SIS-FENStatS 2026

valgrAI

Biarchetype analysis for univariate functional data. An application to macroeconomic financial time series

Aleix Alcacer, Rafael Benítez, Vicente J. Bolós, Irene Epifanio

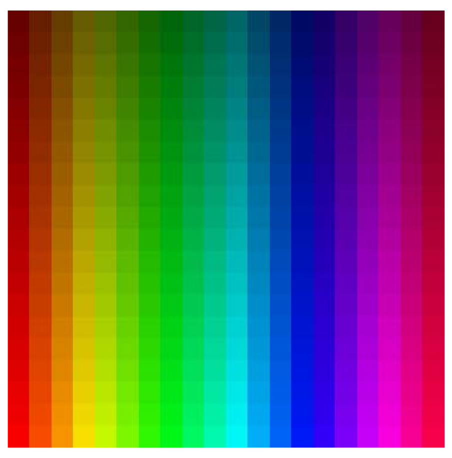
UNIVERSITAT JAUME I

UNIVERSITAT DE VALÈNCIA

1 BACKGROUND: ARCHETYPE ANALYSIS (AA) and BiAA

Archetype Analysis is an unsupervised statistical learning method that represents data as mixtures of "extreme" or pure types (archetypes), which are mixtures of the actual observations themselves. See [2,3] and previous poster for examples. 👀

Biarchetype Analysis (BiAA) extends AA by simultaneously identifying archetypes for both observations and variables, with each represented as a mixture of biarchetypes (see [1]). It provides a more interpretable alternative to biclustering, focusing on extreme values in data for better insight.



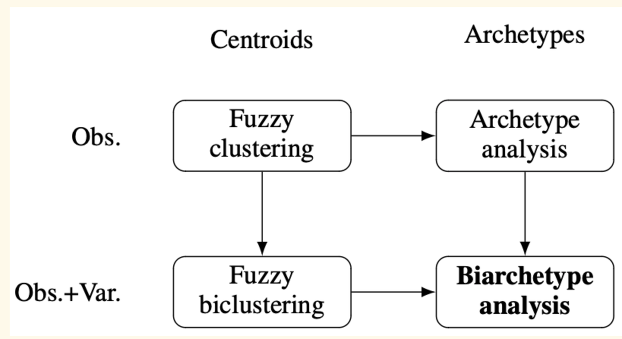
Dataset representing a set of colors where axis X represents the hue and axis Y the brightness.

Archetypal Analysis with 2 archetypes in Y axis:

Archetypal Analysis with 3 archetypes in X axis:

Biarchetypal Analysis with 2 archetypes in Y axis and 3 archetypes in X axis:

Diagram of the relationship between BiAA and other unsupervised methods:



PENDING

BiAA has so far only been applied to multivariate data and has **NOT** yet been adapted to functional data analysis.

2 BACKGROUND: FUNCTIONAL ARCHETYPAL ANALYSIS (FAA)

Let $\{x_i(t)\}_{i=1}^{n}$ denote a sample of univariate functional observations defined on a common domain $\mathcal{T}$.

The k archetypes $\{z_g(t)\}_{g=1}^{k}$ are defined as convex combinations of the observed functions (see [4]):

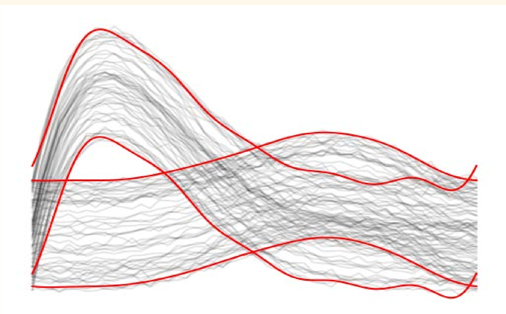
$$z_g(t) = \sum_{l=1}^{n} \beta_{gl} x_l(t), \qquad \sum_{l=1}^{n} \beta_{gl} = 1, \qquad \beta_{gl} \geq 0, \qquad g = 1, \dots, k.$$

Each functional observation is then approximated as a convex combination of the archetypes:

$$x_i(t) \approx \hat{x}_i(t) = \sum_{g=1}^{k} \alpha_{ig} z_g(t), \qquad \sum_{g=1}^{k} \alpha_{ig} = 1, \qquad \alpha_{ig} \geq 0, \qquad i = 1, \dots, n.$$

Under these constraints, the residual sum of squares to be minimized is

$$RSS = \sum_{i=1}^{n} \int_{\mathcal{T}} \left(x_i(t) - \sum_{g=1}^{k} \alpha_{ig} \sum_{l=1}^{n} \beta_{gl} x_l(t) \right)^2 dt \, .$$



Four functional archetypes in red.

