is required at alower level (e.e. annual).

For an exmpel, please refer to page 153 of Profit from Your Forecasting Software, by Paul Goodwin.

Let's examine the performance of aggregating data versus aggregating forecat approaches using M4 competition dataset

Outline

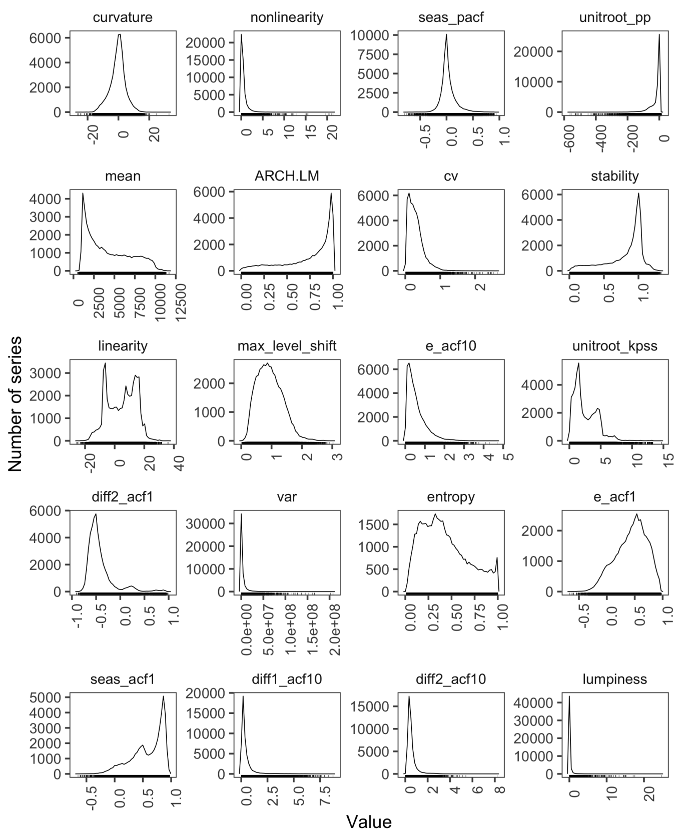
- Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?
- How does temporal aggregation approaches perfrom on M4 competition data?
- Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (**Working paper 1**)?
- How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF? (**Working paper 2**)

Time series data

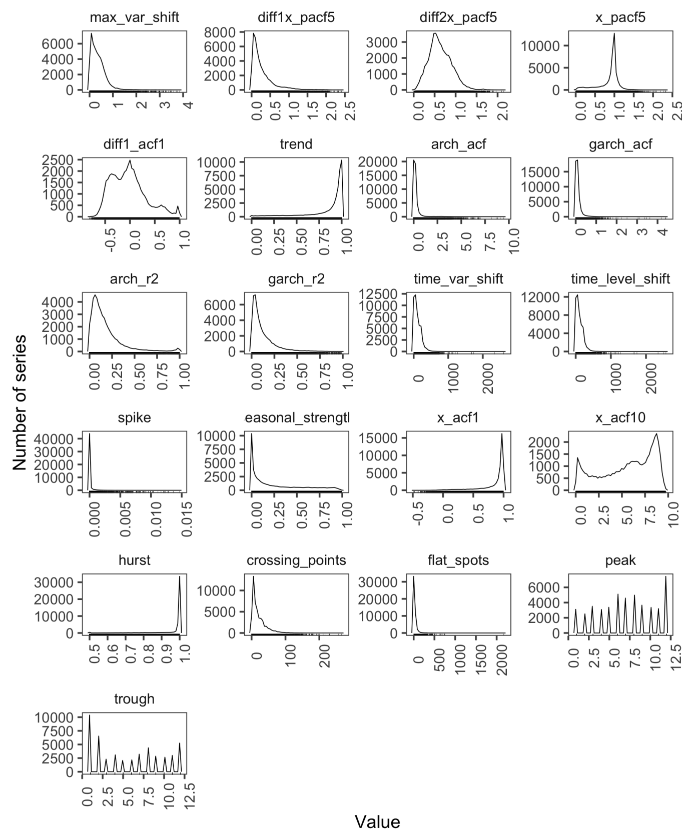
- M4 competition data time series
    - 24,000 Quarterly
    - 48,000 monthly
    - 4,227 daily
- Time series features
    - 42 features
    - Extract features using `tsfeatures::tsfeatures()` in R
- Forecasting methods: Exponential Smoothing State Space (ETS) (ARIMA is also considered).
- Point forecast accuracy measure: Mean Absolute Scaled Error (MASE), Root Mean Squared Scaled Error (RMSSE), and more.
- Time series cross validation is performed.

https://supplychainanalytics.shinyapps.io/Evaluation_of_ML_models/.

