The relationship between clustering and forecast reconciliation

Raffaele Mattera[a]

[a]Sapienza University of Rome, Italy

Outline

Reconciliation with unknown hierarchies

The problem

- Reconciliation is used to forecast time series adhering to *known* linear constraints [Panagiotelis et al., 2021];
- Let $\mathbf{b}_t$ be the vector of all N stocks of interest observed at time t, and let $\mathbf{a}_t$ be a corresponding vector of n_a aggregated time series:

$$\mathbf{a}_t = \mathbf{A}\mathbf{b}_t. \tag{1}$$

- Usually, $\mathbf{A}$ is known since time series can be naturally disaggregated by various attributes of interest (e.g. geographic divisions);
- What if we do not know how the matrix $\mathbf{A}$ looks like?
- What if time series reasonably adhere to multiple unknown linear constraints?
- We know that "hierarchical" and "grouped" aggregations are possible [Hyndman and Athanasopoulos, 2021]. Why not also "clustered" time series too?

Empirical instance I - epidemiology

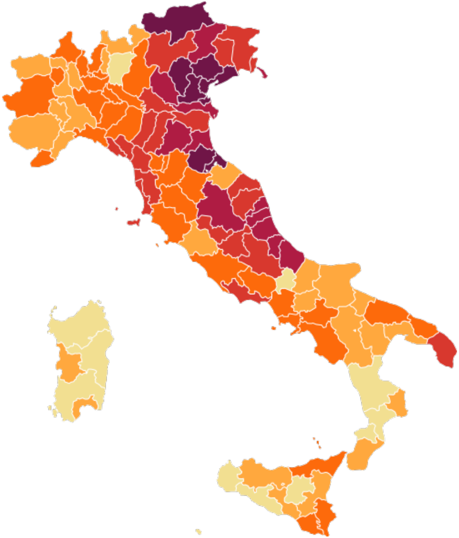

Figure: Number of COVID-19 cases – Italian provinces

Empirical instance II - economics

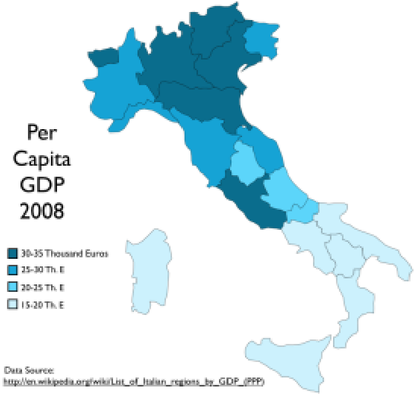

Figure: GDP per capita – Italian regions

Empirical instance III - finance?

- Many instances:
  1. Industry sector (Industry-based portfolios);
  2. Market exchange (Market portfolios);
  3. Intrinsic stock similarities [see clustered the portfolios of Raffinot, 2017];
  4. etc.
- We do not have a clear hierarchy;
- The hierarchy depends by the market topology!

The main problem of time series clustering: defining similarity

- The main issue when dealing with time series is how we define hierarchies;
- We propose to use clustering, but perhaps other approaches can also be exploited (e.g. complex networks);
- One of the most important aspects of time series clustering is the definition of the dissimilarity measure;
- Dissimilarities can be divided into three main classes: observation-based, future-based and model-based [Maharaj et al., 2019].

Some interesting time series features

Features	Examples
Auto-correlation function	D'Urso and Maharaj [2009], Caiado and Crato [2010]
Peridiogram ordinates	Vilar and Pértega [2004], Caiado et al. [2006]
Cepstral coefficients	Savvides et al. [2008], Maharaj and D'Urso [2011]
	D'Urso et al. [2020]
ARMA coefficients	Piccolo [1990], Maharaj [1996]
GARCH coefficients	Otranto [2008], D'Urso et al. [2013], D'Urso et al. [2016]
Spline coefficients	Iorio et al. [2016, 2018]
Distribution parameters	D'Urso et al. [2017], Mattera et al. [2021], Cerqueti et al. [2021]
	Cerqueti et al. [2022a,b]
Global *static* features	Wang et al. [2006], Lubba et al. [2019], Bastos and Caiado [2021]

Table: Main features used for clustering time series

Using clustering for better forecasting is not only a reconciliation story

- There exist load forecasting approaches [e.g. Goia et al., 2010] using clustering for capturing regular behaviours in load time series to perform more accurate forecasts;
- Usually, k-means clustering is used to group customers prior to forecasting. Different statistical methods are used for each cluster and then the results are combined;
- There is also literature developing methods for forecasting time series with clustered factor-augmented regressions [Ando and Bai, 2017, Alonso et al., 2020, Mattera and Franses, 2023b];
- We take a different perspective as we aim at making the bottom forecasts coherent with their similarity structure.