FAA by discretizing the functions:

In practice, functions $\{x_i(t)\}_{i=1}^{n}$ are observed on $t_1, \dots, t_m$, yielding the data matrix $\mathbf{X}_{n\times m}$. The archetype matrix $\mathbf{Z}_{k\times m}$ is defined by

$$\mathbf{Z}_{k\times m} = \boldsymbol{\beta}_{k\times n} \cdot \mathbf{X}_{n\times m}, \qquad \sum_{l=1}^{n} \beta_{gl} = 1, \qquad \beta_{gl} \geq 0, \qquad g = 1, \dots, k.$$

The data matrix is approximated:

$$\mathbf{X}_{n\times m} \approx \hat{\mathbf{X}}_{n\times m} = \boldsymbol{\alpha}_{n\times k} \cdot \mathbf{Z}_{k\times m}, \qquad \sum_{g=1}^{k} \alpha_{ig} = 1, \qquad \alpha_{ig} \geq 0, \qquad i = 1, \dots, n.$$

Under these constraints, the residual sum of squares to be minimized is

$$RSS = \left\| \mathbf{X}_{n\times m} - \hat{\mathbf{X}}_{n\times m} \right\|^2 = \left\| \mathbf{X}_{n\times m} - \boldsymbol{\alpha}_{n\times k} \cdot \boldsymbol{\beta}_{k\times n} \cdot \mathbf{X}_{n\times m} \right\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(x_i(t_j) - \sum_{g=1}^{k} \alpha_{ig} \sum_{l=1}^{n} \beta_{gl} x_l(t_j) \right)^2 .$$

3 PROPOSAL: FUNCTIONAL BiAA NEW

Functional Biarchetype Analysis simultaneously identifies k archetypes in observations and c archetypes in times. The biarchetype matrix $\mathbf{Z}_{k\times c}$ is defined by

$$\mathbf{Z}_{k\times c} = \boldsymbol{\beta}_{k\times n} \cdot \mathbf{X}_{n\times m} \cdot \boldsymbol{\theta}_{m\times c}, \qquad \sum_{l=1}^{n} \beta_{gl} = 1, \qquad \beta_{gl} \geq 0, \qquad g = 1, \dots, k,$$
$$\sum_{r=1}^{m} \theta_{rh} = 1, \qquad \theta_{rh} \geq 0, \qquad h = 1, \dots, c.$$

The data matrix is approximated:

$$\mathbf{X}_{n\times m} \approx \hat{\mathbf{X}}_{n\times m} = \boldsymbol{\alpha}_{n\times k} \cdot \mathbf{Z}_{k\times c} \cdot \boldsymbol{\gamma}_{c\times m}, \qquad \sum_{g=1}^{k} \alpha_{ig} = 1, \qquad \alpha_{ig} \geq 0, \qquad i = 1, \dots, n,$$
$$\sum_{h=1}^{c} \gamma_{hj} = 1, \qquad \gamma_{hj} \geq 0, \qquad j = 1, \dots, m.$$

Under these constraints, the residual sum of squares to be minimized is

$$RSS = \left\| \mathbf{X}_{n\times m} - \hat{\mathbf{X}}_{n\times m} \right\|^2 = \left\| \mathbf{X}_{n\times m} - \boldsymbol{\alpha}_{n\times k} \cdot \boldsymbol{\beta}_{k\times n} \cdot \mathbf{X}_{n\times m} \cdot \boldsymbol{\theta}_{m\times c} \cdot \boldsymbol{\gamma}_{c\times m} \right\|^2$$
$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \left(x_i(t_j) - \sum_{g=1}^{k} \sum_{h=1}^{c} \alpha_{ig} \left(\sum_{l=1}^{n} \sum_{r=1}^{m} \beta_{gl} \left(x_l(t_r) \right) \theta_{rh} \right) \gamma_{hj} \right)^2 .$$

Summary:
- $\mathbf{Z}_{k\times c}$ are the biarchetypes.
- $\boldsymbol{\beta}_{k\times n}$ defines archetypes of functional observations (countries in our example).
- $\boldsymbol{\theta}_{m\times c}$ defines archetypes of temporal structure.
- $\boldsymbol{\alpha}_{n\times k}$ shows the proportion of each archetypal country in each approx. country.
- $\boldsymbol{\gamma}_{c\times m}$ shows the proportion of each archetypal time at each instant.