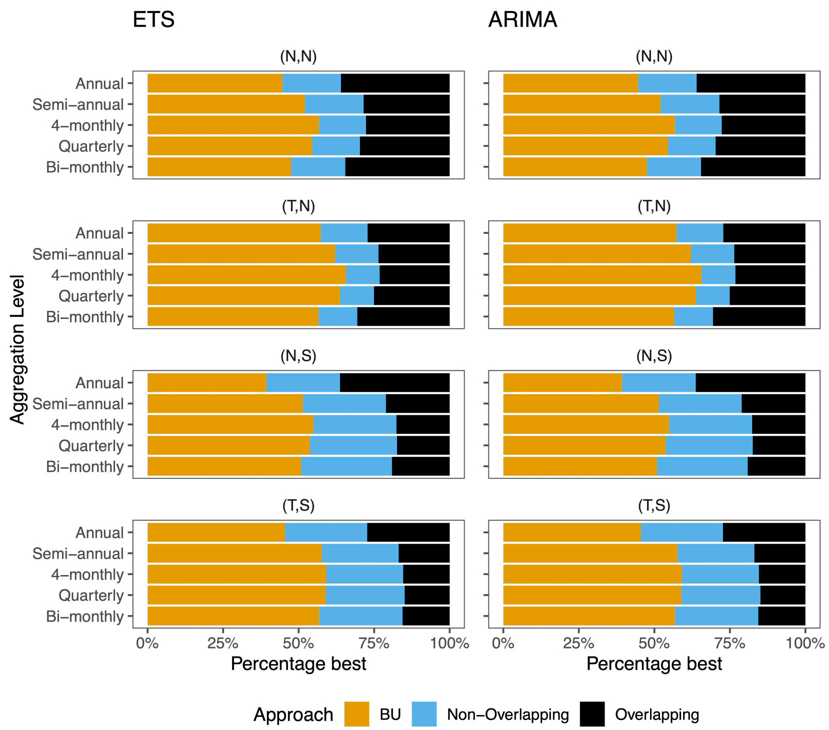
M4 Monthly time series features



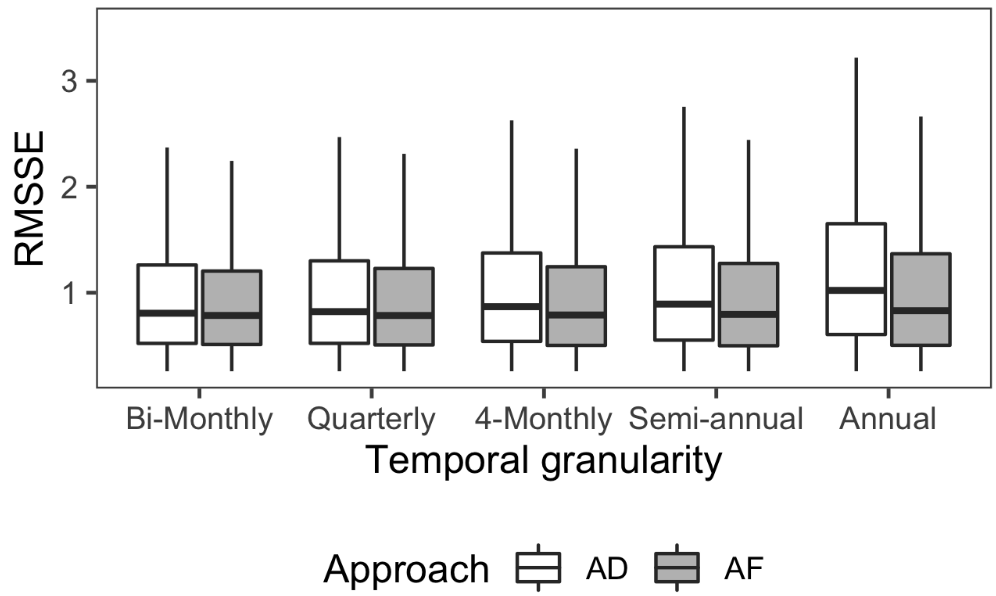
M4 Monthly time series features



Percentage of series for which each approach was more accurate (using MASE)