Some open problems

Are known hierarchies better than those uncovered?

- We know that different aggregation structures could exist for the same set of bottom series, but which of them provides better reconciled forecasts empirically (probably case-dependent)?
- Are we sure that, for example, spatial reconciliation provides superior forecasts than those implied by clustering?

What clustering approach?

1. Hierarchical clustering:
   - Very natural setting for defining hierarchies;
   - We do not need to define the number of clusters;
   - but the implied summation matrix is very large, also with not very large bottom series [not big issue Hyndman et al., 2016].
2. We can use either *k*-means or PAM, but also:
   - Fuzzy methods [Bezdek et al., 1984, Krishnapuram et al., 2001];
   - Future-weighting [de Amorim, 2016];
   - Spatio-temporal [D'Urso et al., 2019, Mattera, 2022, Mattera and Franses, 2023a];
   - Mixed data-type [Hunt and Jorgensen, 2011, D'Urso and Massari, 2019];
   - and many others...
3. **Defining a novel clustering approach explicitly taught for enhancing bottom series forecast accuracy?**

Hierarchies combination

- Is there an optimal way of combining different hierarchies?

$$\mathbf{A}_{comb} = \sum_j w_j \mathbf{A}_j \tag{2}$$

and what is the relationship with forecast combination in this case?

- An idea could be STATIS algorithm [Abdi et al., 2012]

Fuzzy clustering and fuzzy forecast reconciliation

- Fuzzy clustering is based on the iterative solution of the following minimization problem:

$$\min : \quad \sum_{i=1}^{N} \sum_{g=1}^{G} u_{i,g}^{m} d^2(\boldsymbol{b}_i, \boldsymbol{b}_g), \tag{3}$$

  where $m > 1$ is the fuzzifier parameter and $u_{i,g}$ the membership degree of the ith unit to the gth cluster;
- In our base framework we deal with binary membership matrices $\boldsymbol{\mathcal{C}}$ of dimension $N \times G$ with element $c_{i,g} = 1$ if time series i is in the c-th cluster, and 0 otherwise.

Fuzzy clustering and fuzzy forecast reconciliation (cont'd)

- With fuzziness, we just need substituting $\mathcal{C}$ with $\mathbf{U}$, i.e. the membership matrix with generic element $u_{i,g}$;
- Thus we get

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}' \\ \mathbf{U}' \end{bmatrix}, \text{ and } \mathbf{S} = \begin{bmatrix} \mathbf{A} \\ \mathbf{I}_N \end{bmatrix} \tag{4}$$

- As a result, given $G = 2$ clusters, the reconciled forecast for the i-th series is equal to a combination between the forecasts coherent with the $G = 1$ cluster, with weight $u_{i,1}$ and the one coherent with the $G = 2$ cluster with weight $u_{i,2}$.

Fuzzy clustering and fuzzy forecast reconciliation (cont'd)

- Let us consider the standard example with $n_b = 5$ bottom series that aggregate into two clusters $C_1 = \{y_{1,t}, y_{2,t}, y_{3,t}\}$ and $C_2 = \{y_{4,t}, y_{5,t}\}$

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t = \begin{bmatrix} y_t \\ y_{\mathrm{C}_1,t} \\ y_{\mathrm{C}_2,t} \\ y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{bmatrix} \tag{5}$$

- which means, e.g.

$$y_{\mathrm{C}_1,t} = y_{1,t} + y_{2,t} + y_{3,t} \tag{6}$$

Fuzzy clustering and fuzzy forecast reconciliation (cont'd)

- Let us assume now

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t = \begin{bmatrix} y_t \\ y_{\mathrm{C}_1,t} \\ y_{\mathrm{C}_2,t} \\ y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.8 & 0.3 & 0.9 & 0.2 & 0.1 \\ 0.2 & 0.7 & 0.1 & 0.8 & 0.9 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{bmatrix} \tag{7}$$

- which implies

$$y_{\mathrm{C}_1,t} = u_{1,1}y_{1,t} + u_{2,1}y_{2,t} + u_{3,1}y_{3,t} + u_{4,1}y_{4,t} + u_{5,1}y_{5,t}. \tag{8}$$

Is probabilistic reconciliation an opportunity for mixture clustering?