4 APPLICATION

We apply **Functional BiAA** to monthly time series of **10-year government bond yields** of 23 European countries (from Eurostat):

AT (Austria), **BE** (Belgium), **CY** (Cyprus), **CZ** (Czech Republic), **DE** (Germany), **DK** (Denmark), **EL** (Greece), **ES** (Spain), **FI** (Finland), **FR** (France), **HU** (Hungary), **IE** (Ireland), **IT** (Italy), **LT** (Lithuania), **LU** (Luxembourg), **LV** (Latvia), **MT** (Malta), **NL** (Netherlands), **PL** (Poland), **PT** (Portugal), **SE** (Sweden), **SK** (Slovakia), **UK** (United Kingdom).

Sample period from **January 2001** to **April 2025**.

The bidimensional screeplot shows that the best pair (k, c) is $(3, 3)$, resulting in **3 archetypal countries** and **3 archetypal times**.

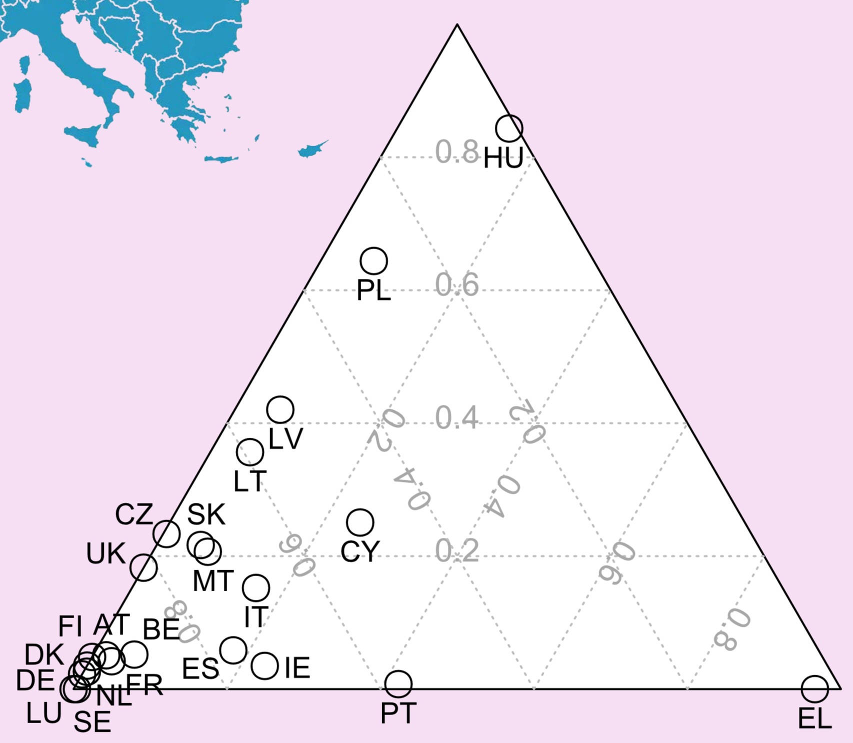
Code and data:
https://epifanio.uji.es/RESEARCH/biaafda.zip
BiAA algorithm in R and Python: see [1].

