Forecast Linear Augmented Projection with Targeted Components

Yangzhuoran Fin Yang

Forecast Linear Augmented Projection (FLAP)

A model-independent post-forecast adjustment method that can reduce forecast error variance.

- Averaging indirect forecasts from linear combinations (components)
- Projecting forecasts of augmented series
- Free lunch: no additional data or information needed

Forecast Linear Augmented Projection (FLAP)

A model-independent post-forecast adjustment method that can reduce forecast error variance.

- Averaging indirect forecasts from linear combinations (components)
- Projecting forecasts of augmented series
- Free lunch: no additional data or information needed

Question

What components?

What to expect

- Intuition with data
- Method
- Properties
- Choice of components
- Empirical applications and simulation

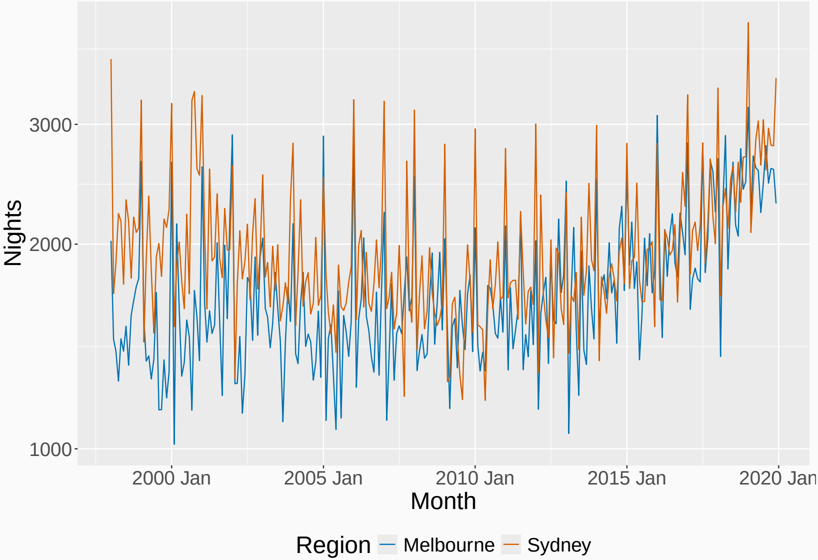
Australian tourism data

- The data include tourism information on seven states and territories which can be divided into 77 regions
  - For example, Melbourne, Sydney, East Coast