- Model-based clustering assumes a probability distribution for the data, typically a finite mixture of G multivariate distributions.
- The probability model is a weighted average of G probability density functions, i.e.,

$$p(\mathbf{y}_i) = \sum_{g=1}^{G} \tau_g f_g(\mathbf{y}_i \mid \boldsymbol{\theta}_g),$$

  where the gth mixing proportion τ_g denotes the probability that observation i 's data were generated by the g th density, where $\tau_g \geq 0$ for $g = 1, 2, \ldots, G$ and $\sum_{g=1}^{G} \tau_g = 1$ and $f_g(\cdot \mid \boldsymbol{\theta}_g)$ is the gth mixture with its parameters collected in $\boldsymbol{\theta}_g$;
- Under the assumption that $f_g(\cdot \mid \boldsymbol{\theta}_g)$ is a multivariate Gaussian distribution, $\boldsymbol{\theta}_g = \{\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g\}$.

Conclusions

Conclusions

- Most reconciliation studies deal with known hierarchies, even if this is not always possible (e.g. finance);
- Allowing for unknown hierarchies could make reconciliation techniques applicable in other domains;
- Deeper studies are required for a better understanding of the potential and usefulness of such approaches;
- It would be really nice to work together on these new possible issues!

References I

H. Abdi, L. J. Williams, D. Valentin, and M. Bennani-Dosse. Statis and distatis: optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2):124–167, 2012.

A. M. Alonso, P. Galeano, and D. Peña. A robust procedure to build dynamic factor models with cluster structure. *Journal of Econometrics*, 216(1):35–52, 2020.

T. Ando and J. Bai. Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, 112(519): 1182–1198, 2017.

J. A. Bastos and J. Caiado. On the classification of financial data with domain agnostic features. *International Journal of Approximate Reasoning*, 138:1–11, 2021.

References II

J. C. Bezdek, R. Ehrlich, and W. Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & geosciences*, 10(2-3):191–203, 1984.

J. Caiado and N. Crato. Identifying common dynamic features in stock returns. *Quantitative Finance*, 10(7):797–807, 2010.

J. Caiado, N. Crato, and D. Peña. A periodogram-based metric for time series classification. *Computational Statistics & Data Analysis*, 50(10): 2668–2684, 2006.

R. Cerqueti, M. Giacalone, and R. Mattera. Model-based fuzzy time series clustering of conditional higher moments. *International Journal of Approximate Reasoning*, 134:34–52, 2021.

R. Cerqueti, P. D'Urso, L. De Giovanni, M. Giacalone, and R. Mattera. Weighted score-driven fuzzy clustering of time series with a financial application. *Expert Systems with Applications*, 198:116752, 2022a.

R. Cerqueti, R. Mattera, and G. Scepi. Multiway clustering with time-varying parameters. *Computational statistics*, pages 1–42, 2022b.

References III

R. C. de Amorim. A survey on feature weighting based k-means algorithms. *Journal of Classification*, 33:210–242, 2016.

P. D'Urso, L. De Giovanni, and R. Massari. Garch-based robust clustering of time series. *Fuzzy Sets and Systems*, 305:1–28, 2016.

P. D'Urso, E. A. Maharaj, and A. M. Alonso. Fuzzy clustering of time series using extremes. *Fuzzy Sets and Systems*, 318:56–79, 2017.

P. D'Urso and E. A. Maharaj. Autocorrelation-based fuzzy clustering of time series. *Fuzzy Sets and Systems*, 160(24):3565–3589, 2009.

