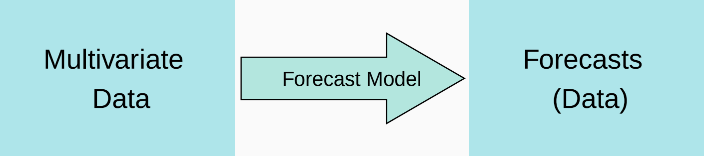
MONASH University

MONASH BUSINESS SCHOOL

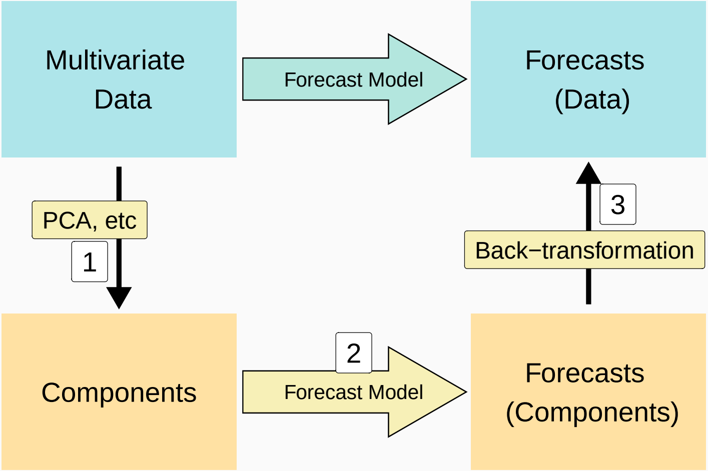
Component-based Approach in Multivariate and Hierarchical Forecasting

Yangzhuoran Fin Yang
Rob J. Hyndman
George Athanasopoulos
Anastasios Panagiotelis

What people do



What we do


Multivariate Data
Forecast Model
Forecasts (Data)
PCA, etc
1
3
Back−transformation
Components
2
Forecast Model
Forecasts (Components)

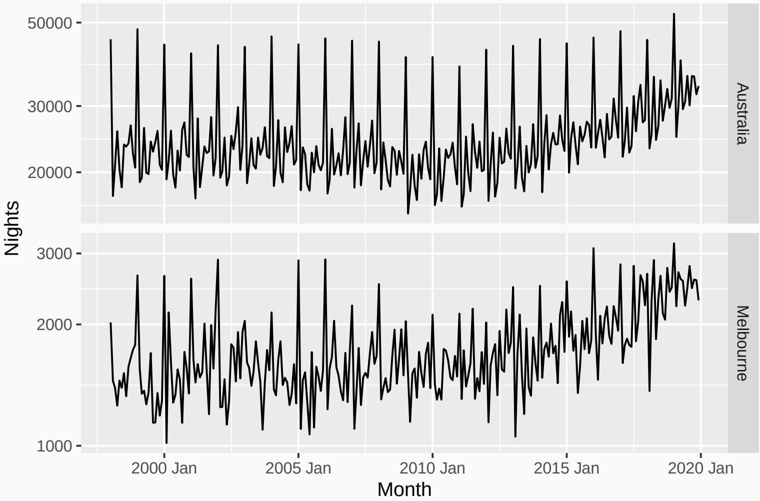
Australian tourism data

- Australia comprises seven states and territories which can be divided into 76 regions
  - For example, Melbourne, Sydney, East Coast

Visitor nights

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

Total and Region



Melbourne and Sydney

Intuition

Observation

1. Better signal-noise ratio in the linear combination.

Intuition

Observation

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

Intuition

Observation

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

One step further

Finding components that have better signal-noise ratio:

1. Easy to forecast;
2. Capturing the common signals;
3. Improving forecast of original series.

Literature

Factor model (Bai and Ng, 2008)