Performance of AF vs. AD (based on non-overlapping temporal aggregation)



Questions

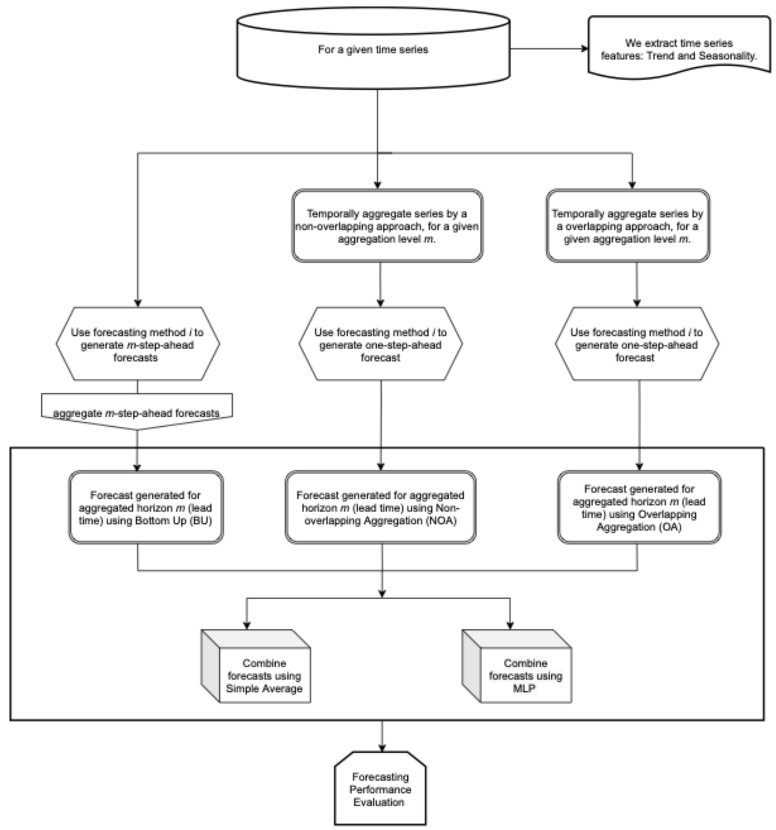
Given the comparative performance of temporal aggregation approaches :

- Whether combining forecasts generated by Bottom-Up (BU), Non-overlapping (NOA) and Overlapping approaches (OA) improves the forecast accuracy? how to combine?
- How data temporal aggregation changes time series features and is there any association between time series features and the forecasting performance of AD versus AF?

Outline

- Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?
- How does temporal aggregation approaches perfrom on M4 competition data?
- Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (**Working paper 1**)?
- How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF (**Working paper 2**)?

Experiment design - 1



Combining algorithm

Algorithm 1 MLP combining rule

Initialization:

- set the vector of learning rates of individual approaches $(\eta_0^{BU}, \eta_0^{NOA}, \eta_0^{OA})$
- set the vector of regrets of individual approaches $(R_0^{BU}, R_0^{NOA}, R_0^{OA}) = (0, 0, 0)$

repeat

At each time (aggregate horizon) in the out-of-sample

1. compute the learning rates η_{t-1}^k according to Equation (1)
2. calculate the combining weights of each individual method by

$$p_t^k = \frac{\eta_{t-1}^k \max(0, R_{t-1}^k)}{\sum_{k=1}^K \eta_{t-1}^k \max(0, R_{t-1}^k)}$$

3. obtain the loss vector $\boldsymbol{\ell}_t = (\ell_t^{BU}, \ell_t^{NOA}, \ell_t^{OA})$ and the weighted loss $\hat{\ell}_t = p_t^{BU}\ell_t^{BU} + p_t^{NOA}\ell_t^{NOA} + p_t^{OA}\ell_t^{OA}$
4. update the regret $R_t^k = R_{t-1}^k + (\hat{\ell}_t - \ell_t^k)$

until *End of the out-of-sample*;

Cesa-Bianchi, Nicolo, and Gábor Lugosi. "Potential-based algorithms in on-line prediction and game theory." Machine Learning 51.3 (2003): 239-261.

