Déjà Vu: Forecasting with Similarity

Feng Li

School of Statistics and Mathematics
Central University of Finance and Economics

Collaborators

		
Yanfei Kang (Beihang)	Fotios Petropoulos (Bath Uni.)	Evangelos Spiliotis (NTUA)

- Feng Li and Yanfei's research are supported by the National Natural Science Foundation of China.

Outline

1. **Forecasting model selection: to be or not to be?**

2. **Forecasting with similarity**

3. **Similarity, forecastability and forecasting uncertainty**

“All models are wrong, but some are useful.”– George Box

- Three sources of uncertainty exist in forecasting: **model**, **parameter**, and **data**.
    - Merely tackling the model uncertainty is sufficient to bring most of the performance benefits (Petropoulos et al., 2018)
- “All models are wrong, but **some** are useful.”
    - Researchers increasingly avoid using a single model, and opt for combinations of forecasts from multiple models (Jose and Winkler, 2008; Kolassa, 2011; Bergmeir et al., 2016; Montero-Manso et al., 2019).


GAME OVER
CONTINUE? 09
MemeCenter.com

Déjà Vu: Forecasting with Similarity

- We argue that there is another way to avoid selecting a single model: **to select no models at all**.
- We provide a new way to forecasting that does not require the estimation of any forecasting models, while also exploiting the benefits of cross-learning (Makridakis et al., 2019)

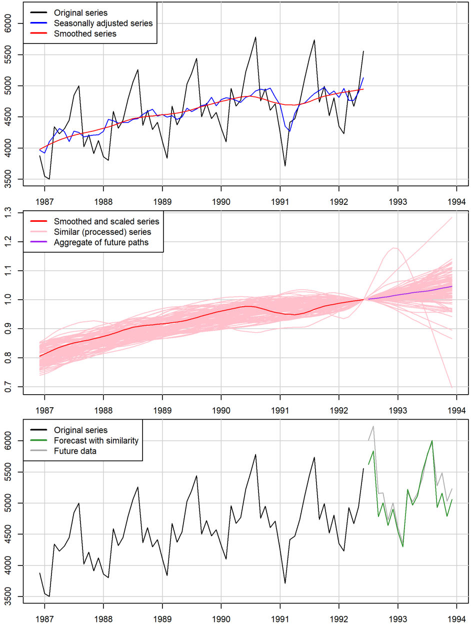
The idea for déjà vu

1. A target series is compared against a set of reference series attempting to identify similar ones (déjà vu).
2. The point forecasts for the target series are the average of the future paths of the most similar reference series.
3. The prediction intervals are based on the distribution of the reference series, calibrated for low sampling variability. Note that no model extrapolations take place in our approach.
4. The proposed approach has several advantages compared to existing methods, namely
   - it tackles both model and parameter uncertainties
   - it does not use time series features or other statistics as a proxy for determining similarity, and
   - no explicit assumptions are made about the DGP as well as the distribution of the forecast errors.

Methodology

- The objective of "forecasting with similarity" is to find the most similar ones to a target series, average their future paths, and use this average as the forecasts for the target series.
  1. **Removing seasonality**, if a series is identified as seasonal.
  2. **Smoothing** by estimating the trend component through time series decomposition.
  3. **Scaling** to render the target and possible similar series comparable.
  4. **Measuring similarity** by using a set of distance measures.
  5. **Forecasting** by aggregating the paths of the most similar series.
  6. **Inverse scaling** to bring the forecasts for the target series back to its original scale.
  7. **Recovering seasonality**, if the target series is found seasonal in Step 1.
- We use the yearly, quarterly, and monthly subsets of the M4 competition (Makridakis et al., 2019), which consist of 23000, 24000, and 48000 series, respectively.