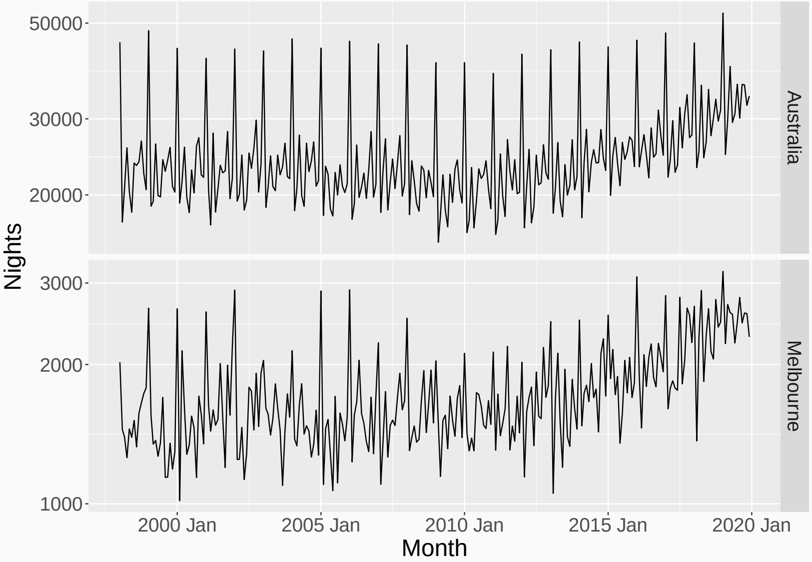
Visitor nights

The total number of nights spent by Australians away from home recorded monthly

Melbourne and Sydney



Total and Region



Intuition

Observation

1. Similar patterns are shared by different series.
2. Better signal-noise ratio in the linear combination.

Intuition

Observation

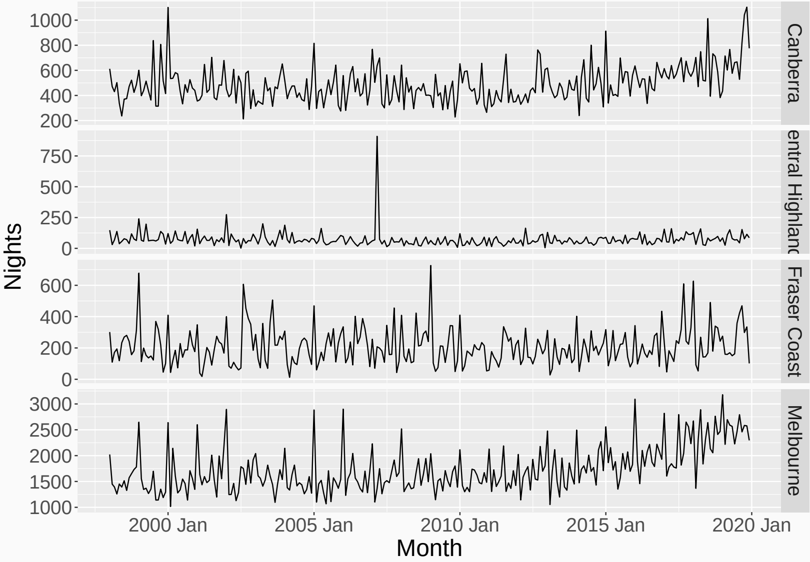
1. Similar patterns are shared by different series.
2. Better signal-noise ratio in the linear combination.

One step further

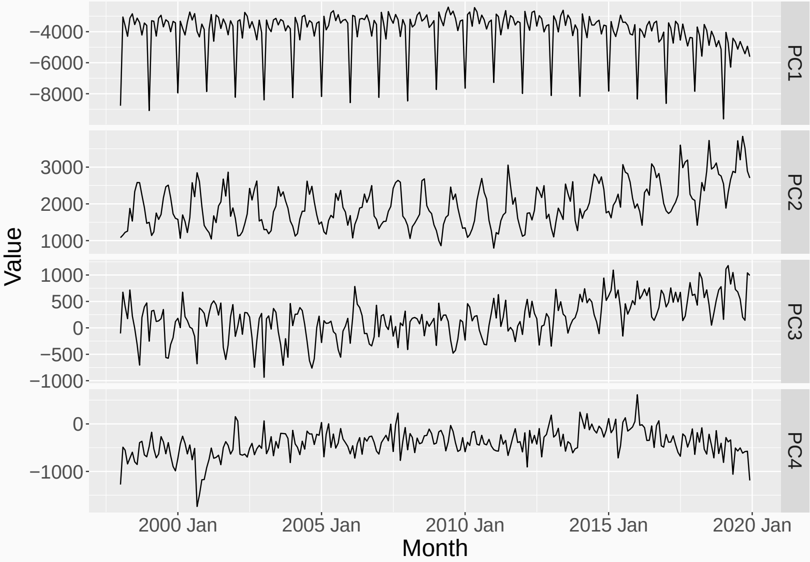
Finding components that

1. are easy to forecast;
2. can capture the common signals;
3. can improve forecast of original series.