P. D'Urso and R. Massari. Fuzzy clustering of mixed data. *Information Sciences*, 505:513–534, 2019.

P. D'Urso, C. Cappelli, D. Di Lallo, and R. Massari. Clustering of financial time series. *Physica A: Statistical Mechanics and its Applications*, 392 (9):2114–2129, 2013.

P. D'Urso, L. De Giovanni, M. Disegna, and R. Massari. Fuzzy clustering with spatial–temporal information. *Spatial Statistics*, 30:71–102, 2019.

References IV

P. D'Urso, L. De Giovanni, R. Massari, R. L. D'Ecclesia, and E. A. Maharaj. Cepstral-based clustering of financial time series. *Expert Systems with Applications*, 161:113705, 2020.

A. Goia, C. May, and G. Fusai. Functional clustering and linear regression for peak load forecasting. *International Journal of Forecasting*, 26(4): 700–711, 2010.

L. Hunt and M. Jorgensen. Clustering mixed data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4):352–361, 2011.

R. J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Melbourne, Australia, 3rd edition, 2021. URL `OTexts.com/fpp3`.

R. J. Hyndman, A. J. Lee, and E. Wang. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis*, 97:16–32, 2016.

C. Iorio, G. Frasso, A. D'Ambrosio, and R. Siciliano. Parsimonious time series clustering using p-splines. *Expert Systems with Applications*, 52: 26–38, 2016.

C. Iorio, G. Frasso, A. D'Ambrosio, and R. Siciliano. A p-spline based clustering approach for portfolio selection. *Expert Systems with Applications*, 95:88–103, 2018.

R. Krishnapuram, A. Joshi, O. Nasraoui, and L. Yi. Low-complexity fuzzy relational clustering algorithms for web mining. *IEEE transactions on Fuzzy Systems*, 9(4):595–607, 2001.

C. H. Lubba, S. S. Sethi, P. Knaute, S. R. Schultz, B. D. Fulcher, and N. S. Jones. catch22: Canonical time-series characteristics. *Data Mining and Knowledge Discovery*, 33(6):1821–1852, 2019.

E. A. Maharaj. A significance test for classifying arma models. *Journal of Statistical Computation and Simulation*, 54(4):305–331, 1996.

References VI

E. A. Maharaj and P. D'Urso. Fuzzy clustering of time series in the frequency domain. *Information Sciences*, 181(7):1187–1211, 2011.

E. A. Maharaj, P. D'Urso, and J. Caiado. *Time series clustering and classification*. Chapman and Hall/CRC, 2019.

R. Mattera. A weighted approach for spatio-temporal clustering of covid-19 spread in italy. *Spatial and Spatio-temporal Epidemiology*, 41: 100500, 2022.

R. Mattera and P. H. Franses. Are african business cycles synchronized? evidence from spatio-temporal modelling. *Economic Modelling*, page 106485, 2023a.

R. Mattera and P. H. Franses. Forecasting spatio-temporal data with clustered factor models with an application to house prices. *Working paper*, 2023b.

R. Mattera, M. Giacalone, and K. Gibert. Distribution-based entropy weighting clustering of skewed and heavy tailed time series. *Symmetry*, 13(6):959, 2021.

E. Otranto. Clustering heteroskedastic time series by model-based procedures. *Computational Statistics & Data Analysis*, 52(10): 4685–4698, 2008.

A. Panagiotelis, G. Athanasopoulos, P. Gamakumara, and R. J. Hyndman. Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, 37(1):343–359, 2021.

D. Piccolo. A distance measure for classifying arima models. *Journal of Time Series Analysis*, 11(2):153–164, 1990.

T. Raffinot. Hierarchical clustering-based asset allocation. *Journal of Portfolio Management*, 44(2):89–99, 2017.

References VIII

A. Savvides, V. J. Promponas, and K. Fokianos. Clustering of biological time series by cepstral coefficients based distances. *Pattern Recognition*, 41(7):2398–2412, 2008.

J. A. Vilar and S. Pértega. Discriminant and cluster analysis for gaussian stationary processes: Local linear fitting approach. *Journal of Nonparametric Statistics*, 16(3-4):443–462, 2004.

X. Wang, K. Smith, and R. Hyndman. Characteristic-based clustering for time series data. *Data mining and knowledge Discovery*, 13(3):335–364, 2006.