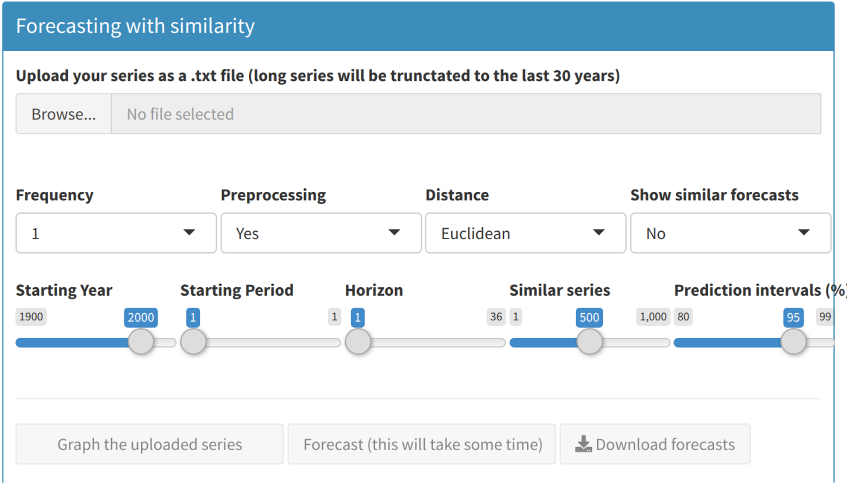
Toy example



Original series
Seasonally adjusted series
Smoothed series
Smoothed and scaled series
Similar (processed) series
Aggregate of future paths
Original series
Forecast with similarity
Future data
1987
1988
1989
1990
1991
1992
1993
1994

Online APP

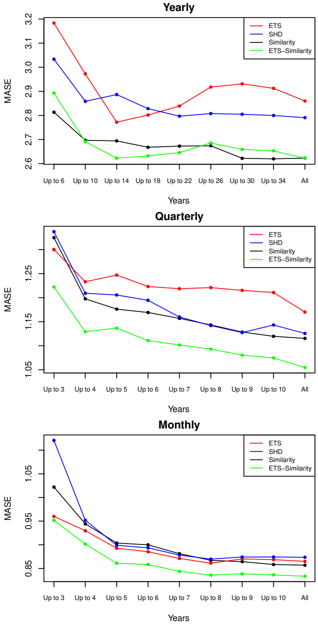
https://fotpetr.shinyapps.io/similarity/



Performance of *Similarity* for different distance measures and pool sizes of similar reference series

Frequency	Distance Measure	Number of aggregated reference series (k)						
		1	5	10	50	100	500	1000
Yearly	$\mathcal{L}_1$	3.289	2.837	2.787	2.689	2.668	2.632	2.634
	$\mathcal{L}_2$	3.333	2.866	2.785	2.703	2.684	2.638	2.639
	DTW	3.270	2.835	2.730	2.656	2.641	**2.623**	2.637
Quarterly	$\mathcal{L}_1$	1.312	1.205	1.175	1.136	1.135	1.127	1.126
	$\mathcal{L}_2$	1.336	1.199	1.162	1.138	1.134	1.126	1.127
	DTW	1.293	1.177	1.158	1.117	1.115	**1.115**	1.116
Monthly	$\mathcal{L}_1$	1.004	0.908	0.887	0.871	0.870	0.867	0.869
	$\mathcal{L}_2$	1.008	0.910	0.891	0.871	0.869	0.866	0.868
	DTW	1.001	0.895	0.875	0.861	0.861	**0.857**	0.857
Total	$\mathcal{L}_1$	1.607	1.427	1.397	1.356	1.351	1.339	1.340
	$\mathcal{L}_2$	1.626	1.433	1.395	1.360	1.354	1.339	1.341
	DTW	1.597	1.413	1.373	1.339	1.335	**1.329**	1.332

Performance of *Similarity* against model-based approach



Similarity, forecastability and forecasting uncertainty I

- Time series forecasting uncertainty is usually quantified by prediction intervals, which depend on the forecastability of the target time series.
    - With a model-based forecasting approach, although one could usually obtain a theoretical prediction interval, the performance of such interval depends upon the length of series, accuracy of the model, and variability of model parameters.
- We use the variability information from the rescaled and reseasonalised reference series, $\check{Q}_t$, as the source of prediction interval bounds.
    - Directly using the quantiles or variance of reference series may lead to lower-than-nominal coverage due to the similarity (or low sampling variability) of reference series.
- We propose a straightforward data-driven approach, in which the $(1-\alpha)100\%$ prediction interval for a forecast f_t is based on the a calibrated $\alpha/2$ and $1-\alpha/2$ quantiles of the selected reference series $\check{Q}_t$ for the target y_t.