Series $\boldsymbol{y}_t \in \mathbb{R}^m$



Components $c_t = \Phi y_t \in \mathbb{R}^p$



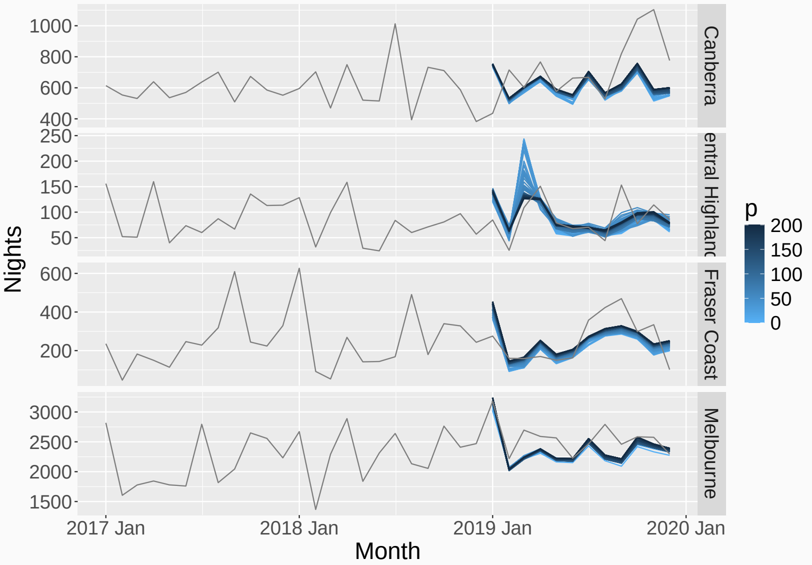
FLAP

$$\boldsymbol{z}_t = \begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{c}_t \end{bmatrix} \qquad \tilde{\boldsymbol{z}}_{t+h} = \boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}$$

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{J}\tilde{\boldsymbol{z}}_{t+h} = \boldsymbol{J}\boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}$$

$$\begin{aligned} \boldsymbol{M} &= \boldsymbol{I}_{m+p} - \boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C} \\ \boldsymbol{J} &= \boldsymbol{J}_{m,p} = \begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{O}_{m\times p} \end{bmatrix} \\ \boldsymbol{C} &= \begin{bmatrix} -\Phi & \boldsymbol{I}_p \end{bmatrix} \\ \boldsymbol{W}_h &= \text{Var}(\boldsymbol{z}_{t+h} - \hat{\boldsymbol{z}}_{t+h}) \end{aligned}$$

Forecasts and FLAP of series



Properties of FLAP

1. The variance reduction is **positive semi-definite**:

$$\mathrm{Var}(\boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}) - \mathrm{Var}(\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h}) \\ = \boldsymbol{J}\boldsymbol{W}_h\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{W}_h\boldsymbol{J}' \quad (1)$$

2. The forecast error variance reductions, i.e. the diagonal elements of Equation 1 is **non-decreasing** as p increases.

3. The projection is the solution to separable optimisation objective functions that **minimise** forecast error variance.

Properties of FLAP

1. The forecast error variance is **reduced** with FLAP
2. The forecast error variance **monotonically** decreases with increasing number of components
3. The forecast projection is **optimal** to achieve minimum forecast error variance of each series

In practice, we need to

$$\tilde{\boldsymbol{y}}_{t+h} = \boldsymbol{J}\tilde{\boldsymbol{z}}_{t+h} = \boldsymbol{J}\boldsymbol{M}\hat{\boldsymbol{z}}_{t+h}$$

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

$$\boldsymbol{W}_h = \text{Var}(\boldsymbol{z}_{t+h} - \hat{\boldsymbol{z}}_{t+h})$$
$$\boldsymbol{C} = \begin{bmatrix} -\Phi & \boldsymbol{I}_p \end{bmatrix}$$

- Estimate $\boldsymbol{W}_h$
- Construct Φ

Estimation of W_h

Shrinking variance towards their median (Opgen-Rhein and Strimmer 2007) and **shrinking covariance** towards zero (Schäfer and Strimmer 2005).

The shrinkage estimator is

- Positive definite, and
- Numerically stable.

In empirical applications, we assume

$$\widehat{\boldsymbol{W}}_h^{shr} = \eta_h \widehat{\boldsymbol{W}}_1^{shr}.$$

Components: construction of Φ

1. Reduce dependency between components
    - TS-PCA (Chang, Guo, and Yao 2018)
    - ICA (Bell and Sejnowski 1995)
2. *also* Emphasise features relevant to forecasting
    - PCA (Jolliffe 2002)
    - CC (Box and Tiao 1977)
    - ForeCA (Goerg 2013)

Components: construction of Φ

1. Reduce dependency between components
   - TS-PCA (Chang, Guo, and Yao 2018)
   - ICA (Bell and Sejnowski 1995)
2. *also* Emphasise features relevant to forecasting
   - PCA (Jolliffe 2002)
   - CC (Box and Tiao 1977)
   - ForeCA (Goerg 2013)

Simulation

Generating values from a normal distribution and normalising them to unit vectors

Reduce dependency between components

Principal component analysis for stationary vector time series (TS-PCA)

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**.