1. Linear transformation
2. VAR models

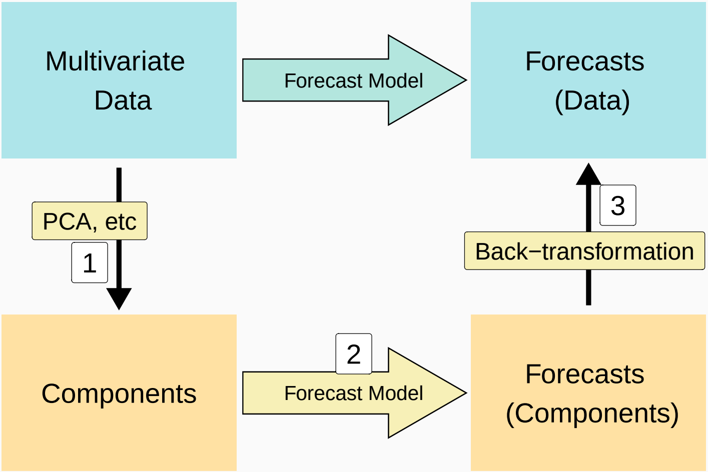
Dynamic Factor Machine Learning (DFML, De Stefani and Bontempi, 2021)

1. Nonlinear transformations with an inherent two-way mapping
   - Autoencoder
2. Machine learning forecast methods

Our differences

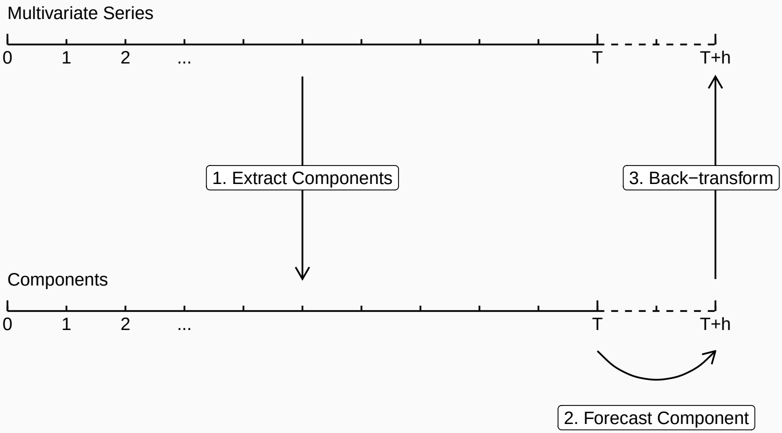
1. Allowing nonlinear transformations
2. Allowing transformations without an inverse function
3. Mappings between forecasts of the components and forecasts of the original series
4. Arbitrary forecast models

Overview



Multivariate Data
Forecast Model
Forecasts (Data)
PCA, etc
1
3
Back−transformation
Components
2
Forecast Model
Forecasts (Components)

Overview


Multivariate Series
0
1
2
...
T
T+h
1. Extract Components
3. Back−transform
Components
0
1
2
...
T
T+h
2. Forecast Component

1. Components: Linear

Taking the first q linear combinations

$$\underset{T\times k}{\boldsymbol{Y}}\underset{k\times q}{\boldsymbol{W}} = \underset{T\times q}{\boldsymbol{C}},$$

where $\boldsymbol{C}$ is the first q components, $\boldsymbol{W}$ is the weighting matrix.

Principal Component Analysis (PCA)

Finding the weights matrix so that the resulting components **maximise variance**:

$$\widehat{\boldsymbol{w}}_s = \underset{\boldsymbol{w} \in \mathbb{R}^k}{\operatorname{argmax}} ||\boldsymbol{Y}\boldsymbol{w}||_2, \quad s = 1, \ldots, q$$

$$\text{subject to } \boldsymbol{Y}\boldsymbol{w}_s \perp \left\{\boldsymbol{Y}\boldsymbol{w}_1, \ldots, \boldsymbol{Y}\boldsymbol{w}_{s-1}\right\},$$

where $\boldsymbol{Y}$ is centred and $|| \cdot ||_2$ denotes the L2 norm.