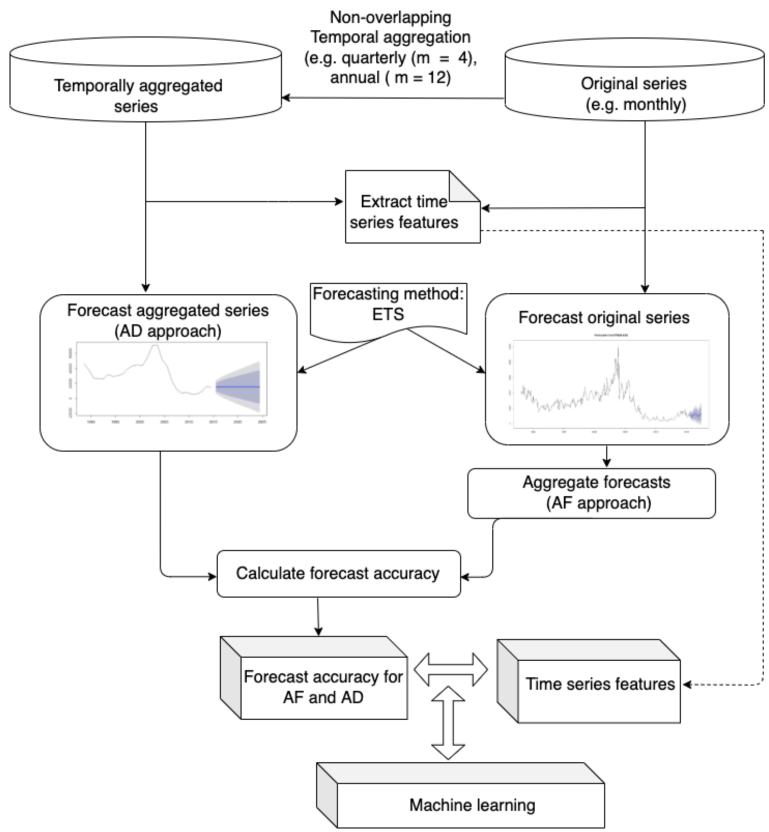
Mean (median) MASE for M4 monthly series with ETS forecasting method

Aggregation level	Pattern	Approach				
		MLP	Average	Overlapping	Non-overlapping	BU
Annual	(N, N)	**6.675** (**4.456**)	7.958 (5.38)	8.491 (5.723)	9.822 (6.939)	7.673 (5.590)
	(T, N)	**6.483** (**4.299**)	8.785 (5.834)	11.161 (7.643)	11.472 (7.66)	7.603 (5.495)
	(N, S)	**6.826** (**4.760**)	7.883 (5.579)	8.288 (5.857)	9.191 (6.585)	8.119 (6.252)
	(T, S)	**6.215** (**4.394**)	8.064 (5.957)	10.23 (7.96)	9.417 (6.820)	7.396 (5.671)
Semi-annual	(N, N)	**3.023** (**2.500**)	3.236 (2.635)	3.398 (2.800)	3.642 (2.871)	3.170 (2.637)
	(T, N)	**2.440** (**1.809**)	2.893 (2.153)	3.502 (2.640)	3.511 (2.607)	2.714 (2.045)
	(N, S)	**3.421** (**2.803**)	3.875 (3.167)	5.161 (3.992)	4.025 (3.315)	3.82 (3.156)
	(T, S)	**2.835** (**2.290**)	3.366 (2.695)	5.105 (3.833)	3.442 (2.768)	3.156 (2.548)
4-monthly	(N, N)	**1.886** (**1.620**)	1.951 (1.668)	2.041 (1.751)	2.155 (1.819)	1.916 (1.635)
	(T, N)	**1.430** (**1.063**)	1.574 (1.181)	1.845 (1.421)	1.837 (1.420)	1.499 (1.115))
	(N, S)	**2.288** (**1.937**)	2.538 (2.115)	3.501 (2.752)	2.551 (2.146)	2.457 (2.056)
	(T, S)	**1.824** (**1.497**)	2.121 (1.693)	3.330 (2.354)	2.057 (1.691)	1.948 (1.578)
Quarterly	(N, N)	**1.358** (1.179)	1.382 (1.198)	1.446 (1.259)	1.496 (1.292)	1.366 (**1.167**)
	(T, N)	**0.998** (**0.731**)	1.054 (0.772)	1.198 (0.909)	1.195 (0.917)	1.012 (0.733)
	(N, S)	**1.722** (**1.471**)	1.862 (1.567)	2.515 (1.986)	1.854 (1.553)	1.807 (1.529)
	(T, S)	**1.338** (**1.093**)	1.524 (1.198)	2.372 (1.644)	1.448 (1.181)	1.394 (1.121)
Bi-monthly	(N, N)	0.881 (0.758)	0.886 (0.764)	0.931 (0.799)	0.926 (0.799)	**0.877** (**0.745**)
	(T, N)	0.615 (0.423)	0.627 (0.429)	0.684 (0.470)	0.679 (0.493)	**0.608** (**0.412**)
	(N, S)	**1.191** (**1.030**)	1.218 (1.050)	1.504 (1.241)	1.223 (1.043)	1.208 (1.031)
	(T, S)	**0.895** (0.717)	0.951 (0.745)	1.329 (0.951)	0.919 (0.737)	0.900 (**0.713**)

