Forecast Linear Augmented Projection with Targeted Components

Yangzhuoran Fin Yang

Department of Quantitative Economics, Maastricht University, The Netherlands
Yangzhuoran.Yang@maastrichtuniversity.nl
https://yangzhuoranyang.com

Abstract. Forecast Linear Augmented Projection (FLAP) is a post-forecast adjustment method that can reduce forecast error variance in multivariate time series. In FLAP, components containing information about shared features are constructed as linear combinations of the original time series. The forecasts of the original time series and the components are then projected such that the linear relationship between the historical data is enforced on the forecasts. While forecast error variance reduction has been theoretically proven regardless of the linear combination, the empirical performance of different component types is less clear and is examined in this paper. Components considered in this paper are estimated by maximising measures of information and/or by minimising the dependency between components. Among other methods, using FLAP with Principal Component Analysis is recommended for its stable performance across settings, while Forecastable Component Analysis offers a strong alternative, as demonstrated by simulations and application to Australian tourism data.

Keywords: Forecast combinations, High-dimensional time series, Components, Forecast reconciliation

1 Introduction

Yang et al. [2024] show that Forecast Linear Augmented Projection (FLAP) can reduce the forecast error variance in multivariate time series forecasts. FLAP is a post-processing framework that utilises information shared between different series—information often overlooked by standard forecasting models. The process involves: (1) constructing components as linear combinations of the series to capture underlying signals; (2) generating forecasts for both the original series and the components using arbitrary forecasting method; and (3) applying a projection to ensure that the constraints defined by the linear combinations between the components and the series are preserved on the forecasts. This approach provides a systematic method for improving forecast accuracy by incorporating interdependencies between series that may otherwise be neglected.

The forecast error variance reduction property has been shown to be theoretically agnostic to the choice of components. Yang et al. [2024] support this with two key results: (1) irrespective of the weights used in the linear combinations, the forecast error variance for each series is non-increasing as the number of components increases (including from 0 to 1); and (2) the condition under which no reduction occurs is theoretically possible but rarely encountered in practice, and even if it does arise, the effect is likely to be overshadowed by estimation error. While any reasonable choice of components can improve forecast accuracy in theory, the degree of improvement can vary in practice. Because the optimal component weights are difficult to determine analytically, the authors propose using principal component analysis (PCA, Jolliffe [2002]), which they show performs better than randomly generated linear combinations on both simulated and real data. Nevertheless, it remains unclear how to best construct components in applied settings, as theoretical guidance is limited. This paper addresses this issue by evaluating and comparing the performance of various components in the FLAP framework, using simulations and Australian tourism data.

Component estimation has been regarded as an effective approach to dimension reduction, aiming to uncover simplified structures in multivariate time series modelling. Although FLAP does not require the number of components to be small, it can benefit from the way components summarise information about the data-generating process. In this paper, I focus on component estimation methods that reduce dependency between components, with some also enhancing features relevant to forecasting. The first group includes principal component analysis for stationary vector time series (TS-PCA, Chang et al. [2018]) and independent component analysis (ICA, Bell and Sejnowski [1995]). The second group includes forecastable component analysis (ForeCA, Goerg [2013]) and canonical components (CC, Box and Tiao [1977]). When dependence between components is reduced, forecasting models can be estimated more efficiently within a lower-dimensional parameter space, making the use of univariate models more appealing. Likewise, components that highlight forecastable signals may improve performance by directing the model's focus toward patterns with greater predictive value.

2 Forecast Linear Augmented Projection (FLAP)

In this section, I briefly review some key results related to FLAP. For a more detailed discussion, see Yang et al. [2024]. The implementation of FLAP is provided by the `flap` package [Yang, 2024] in R [R Core Team, 2024]. Let $\boldsymbol{y}_t \in \mathbb{R}^m$ be a vector of m observed time series to be forecasted. Let $\boldsymbol{I}_n$ to be a $n \times n$ identity matrix. The FLAP method has three steps:

1. Construct components $\boldsymbol{c}_t = \boldsymbol{\Phi}\boldsymbol{y}_t \in \mathbb{R}^p$, a vector of p linear combinations of $\boldsymbol{y}_t$. The component weights $\boldsymbol{\Phi} \in \mathbb{R}^{p\times m}$ are specific to the method of the component construction.
2. Generate forecasts of the series and the components. Define $\boldsymbol{z}_t$ as the stacked vector of the series $\boldsymbol{y}_t$ and the components $\boldsymbol{c}_t$, i.e., $\boldsymbol{z}_t = \left[\boldsymbol{y}_t', \boldsymbol{c}_t'\right]'$. The h-step-ahead base forecast of $\boldsymbol{z}_t$ are generated from arbitrary methods and denoted as $\hat{\boldsymbol{z}}_{t+h}$.
3. Impose the linear constraint through projection. The linear constraints $\boldsymbol{C}\boldsymbol{z}_t = \boldsymbol{c}_t - \boldsymbol{\Phi}\boldsymbol{y}_t = \boldsymbol{0}$ where $\boldsymbol{C} = \begin{bmatrix} -\boldsymbol{\Phi} & \boldsymbol{I}_p \end{bmatrix}$ hold for $\boldsymbol{z}_t$ but not necessary for the

forecast $\hat{\boldsymbol{z}}_t$. The base forecasts are projected to produced the FLAP forecast $\tilde{\boldsymbol{z}}_{t+h}$ such that $\tilde{\boldsymbol{z}}_{t+h} = \boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}$ with projection matrix

$$\boldsymbol{M} = \boldsymbol{I}_{m+p} - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}, \tag{1}$$

where $\mathrm{Var}(\boldsymbol{z}_{t+h} - \hat{\boldsymbol{z}}_{t+h}) = \boldsymbol{W}_h$ is the forecast error covariance matrix.

To extract the FLAP forecast of $\boldsymbol{y}_t$ and not the component, define selection matrix $\boldsymbol{J}_{m,p} = \begin{bmatrix}\boldsymbol{I}_m & \boldsymbol{O}_{m\times p}\end{bmatrix}$, so the projected forecast of $\boldsymbol{y}_t$ can be selected by $\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{J}\tilde{\boldsymbol{z}}_{t+h} = \boldsymbol{J}\boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}$. Yang et al. [2024] have shown that the difference between the forecast error variances of the base and FLAP forecasts, $\mathrm{Var}(\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}) - \mathrm{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h})$ is positive semi-definite. In addition, it has non-decreasing diagonal elements as p increases, which means FLAP does not make forecasts worse. This provides the theoretical justification for FLAP.

In practice, the key consideration is constructing the projection matrix $\boldsymbol{M}$ in Equation 1. I adopt a shrinkage estimator for $\boldsymbol{W}_h$ following Yang et al. [2024]. The constraint matrix $\boldsymbol{C}$ is determined by the choice of component construction method and is the focus of this paper. Yang et al. [2024] demonstrate that for a new component to be beneficial—beyond those already included in FLAP—the information it introduces, as reflected in the error covariance, must not be a linear combination of the information already captured by existing components. Heuristically, each component should bring in new information. This is more likely when components are uncorrelated, which motivates the use of principal components in Yang et al. [2024], and more generally, the consideration of methods in this paper that reduce cross-dependencies among components. Building on this same intuition, I also consider components constructed to maximise properties that are directly relevant for forecasting, so that the new information is not only distinct, but also more strongly expressed. The next section introduces the component construction methods examined in this paper.

3 Components

Components obtained from PCA, ForeCA, and CC analysis are uncorrelated with each other, representing a basic level of independence. Due to space constraints, I do not provide a detailed introduction to PCA in this section, but note that it constructs components by maximising variances of the components. TS-PCA extends this idea by producing groups of components with reduced cross-sectional and temporal correlations between different groups. ICA goes a step further by constructing statistically independent components, representing the strongest form of independence considered in this paper.

Forecastable Component Analysis (ForeCA) The idea of ForeCA is to find linear combinations such that the signal in the components is most significant. This is achieved by minimising the entropy of the spectral density of the time series, or maximising the forecastability defined as $\Omega\left(c_t\right) = 1 - \frac{H_{s,a}(c_t)}{\log_a(2\pi)}$, where $H_{s,a}\left(c_t\right)$ is the Shannon entropy [Shannon, 1948] of the spectral density $f_c(\xi)$ of component c_t. At minimum entropy, the spectral density of the components concentrates around a small number of frequencies. As a result, components with high forecastability tend to exhibit near sine wave-like patterns, making them particularly desirable for forecasting. ForeCA components are estimated using the `ForeCA` package [Goerg, 2020].