Forecastable Component (ForeC)

Forecastable components (Goerg, 2013) maximise **forecastability** $\Omega(\cdot)$, finding linear combinations with **most regular patterns**:

$$\widehat{\mathbf{w}}_s = \underset{\mathbf{w}\in\mathbb{R}^k\times 1}{\operatorname{argmax}} \left(\Omega(\mathbf{Y}\mathbf{w})\right), \quad s = 1, \ldots, q$$

$$\text{subject to } \mathbf{Y}\mathbf{w}_s \perp \left\{\mathbf{Y}\mathbf{w}_1, \ldots, \mathbf{Y}\mathbf{w}_{s-1}\right\},$$

and

$$\Omega\left(y_t\right) = 1 - H\left(y_t\right),$$

where $H(y_t)$ is The Shannon entropy (Shannon, 1948) of the spectral density of y_t

1. Component: Nonlinear

Manifold learning

Nonlinear dimension reduction that preserves the distances between points (relative locations of points) on a manifold

- Isomap, Laplacian Eigenmaps
- No back-transformation methods available

2. Forecast model

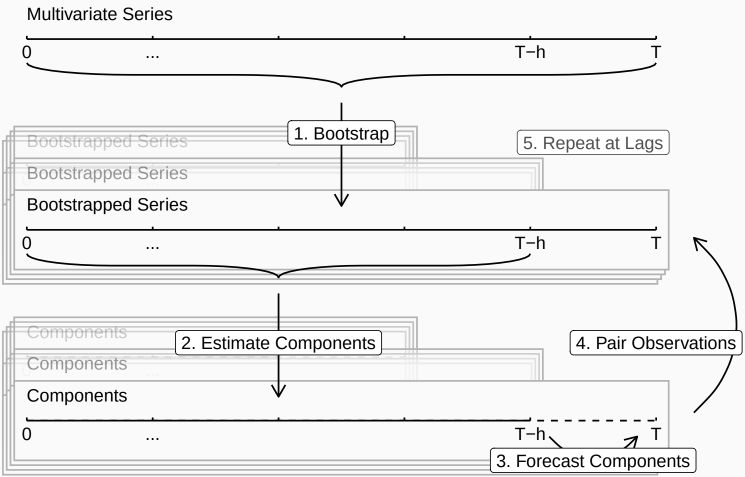
Arbitrary choice of forecast models

- ARIMA
- Exponential smoothing
- Dynamic regression models
- Machine learning methods
- etc

3. Back-transformation

1. Construct a training set
    - Bootstrap to increase the sample size
    - Expanding window to cover more sample values
    - Redo Component Extraction and Component Forecast on each bootstrapped set
2. Fit a back-transformation model using the above as the sample