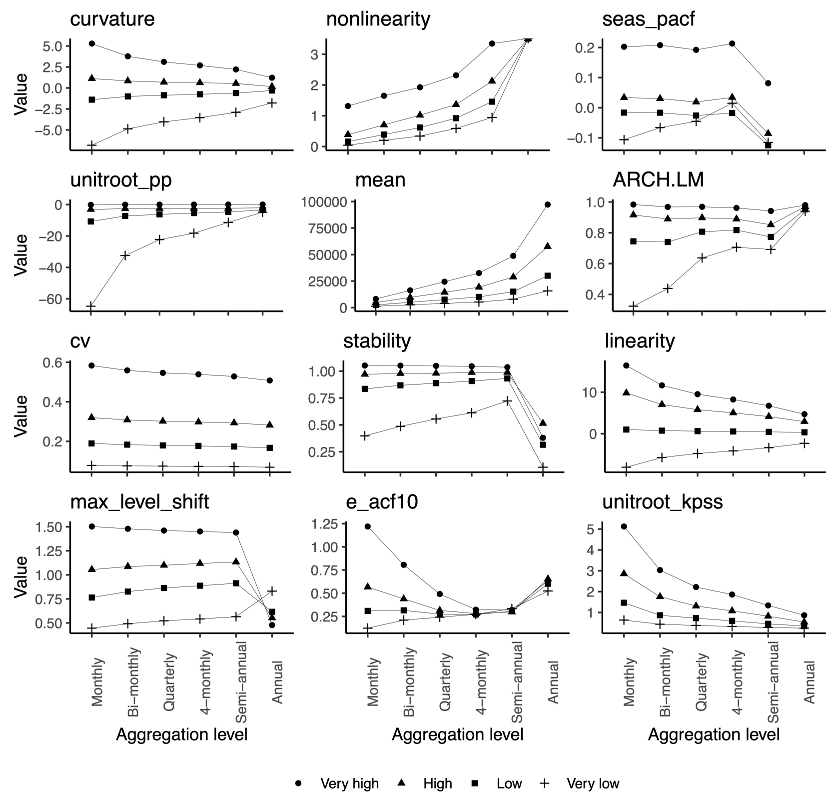
Outline

- Temporal aggregation: why do we need it in time series forecasting and what are the common approaches?
- How does temporal aggregation approaches perfrom on M4 competition data?
- Whether combining forecasts generated by temporal aggregation improves the forecast accuracy? how to combine (**Working paper 1**)?
- How data temporal aggregation changes time series features and how might time series features affect the forecasting performance of AD versus AF (**Working paper 2**)?