5 COUNTRY RESULTS

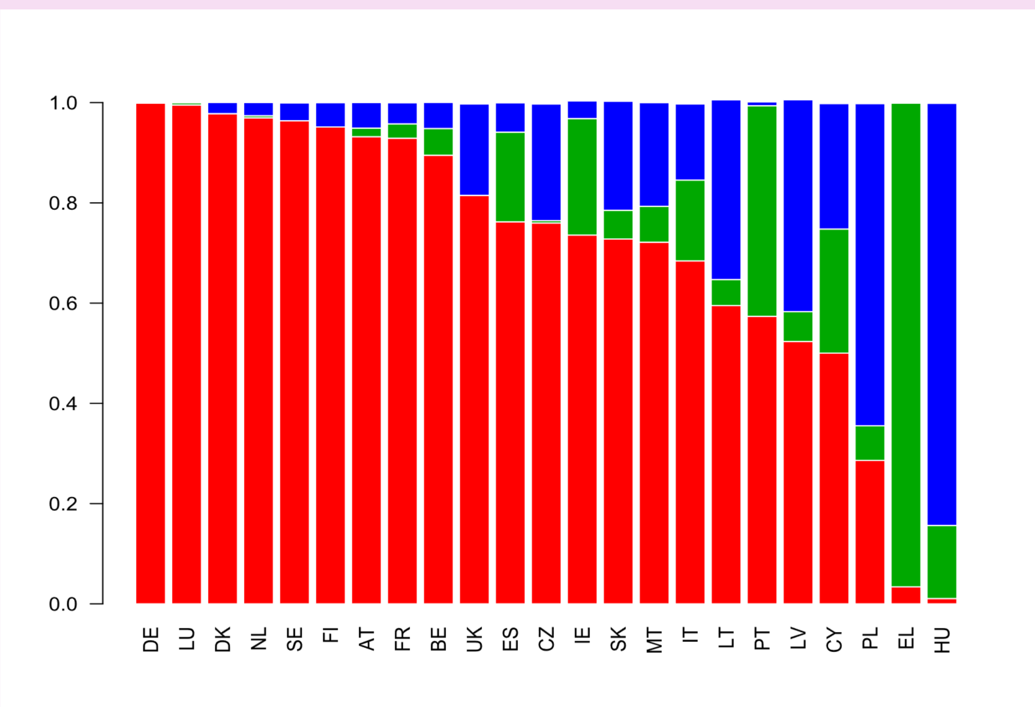
Archetype 1 is exemplified by **DE**, which is composed entirely of this archetype. Other economically advanced countries also exhibit very high shares of archetype 1: **LU**, **DK**, **NL**, **SE**, **FI**, **AT**, **FR**, **BE**.

Archetype 2 is best represented by **EL**, with a share of 96.6%. Several Southern European countries also display substantial proportions of archetype 2: **PT**, **CY**, **IE**, **ES**, **IT**; however, in all these cases, archetype 1 remains dominant.

Archetype 3 is exemplified by **HU**, with a share of 84%. Other countries (primarily located in Central Europe and the Baltic region) also exhibit notable shares: **PL**, **LV**, **LT**, **CY**, **CZ**, **SK**, **MT**, and **UK**. Among these countries, only **PL** displays a majority share of archetype 3. All remaining countries are dominated by archetype 1.

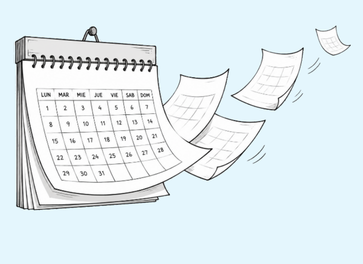


Ternary plot of matrix $\boldsymbol{\alpha}_{n\times k}$. The 3 archetypal countries are represented by the vertices of the triangle.



Parameter matrix $\boldsymbol{\alpha}_{n\times k}$ shows the proportion of each archetypal country (1: red, 2: green, or 3: blue) in each country approximation.