Construct Training Set

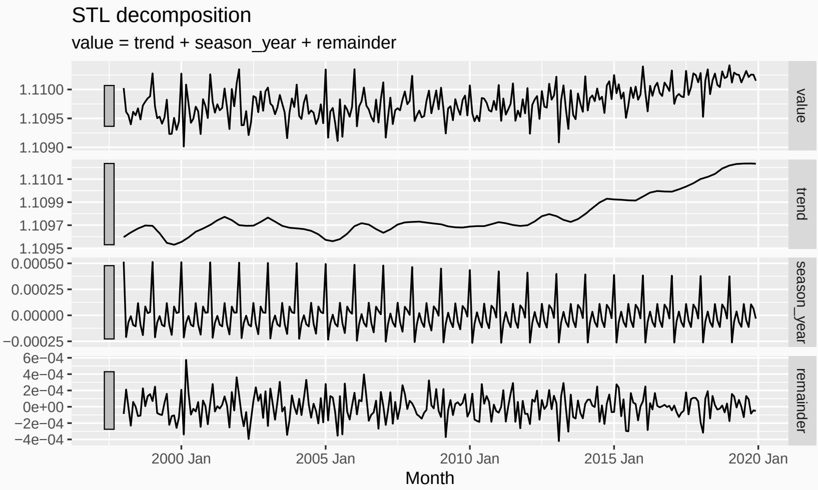


Bootstrap

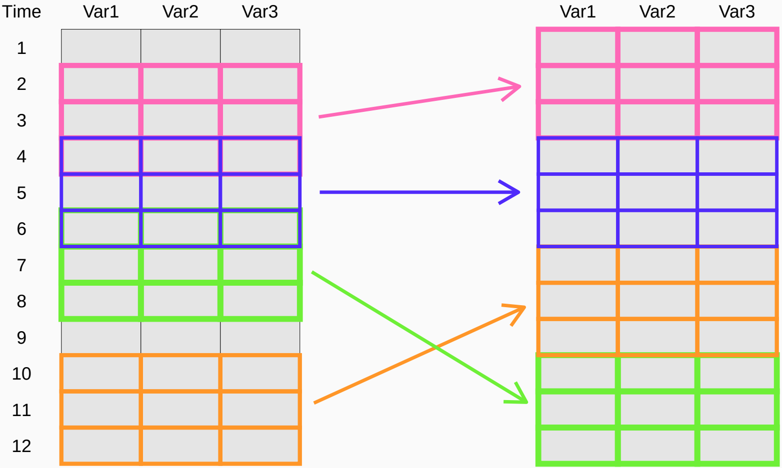
Bergmeir et al. (2016)

1. Box Cox Transformation
   - Stabilising variance
2. Seasonal and Trend decomposition using Loess (STL)
   - Separating series into trend, seasonality and the stationary remainder
3. Moving Block Bootstrap (MBB)
   - Bootstrapping stationary remainder
4. Adding back trend and seasonality. Inverting Box Cox transformation.

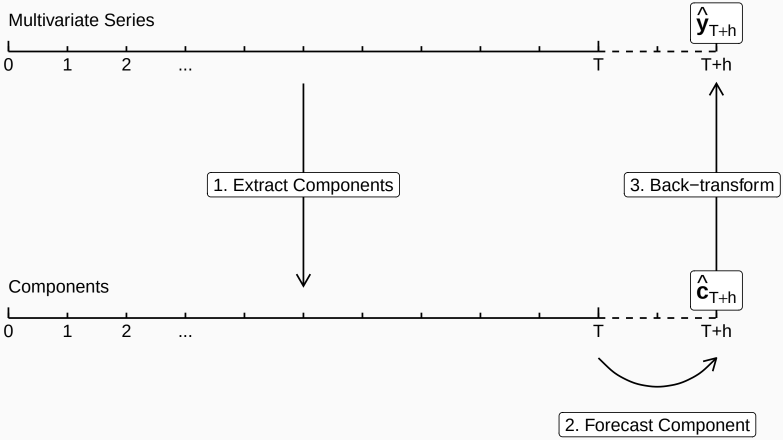
STL Decomposition



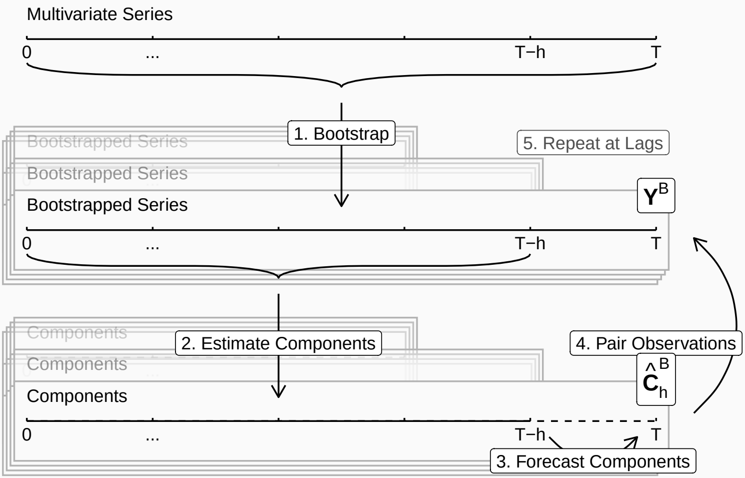
MMB on the remainder



Overview


Multivariate Series
$\hat{y}_{T+h}$
0 1 2 ... T T+h
1. Extract Components
3. Back−transform
Components
$\hat{c}_{T+h}$
0 1 2 ... T T+h
2. Forecast Component