Independent component analysis (ICA)

Statistically **independent** latent components.

Emphasise features relevant to forecasting

Principal component analysis (PCA)

Maximise variance

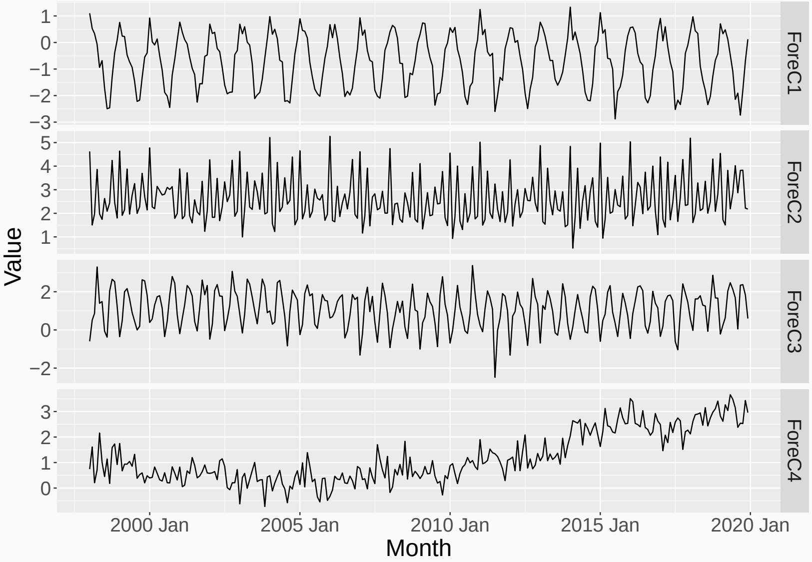
Canonical Component analysis (CC)

Maximise the ratio of explained variance to total variance in an AR process of the component.

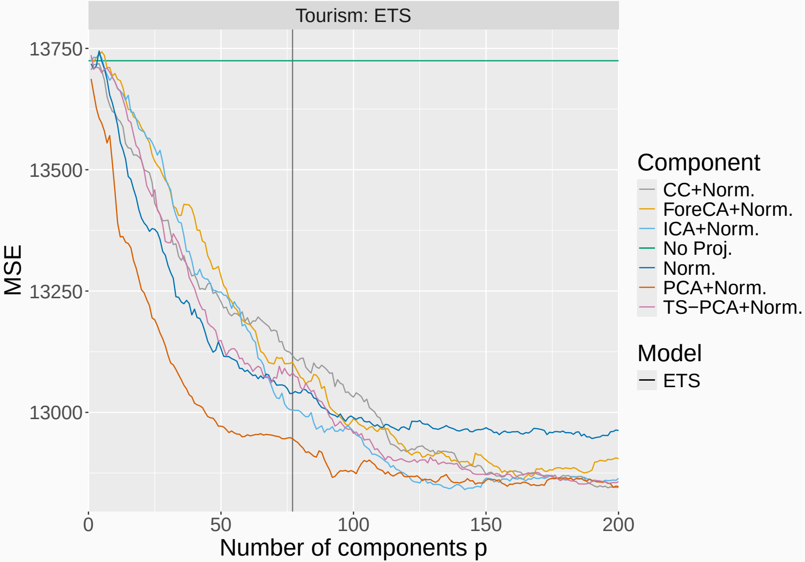
Forecastable component analysis (ForeCA)

Maximised forecastability $\Omega(c_t) = 1 - \frac{H_{s,a}(c_t)}{\log_a(2\pi)}$, where $H_{s,a}(c_t)$ is the Shannon entropy (Shannon 1948) of the spectral density of the component.