Experiment design - 2

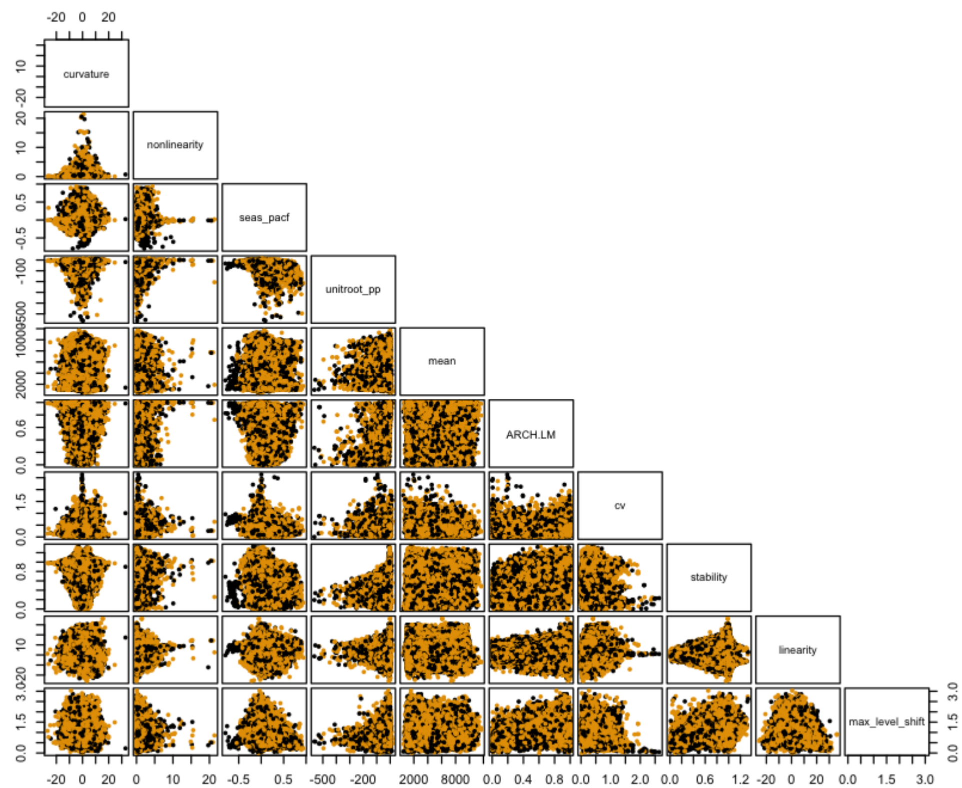


How does non-overlapping TA change time series features?

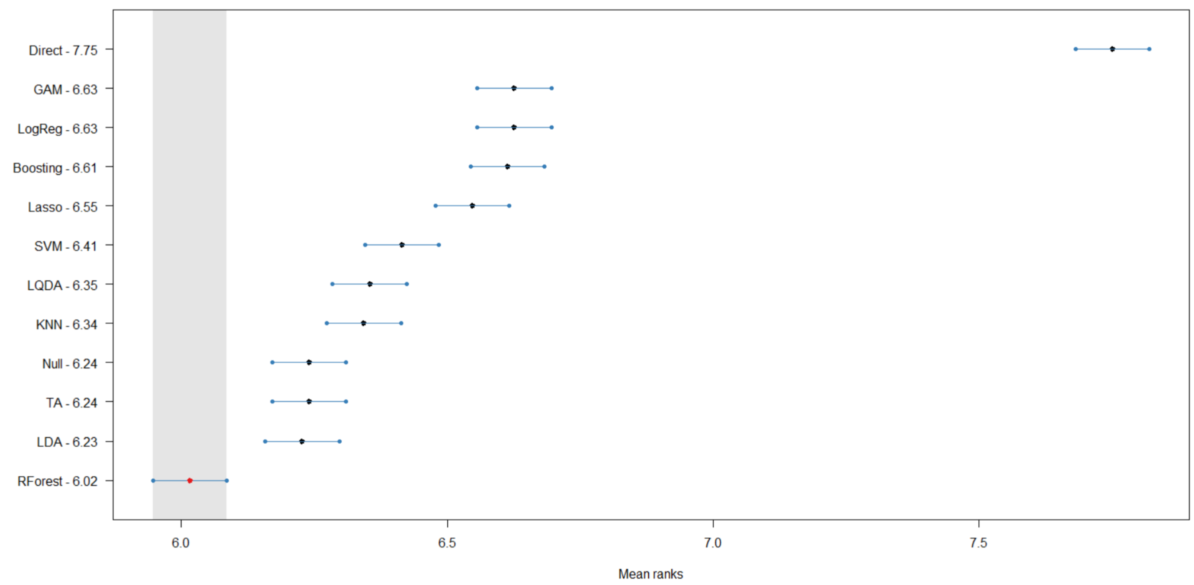


How does non-overlapping TA change time series features (continue)?