Construct Training Set



Back-transformation Model

- $\hat{\boldsymbol{C}}_h^{\mathrm{B}}$: h-step-ahead forecasts of components from different bootstraps at different lags
- $\boldsymbol{Y}^{\mathrm{B}}$: the corresponding “real” values of the original series from bootstraps
- $\hat{\boldsymbol{c}}_{T+h}$: h-step-ahead forecasts of components of the original series
- $\hat{\boldsymbol{y}}_{T+h}$: h-step-ahead forecasts of the original series
- $\boldsymbol{S}$: collection of seasonal dummies corresponding to $\hat{\boldsymbol{C}}_h^{\mathrm{B}}$
- $\boldsymbol{s}_{T+h}$: seasonal dummies at time $T+h$
- $\boldsymbol{X} = \left[\hat{\boldsymbol{C}}_h^{\mathrm{B}} \quad \boldsymbol{S}\right]$

Back-transformation Model

$$\hat{\boldsymbol{y}}_{T+h} = f\left(\hat{\boldsymbol{c}}_{T+h}, \boldsymbol{s}_{T+h}\right)$$

Discounted Least Squares (DLS)

$$\hat{\boldsymbol{y}}_{T+h} = \hat{\boldsymbol{B}}' \begin{bmatrix} \hat{\boldsymbol{c}}_{T+h} \\ \boldsymbol{s}_{T+h} \end{bmatrix}$$

$$\hat{\boldsymbol{B}} = \left(\boldsymbol{X}'\boldsymbol{U}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{U}\boldsymbol{Y}^{\mathrm{B}},$$

Results

Performance Measure (cross-validation)

$$mRMSSE = \frac{1}{Mk}\sum_{j}^{M}\sum_{i}^{k}\sqrt{\frac{\left(y_{T-j+h,i} - \hat{y}_{T-j,h,i}\right)^2}{\frac{1}{T-j-\nu}\sum_{t=1+\nu}^{T-j}\left(y_{ti} - y_{t-\nu,i}\right)^2}}.$$

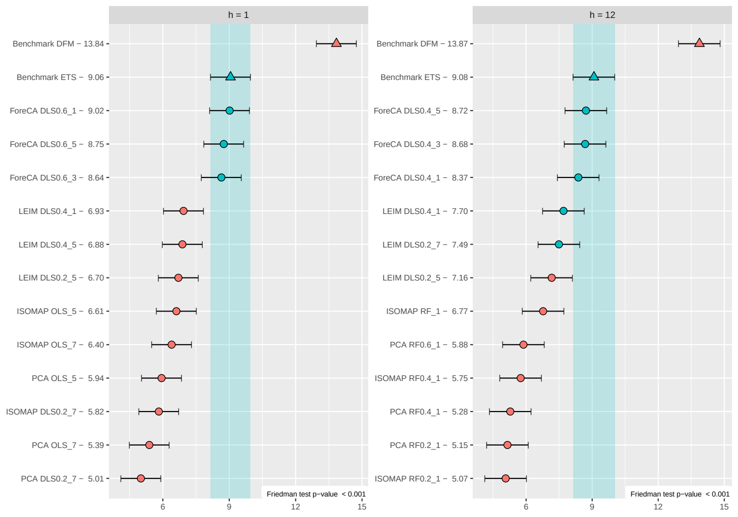
Multiple Comparisons with the Best (MCB)

Compare Average ranks of mRMSSE from independent simulation or cross-validations (Koning et al., 2005)

Forecast model

Automatically selected ExponenTial Smoothing (ETS) model using AICc

Australian tourism: PCA



Conclusion

- Generic method to forecast using lower dimensional components with arbitrary choices of components and forecast models
- Robust to the number of components
- PCA and ISOMAP are competitive in short-term forecasts
- Laplacian Eigenmaps show better performance in longer-term forecasts