Canonical Component (CC) CCs are obtained by fitting a Vector AutoRegressive (VAR) model on $\boldsymbol{y}_t$ first, then the weight matrix $\boldsymbol{\Phi}$ is generated as the eigenvectors corresponding to the largest magnitude eigenvalues of $\Sigma^{-1}(\boldsymbol{y}_t)\Sigma(\hat{\boldsymbol{y}}_t)$, where $\Sigma(\boldsymbol{y}_t)$ and $\Sigma(\hat{\boldsymbol{y}}_t)$ are the covariance matrix of the data $\boldsymbol{y}_t$ and the predicted value $\hat{\boldsymbol{y}}_t$ from the VAR model respectively. This procedure can be interpreted as maximising the predictability of the components, defined as the ratio of explained variance to total variance in an AR process of the component. The VAR lag is set to be 3 in the simulation, matching the true specification, and 1 in the tourism data application, due to sample size limit.

PCA for second-order stationary vector time series (TS-PCA) Chang et al. [2018] propose finding linear combinations of the time series such that the resulting components can be segmented into lower-dimensional subseries that are uncorrelated both contemporaneously and serially. This leads to a block-diagonal structure in the autocovariance matrices of the transformed series: components within the same block may be correlated, but components across different blocks are uncorrelated. This is achieved by computing the eigenvectors of the estimator of the sum of cross-products of the autocovariance matrices across lags, $\sum_{k=0}^{k_0} \operatorname{Cov}(\boldsymbol{y}_{t+k}, \boldsymbol{y}_t) \operatorname{Cov}(\boldsymbol{y}_{t+k}, \boldsymbol{y}_t)'$, where the eigenvectors are then used as the weights in the linear combinations. TS-PCA components are estimated using the `PCA4TS` package [Chang et al., 2015].

Independent Component Analysis (ICA) ICA assumes that the process $\boldsymbol{y}_t$ consists of statistically independent latent components $\boldsymbol{c}_t$, mixed through a linear transformation. It aims to recover these components by "demixing" the observed process. ICA has been extensively studied; in this paper, I adopt the information maximisation approach proposed by Bell and Sejnowski [1995]. The ICA components are estimated using the `ica` package [Helwig, 2022].

4 Applications and Results

The Australian Tourism Data Set from Tourism Research Australia contains the total number of nights spent by Australians away from home. The monthly visitor nights are recorded for $m = 77$ regions from January 1998 to December 2019. An expanding window time series cross-validation is performed with a step size of 1 and initial set of 156 observations. Base forecasts are generated from univariate ExponenTial Smoothing (ETS) models.

In the simulation, I generate time series of length $T = 400$ from a $m = 70$ variable VAR(3) data generating process (DGP) and the performance of FLAP with different components are evaluated. This process is repeated 220 times. The coefficients for the VAR DGP are estimated from the first 70 series in the Australian tourism data set. The innovations are simulated from a multivariate normal distribution with an identity covariance matrix. The estimation and simulation is performed using the `tsDyn` package [Fabio Di Narzo et al., 2009]. The base forecasts are generated from the univariate ARIMA model and the dynamic factor model (DFM) following Stock and Watson [2002].

Since at most m components of a given type can be extracted from an m-dimensional series, I supplement with components whose weights are randomly generated from a normal distribution once the original components are exhausted. For completeness, I also include a comparison with these random components from the outset.