Features relationship and AD/AF performance

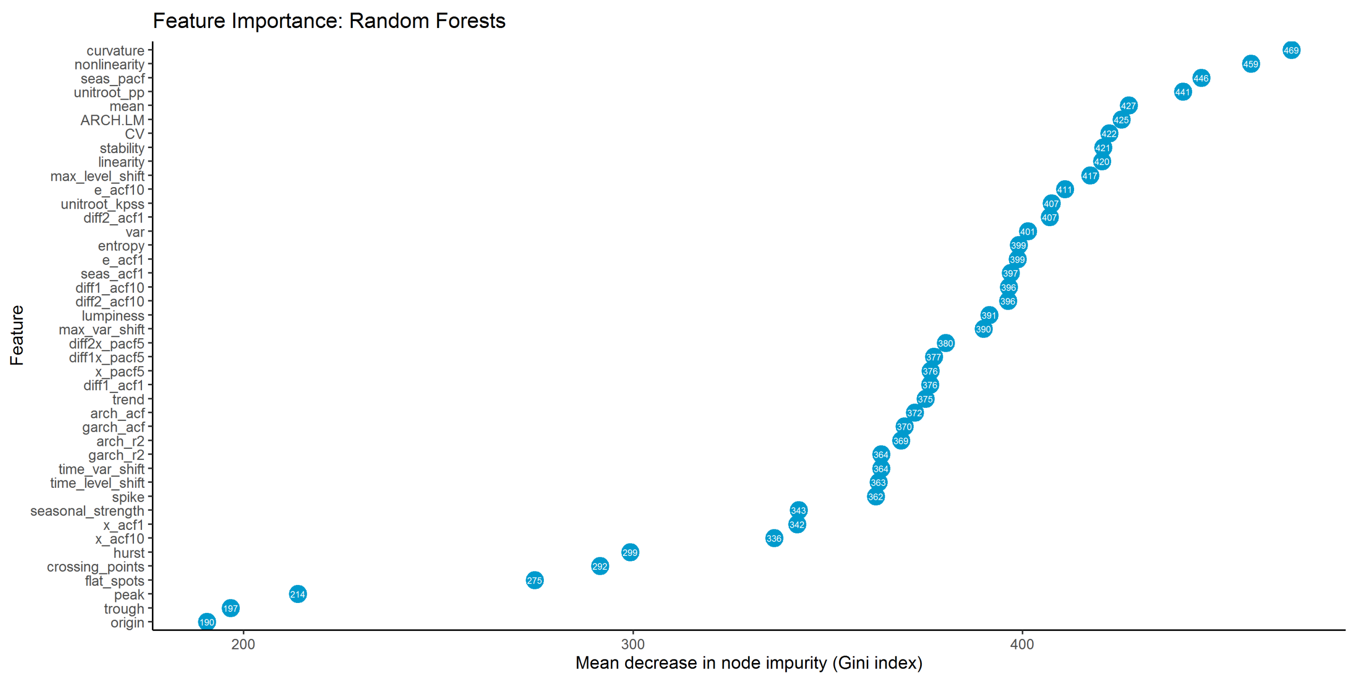


MCB test for all classiefiers



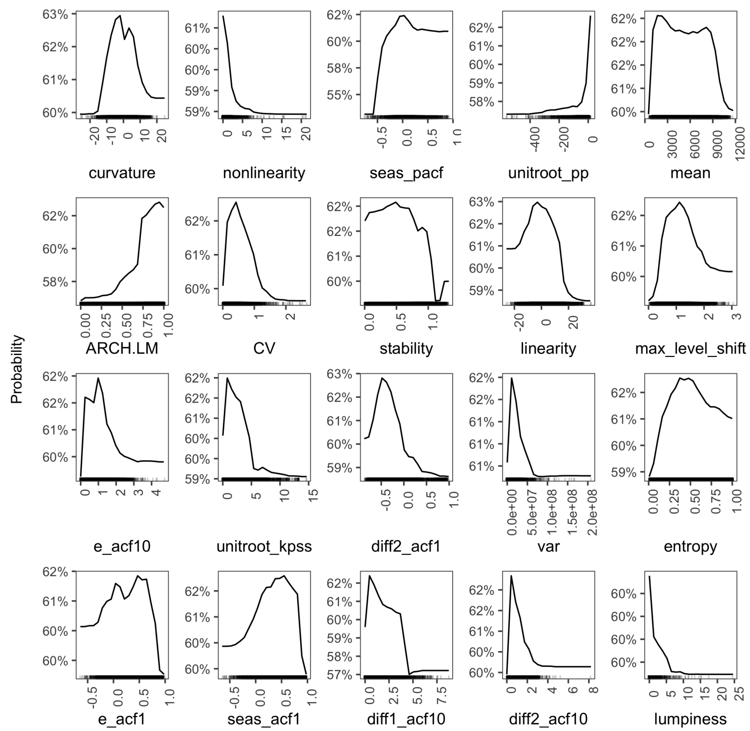
We also use missclassification error, F-statistics and Area under the Curve(AUC).

Important features



Partial dependence plot

Probability of AF performing better



Partial dependence plot (continue)

Probability of AF performing better

Summary and conclusions (continue)

- Although aggregating time series seems to be intutive, it might not always improve forecast accuracy. Our results indicate that Aggregate Forecast is a competitive approach, but neither of them dominate. They both have a merit.
- Combining aggregate data (non-overlapping and overlapping) and aggregate forecast approaches improve forecast accuracy. Combination again works here.
- Aggregate data using temporal aggregation changes the features of time series. The magnitude of the change varies for different features. In particular, we observe that with increase in the aggregation level, the strength of seasonality, the autocorrelation, coefficient of variation, linearity, curvature and KPSS unitroot statistic decrease. However, non-linearity, mean, variance, ARCH.LM, trend , unitroot pp statistics increase. Entropy is the only measure that both increases and decreases based on its initial value.

Summary and conclusions

- Random Forest model is the most accurate classifier among ML algorithm in predicting which approach provides more accurate forecast given a set of time series features as input.
- The most important features for predicting whether AF or AD should be used for a given monthly time series in M4 competition include *curvature*, *nonlinearity*, *seas_pacf*, *unitroot_up*, *mean*, *ARCHM.LM*, *Coifficient of Variation*, *stability*, *linearity* and *max_level_shift*.