Appendix

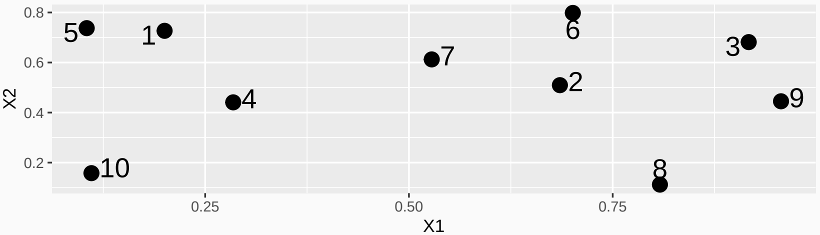
Isomap

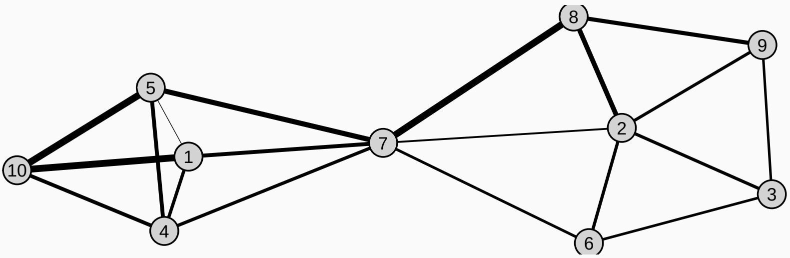
1. Construct Nearest Neighbour Graph
2. Estimate the Geodesic distances (distances along a manifold)
3. Apply Classical MDS
   - Input distances
   - Output coordinates in a lower dimension with similar distances