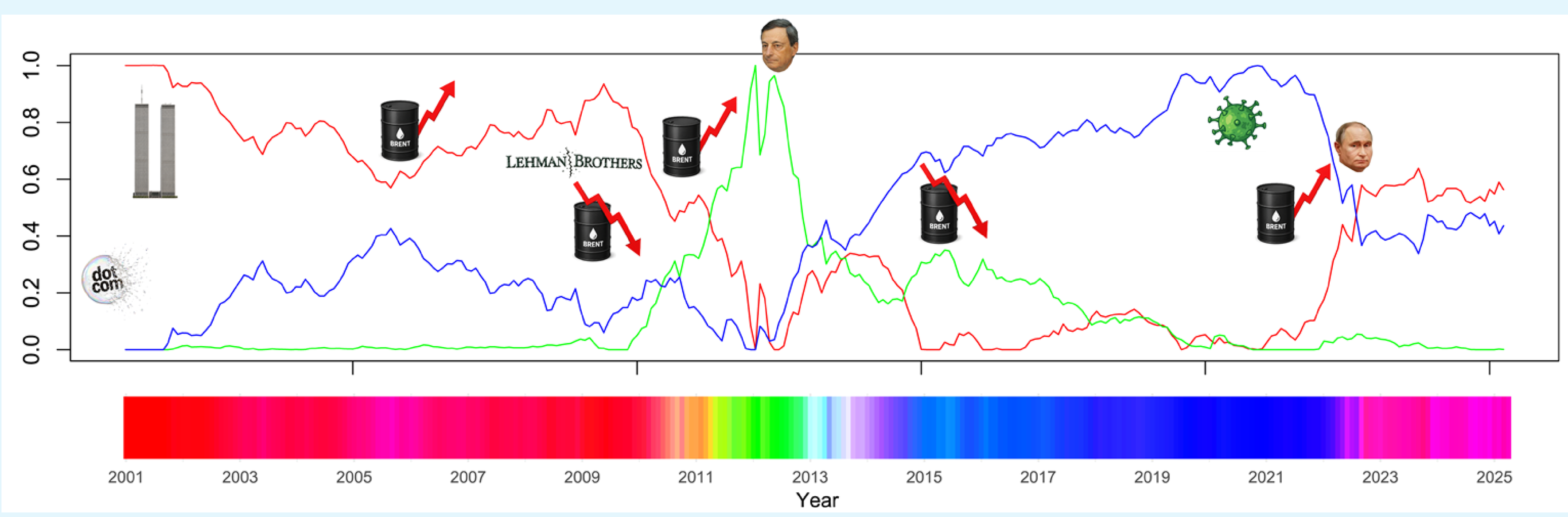
6 TIME RESULTS



Archetype 1 is exemplified by the year **2001**, which is characterized by the dot-com bubble crisis and high yields. It remains dominant until April 2011. Although its influence subsequently declines as yields also decreases, it re-emerges as the dominant archetype from September 2022 onward, jointly with archetype 3.

Archetype 2 is represented by both February and June **2012**, during which it reaches its maximum. This period is associated with the euro area sovereign debt crisis, during which countries such as EL, ES, IT, PT and IE saw yields increase. Notably, in July 2012, Mario Draghi delivered his well-known statement that "*the ECB is ready to do whatever it takes to preserve the euro.*" This archetype dominates the time interval from May 2011 to March 2013.

Archetype 3 remains elevated throughout the COVID-19 crisis and reaches its peak in December **2020**. In this period, yields were very low, even negative for countries such DE or DK. Nevertheless, as previously noted, from September 2022 onwards, archetype 1 regains dominance, albeit with values that are only slightly higher than those of archetype 2. A similar coexistence of these two archetypes was also observed in 2005.



Parameter matrix $\boldsymbol{\gamma}_{c\times m}$, representing the mixture of archetypal times at each instant.

7 CONCLUSIONS

- We introduce biarchetype analysis for the first time in the context of univariate functional data.
- Both cases and time points are expressed as mixtures of biarchetypes.
- Concise and highly interpretable representation of complex functional observations.
- Superior interpretability compared with biclustering approaches, as it is based on extreme, representative patterns rather than centroids.
- We apply the proposed method to 10-year government bond yields of European countries over the period 2001–2025. The results identify three distinct time regimes (the pre-crisis period, the euro-area sovereign debt crisis, and the post-crisis period), and reveal Germany, Greece, and Hungary as country archetypes.

REFERENCES

[1] Alcacer, A., Epifanio, I., Gual-Arnau, X.: Biarchetype analysis: Simultaneous learning of observations and features based on extremes. IEEE Transactions on Pattern Analysis and Machine Intelligence 46(12), 8228–8239 (2024).

[2] Alcacer, A., Epifanio, I., Mair, S., Morup, M.: A survey on archetypal analysis. arXiv preprint arXiv:2504.12392 (2025).

[3] Cutler, A., Breiman, L.: Archetypal Analysis. Technometrics 36(4), 338–347 (1994).

[4] Epifanio, I.: Functional archetype and archetypoid analysis. Computational Statistics & Data Analysis 104, 24 – 34 (2016).

ACKNOWLEDGEMENT

This work was partially supported by the Spanish Ministry of Science and Innovation PID2023-153128NB-I00, PID2022-141699NB-I00, PID2020-118763GA-I00, and Generalitat Valenciana CIPROM/2023/66.