- Increasing trend, ARCH.LM, hurst, autocorrelation lag 1 and unitroot_pp and seas_pacf may increases the chance of AF performing better.
- Increasing lumpiness, entropy, no-linearity, curvature, stremgth of seasonality may increase the chance of AD performing better, so the strong presence of these features may favorite AD over AF.

Wrok in progress

- Rostami-Tabar B., Goltsos T. Wang, S. (2022), Forecasting for lead-time period by temporal aggregation: Whether to combine and how
- Rostami-Tabar B., Mercetic D. (2022), On time series features and the perfromance of emporal aggregation

Published recently

- Mircetic, D., et al. (2021), "Forecasting hierarchical time series in supply chains: an empirical investigation." International Journal of Production Research, 1-20.
- Babai. M.Z., Boylan, J., Rostami-Tabar, B. (2021), "Demand Forecasting in Supply Chains: A Review of Aggregation and Hierarchical Approaches", International Journal of Production Research, 1-25.

References for temporal aggregation forecasting

- An aggregate–disaggregate intermittent demand approach (ADIDA) to forecasting: an empirical proposition and analysis. Journal of the Operational Research Society.
- Improving forecasting via multiple temporal aggregation. International Journal of Forecasting.
- Demand forecasting by temporal aggregation, Naval Research Logistics
- Forecasting with temporal hierarchies, European Journal of Operational Research

- Slides and papers: www.bahmanrt.com
- Check out also www.f4sg.org

Say hello: @Bahman_R_T

Connect: Bahman ROSTAMI-TABAR