Other Components

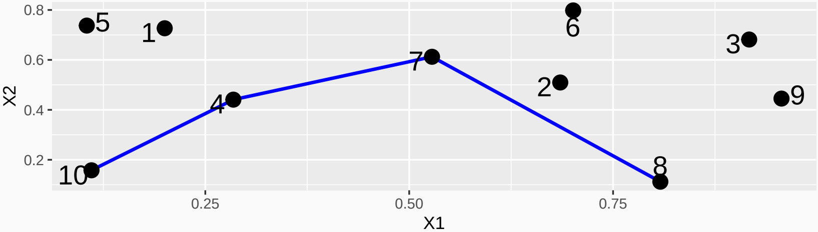
Laplacian Eigenmaps, etc

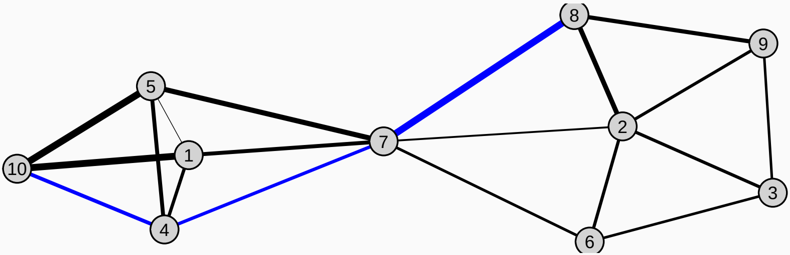
Isomap





Isomap



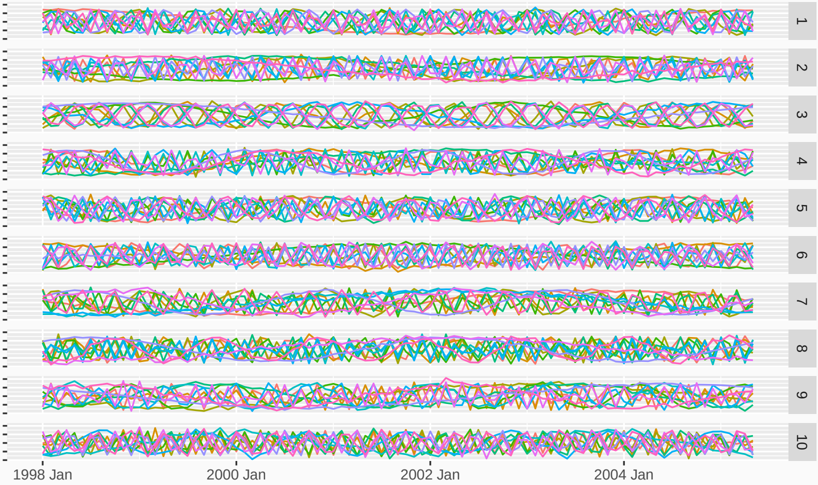


Components Clustering

Problem

- Some components do not have order.
  - e.g. ForeCA
- Components from the bootstraps should provide similar information about the future

Components Clustering: Before

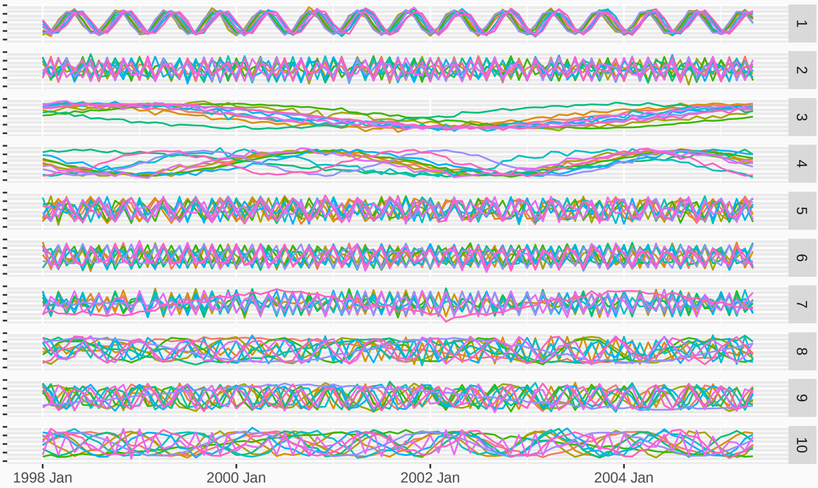


Components Clustering

Solution: Feature-based clustering

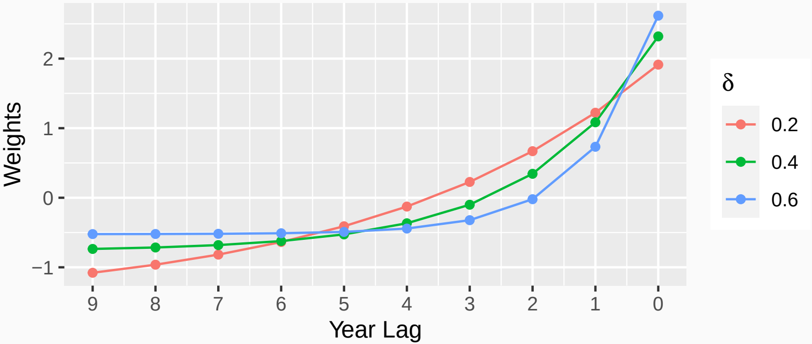
1. Calculate features from each component
   - Highly comparative time-series analysis: Fulcher and Jones (2017)
   - Talagala et al. (2023)
2. Cluster the features
   - K-means with cannot-link constraints: COP kmeans Wagstaff et al. (2001)

Components Clustering: After



Discounted Least Squares (DLS)