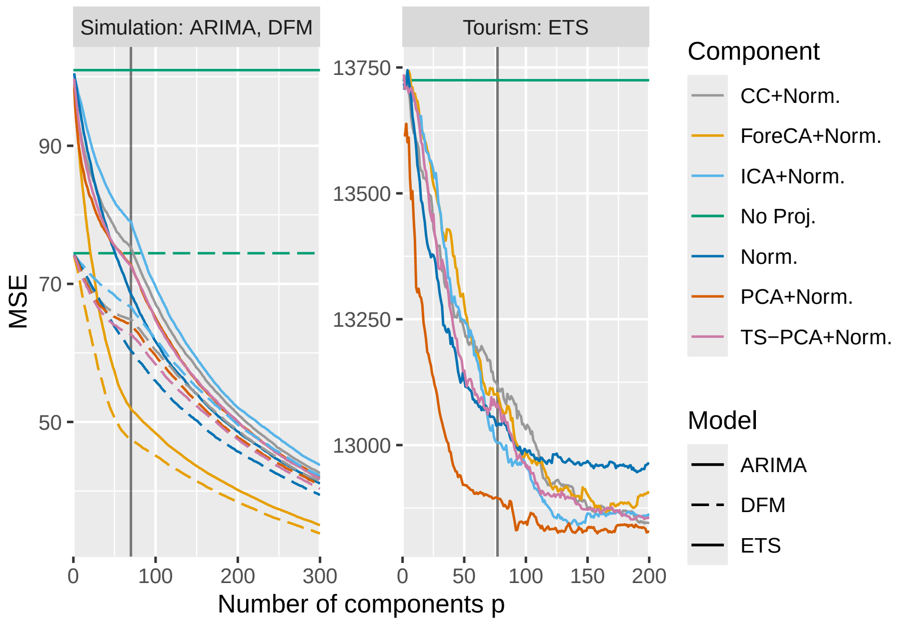

Figure 1: Out-of-sample MSE for base and FLAP forecasts.

The mean squared error (MSE) results are shown in **Figure 1.** As the number of components increases, FLAP consistently reduces MSE relative to the base forecasts, regardless of the component method used. In the simulation study (left panel), ForeCA achieves the largest error reduction, significantly outperforming the other methods. Notably, even when used with simple ARIMA models, ForeCA quickly surpasses the base forecasts from a dynamic factor model (DFM), thanks to its property of amplifying forecast-relevant signals. A statistical test confirms that this performance gain is significant compared to the other component methods. In contrast, in the application to Australian tourism data (right panel), ForeCA does not clearly outperform the alternatives. Instead, PCA yields the fastest MSE reduction. This may reflect challenges in real-world data, such as noisy signals, structural complexity, or violations of stationarity assumptions. Still, ForeCA does not perform poorly. It performs comparably to other methods, including random linear combinations of the series.

In conclusion, PCA is recommended as a robust and computationally efficient default choice for use with FLAP, given its stable performance across settings. ForeCA, however, remains a promising alternative when the data conditions are favourable.

Acknowledgement

I thank Daniele Girolimetto for the kind invitation to present this work at the 2025 Italian Statistical Society Conference (SIS 2025) and for the related discussion. I also thank Rob Hyndman, George Athanasopoulos, and Anastasios Panagiotelis for early conversations that contributed to the development of the idea. I acknowledge the use of OpenAI's ChatGPT-4o and ChatGPT-4o mini for typographical correction. These tools were used solely for lanfuage refinement. All intellectual contributions and final edits were made by me.

References

Yangzhuoran Fin Yang, George Athanasopoulos, Rob J Hyndman, and Anastasios Panagiotelis. Forecast Linear Augmented Projection (FLAP): A free lunch to reduce forecast error variance. *arXiv [stat.ME]*, 2024.

Ian T Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, NY, 2 edition, 2002. ISBN 9780387954424. doi: 10.1007/b98835.

Jinyuan Chang, Bin Guo, and Qiwei Yao. Principal component analysis for second-order stationary vector time series. *Ann. Stat.*, 46:2094–2124, 2018. doi: 10.1214/17-AOS1613.

Anthony J Bell and Terrence J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.*, 7(6): 1129–1159, 1995. doi: 10.1162/neco.1995.7.6.1129.

Georg Goerg. Forecastable Component Analysis. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28, pages 64–72, Atlanta, Georgia, USA, 2013.

G E P Box and G C Tiao. A canonical analysis of multiple time series. *Biometrika*, 64(2):355–365, 1977. doi: 10.1093/biomet/64.2.355.

Yangzhuoran Fin Yang. *flap: Forecast Linear Augmented Projection*, 2024. R package version 0.2.0. doi: 10.32614/CRAN.package.flap.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024.

C E Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

Georg M. Goerg. *ForeCA: An R package for Forecastable Component Analysis*, 2020. R package version 0.2.7. doi: 10.32614/CRAN.package.ForeCA.

Jinyuan Chang, Bin Guo, and Qiwei Yao. *PCA4TS: Segmenting Multiple Time Series by Contemporaneous Linear Transformation*, 2015. R package version 0.1. doi: 10.32614/CRAN.package.PCA4TS.

Nathaniel E. Helwig. *ica: Independent Component Analysis*, 2022. R package version 1.0-3. doi: 10.32614/CRAN.package.ica.

Antonio Fabio Di Narzo, Jose Luis Aznarte, and Matthieu Stigler. *tsDyn: Time series analysis based on dynamical systems theory*, 2009. R package version 0.7. doi: 10.32614/CRAN.package.tsDyn.

James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. *J. Bus. Econ. Stat.*, 20(2):147–162, 2002. doi: 10.1198/073500102317351921.