Similarity, forecastability and forecasting uncertainty II

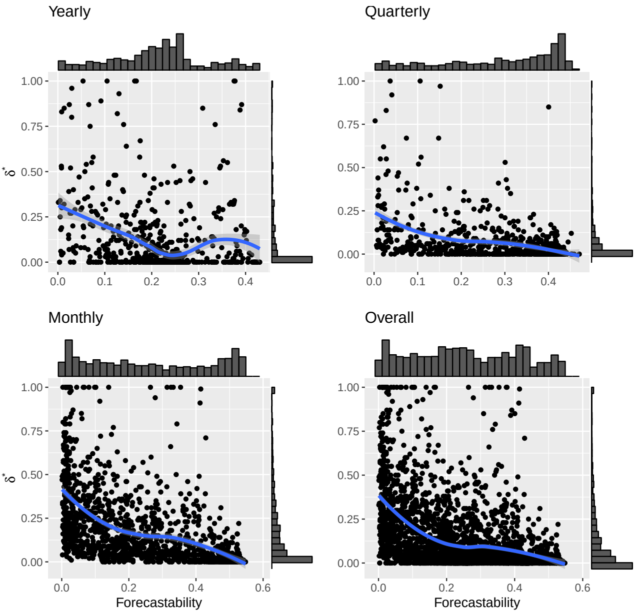
- The lower and upper bounds for the prediction interval are defined as

$$L_t = (1-\delta)\ F^{-1}_{\check{Q}_t}(\alpha/2) \text{ and } U_t = (1+\delta)\ F^{-1}_{\check{Q}_t}(1-\alpha/2), \tag{1}$$

  respectively, where $F^{-1}_{\check{Q}_t}$ is the quantile based on the selected reference series $\check{Q}_t$, and δ is a calibrating factor.
- We follow Kang et al. (2017) and use the spectral entropy to measure the "forecastability" of a time series as

$$\text{Forecastability} = 1 + \int_{-\pi}^{\pi} \hat{f}_y(\gamma) \log \hat{f}_y(\gamma) \mathrm{d}\gamma,$$

Similarity, forecastability and forecasting uncertainty



Performance of *Similarity* in terms of prediction intervals

	MSIS	Coverage (%) Target: 95%	Upper coverage (%) Target: 97.5%	Spread
			Yearly	
ETS	30.616	84.341	89.664	12.346
SHD	35.488	80.439	86.744	**8.782**
Similarity	23.182	88.372	**95.065**	13.591
ETS-Similarity	**22.437**	**90.904**	94.677	12.968
			Quarterly	
ETS	10.717	87.153	92.659	4.688
SHD	11.027	87.219	91.981	**4.398**
Similarity	10.556	87.087	95.635	5.213
ETS-Similarity	**9.240**	**91.402**	**96.114**	4.950
			Monthly	
ETS	6.342	92.032	94.293	4.094
SHD	6.885	90.799	93.600	**4.039**
Similarity	6.666	91.068	96.152	4.563
ETS-Similarity	**5.765**	**94.141**	**96.678**	4.328

Conclusions

- The advantages of our proposition are that it is model-free, in the sense that it does not rely on statistical forecasting models, and, as a result, it does not assume an explicit DGP.
- Instead, we argue that history repeats itself (déjà vu) and that the current data patterns will resemble the patterns of other already observed series.
- The proposed approach is data-centric and relies on the availability of a rich, representative reference set of series – a not so unreasonable requirement in the era of big data.
- Incorporating our GRATIS paper to have a diverse reference set.

References

Bergmeir, C., Hyndman, R. J., and Benítez, J. M. (2016), "Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation," *International Journal of Forecasting*, 32, 303–312.

Jose, V. R. R. and Winkler, R. L. (2008), "Simple robust averages of forecasts: Some empirical results," *International Journal of Forecasting*, 24, 163–169.

Kang, Y., Hyndman, R. J., and Smith-Miles, K. (2017), "Visualising forecasting algorithm performance using time series instance spaces," *International Journal of Forecasting*, 33, 345–358.

Kolassa, S. (2011), "Combining exponential smoothing forecasts using Akaike weights," *International Journal of Forecasting*, 27, 238–251.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2019), "The M4 Competition: 100,000 time series and 61 forecasting methods," *International Journal of Forecasting*, In Press.

Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., and Talagala, T. S. (2019), "FFORMA: Feature-based Forecast Model Averaging," *Monash Econometrics and Business Statistics Working Papers*, 19.

Petropoulos, F., Hyndman, R. J., and Bergmeir, C. (2018), "Exploring the sources of uncertainty: Why does bagging for time series forecasting work?" *European Journal of Operational Research*, 268, 545–554.

Thank you!

feng.li@cufe.edu.cn

https://feng.li/