$$u = \delta(1 - \delta)^{\text{YearLag}}$$



Box Cox Transformation

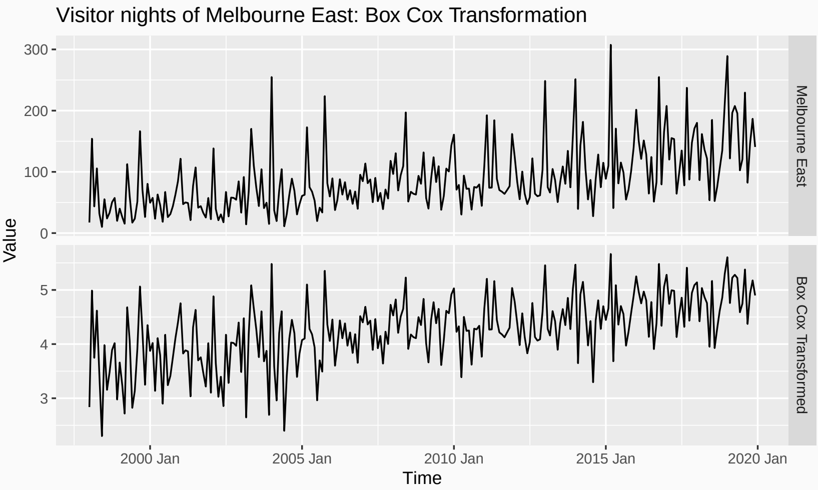
Modified version of Bickel and Doksum (1981)

$$w_t = \begin{cases} \log(y_t) & \text{if } \lambda = 0; \\ (\text{sign}(y_t)|y_t|^\lambda - 1)/\lambda & \text{otherwise,} \end{cases}$$

Reverse transformation

$$y_t = \begin{cases} \exp(w_t) & \text{if } \lambda = 0; \\ \text{sign}(\lambda w_t + 1)|\lambda w_t + 1|^{1/\lambda} & \text{otherwise.} \end{cases}$$

Box Cox Transformation



References i

Jushan Bai and Serena Ng. Large dimensional factor analysis. *Foundations and Trends® in Econometrics*, 3(2):89–163, 2008. ISSN 1551-3076. doi: 10.1561/0800000002. URL http://dx.doi.org/10.1561/0800000002.

Christoph Bergmeir, Rob J Hyndman, and José M Benítez. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International J Forecasting*, 32(2):303–312, 1 April 2016. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2015.07.002. URL https://www.sciencedirect.com/science/article/pii/S0169207015001120.

Peter J Bickel and Kjell A Doksum. An analysis of transformations revisited. *J. American Statistical Association*, 76(374):296–311, June 1981. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1981.10477649. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1981.10477649.

References ii

Jacopo De Stefani and Gianluca Bontempi. Factor-Based framework for multivariate and multi-step-ahead forecasting of large scale time series. *Frontiers in Big Data*, 4:690267, 10 September 2021. ISSN 2624-909X. doi: 10.3389/fdata.2021.690267. URL http://dx.doi.org/10.3389/fdata.2021.690267.

Ben D Fulcher and Nick S Jones. hctsa: A computational framework for automated Time-Series phenotyping using massive feature extraction. *Cell Systems*, 5(5):527–531.e3, 22 November 2017. ISSN 2405-4712. doi: 10.1016/j.cels.2017.10.001. URL http://dx.doi.org/10.1016/j.cels.2017.10.001.

Georg Goerg. Forecastable component analysis. In *Proceedings of the 30th International Conference on Machine Learning*, pages 64–72, Atlanta, Georgia, 2013. URL http://proceedings.mlr.press/v28/goerg13.pdf.

References iii

Alex J Koning, Philip Hans Franses, Michèle Hibon, and H O Stekler. The M3 competition: Statistical tests of the results. *International J. Forecasting*, 21 (3):397–409, 1 July 2005. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2004.10.003. URL https://www.sciencedirect.com/science/article/pii/S0169207004000810.

C E Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948. ISSN 0005-8580. doi: 10.1002/j.1538-7305.1948.tb01338.x. URL http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Thiyanga S Talagala, Rob J Hyndman, and George Athanasopoulos. Meta-learning how to forecast time series. *J Forecasting*, 2023. doi: 10.1002/for.2963. URL https://doi.org/10.1002/for.2963.

Kiri Wagstaff, Claire Cardie, Seth Rogers, and Stefan Schrödl. Constrained k-means clustering with background knowledge. In *Proceedings of the 18th International Conference on Machine Learning*, pages 577–584, 2001. URL https://web.cse.msu.edu/~cse802/notes/ConstrainedKmeans.pdf.