Forecastable component analysis (ForeCA)



Tourism

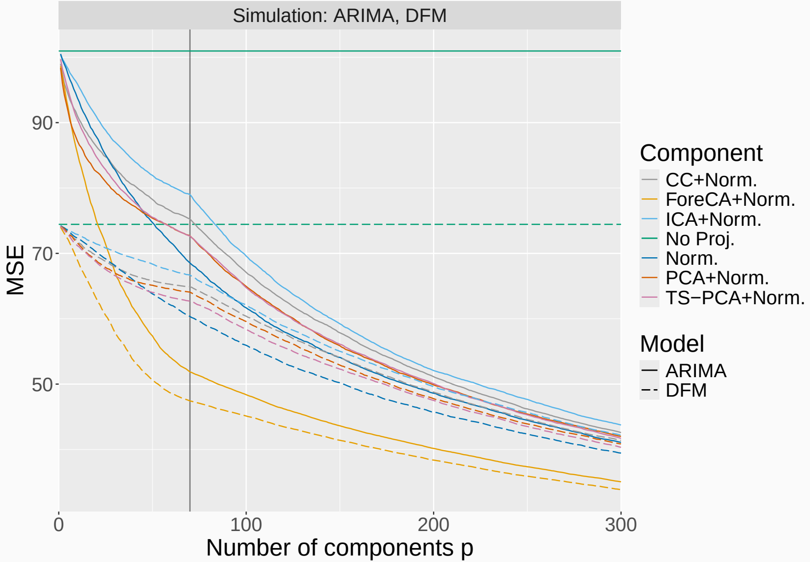


Tourism: ETS
13750
13500
13250
13000
MSE
0
50
100
150
200
Number of components p
Component
CC+Norm.
ForeCA+Norm.
ICA+Norm.
No Proj.
Norm.
PCA+Norm.
TS−PCA+Norm.
Model
ETS

Simulation

- Data generating process (DGP): VAR(3) with $m = 70$ variables
- Sample size: $T = 400$
- Number of repeated samples: 220
- Base model: ARIMA and DFM

Simulation



Simulation: ARIMA, DFM
90
70
50
MSE
0
100
200
300
Number of components p
Component
CC+Norm.
ForeCA+Norm.
ICA+Norm.
No Proj.
Norm.
PCA+Norm.
TS−PCA+Norm.
Model
ARIMA
DFM

R Package `flap`

You can install the stable version from CRAN

```r
## CRAN.R-project.org/package=flap
install.packages("flap")
```

or the development version from Github

```r
## github.com/FinYang/flap
# install.packages("remotes")
remotes::install_github("FinYang/flap")
```

Contact

yangzhuoranyang.com/talk/sis2025/

yangzhuoran.yang@maastrichtuniversity.nl
yangzhuoranyang.com

References i

Bell, Anthony J, and Terrence J Sejnowski. 1995. "An information-maximization approach to blind separation and blind deconvolution." *Neural Comput.* 7 (6): 1129–59. https://doi.org/10.1162/neco.1995.7.6.1129.

Box, G E P, and G C Tiao. 1977. "A canonical analysis of multiple time series." *Biometrika* 64 (2): 355–65. https://doi.org/10.1093/biomet/64.2.355.

Chang, Jinyuan, Bin Guo, and Qiwei Yao. 2018. "Principal component analysis for second-order stationary vector time series." *Ann. Stat.* 46: 2094–124. https://doi.org/10.1214/17-AOS1613.

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

Jolliffe, Ian T. 2002. *Principal Component Analysis*. 2nd ed. Springer Series in Statistics. New York, NY: Springer. https://doi.org/10.1007/b98835.

Opgen-Rhein, Rainer, and Korbinian Strimmer. 2007. "Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach." *Statistical Applications in Genetics and Molecular Biology* 6 (February). https://doi.org/10.2202/1544-6115.1252.

References ii

Schäfer, Juliane, and Korbinian Strimmer. 2005. “A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics.” *Statistical Applications in Genetics and Molecular Biology* 4 (November). https://doi.org/10.2202/1544-6115.1175.

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