

Visualizing and clustering financial portfolios using internal compositions¹

Domenico Vistocco²

Dipartimento di Scienze Economiche, Università di Cassino, e-mail: vistocco@unicas.it

Claudio Conversano

Dipartimento di Economia, Università di Cagliari, e-mail: conversa@unica.it

Abstract: A method to cluster financial portfolio according to their internal composition is introduced. Portfolio internal compositions are estimated through a style analysis model. Starting from such compositions, we cluster portfolios exploiting archetypal data analysis. Visualization tools are based on parallel coordinate plots and on asymmetric maps. The proposed procedure is based both on static and dynamic visualization tools. Final aim of such a procedure should be to compare and classify portfolios according to their similarities and to compare the evolution of the portfolio management along time.

Keywords: Style Analysis, Portfolio compositions, Archetypal analysis

1. Introduction

Style analysis, as originally proposed by Sharpe (12; 13), is a return based analysis aiming at decomposing portfolio performance with respect to the contribution of different portfolio constituents. Each sector is represented by an index whose returns are available. The model regresses portfolio returns on constituents' returns in order to decompose the portfolio performance with respect to each constituent. For a comprehensive review of style analysis models see (9).

Among the different types of style models, different constraints can be imposed on the coefficients; we refer to the so-called *strong style analysis*. In such a model the coefficients are imposed to be positive and to sum up to one. Moreover the model is estimated as a regression through the origin: the use of the double constraint and the absence of the intercept allow to interpret the regression coefficients in terms of portfolio weights and to estimate the internal composition of the portfolio (4; 5).

Aim of this paper is to investigate visualization tools useful for exploring and exploiting style analysis results. In practical applications, style models are indeed carried out on a set of financial portfolios or mutual funds and the style coefficients are used in order to rank portfolios with respect to different criteria. Effective use of the style models requires to analyze style coefficients for the different portfolios and to cross information on portfolio compositions with classical risk/return measures computed for the portfolios. Final aim of such a procedure should be to compare and classify portfolios according to

¹Research work of Domenico Vistocco is supported by Laboratorio di Calcolo ed Analisi Quantitative, Department of Economics, University of Cassino. Research work of Claudio Conversano is supported by the research funds awarded by University of Cagliari within the Young Researchers Start-Up Programme 2007.

²Address of correspondence: Via S. Angelo, Località Folcara, I-03043, Cassino (FR) - ITALY

their similarities and to compare the evolution of the portfolio management along time. A set of classical graphical representation is used for these purposes, exploiting both static and dynamic graphics. Furthermore, some representations based on the compositional data analysis (1; 2; 3) approach are inspected.

2. Style analysis and related visualization tools

Style analysis model regresses portfolio returns on the returns of a variety of investment class returns. The method thus identifies the portfolio style in the time series of its returns and of constituent returns (9).

Let us denote by $\mathbf{r}^{port}$ the random vector of portfolio returns along time and by $\mathbf{R}^{const}$ the matrix containing the returns along time of the i^{th} portfolio constituent on the i^{th} column ($i = 1, \dots, n$). Data are observed on T subsequent time periods. As above stated, the style analysis model regresses portfolio returns on the returns of the n constituents:

$$\mathbf{r}^{port} = \mathbf{R}^{const} \mathbf{w}^{const} + \mathbf{e} \quad \text{s.t.: } \mathbf{w}^{const} \geq 0, \mathbf{1}^T \mathbf{w}^{const} = 1.$$

The vector $\mathbf{e}$ denotes the expected tracking error of the portfolio, where $\mathbb{E}(\mathbf{R}^{const} \mathbf{e} = 0)$. The two constraints allow the coefficients to be exhaustive and non negative, so to interpret them in terms of compositional data: the estimated coefficients correspond then to portfolio weights. The interested reader is referred to (5) for the estimation procedure of the style coefficients $\mathbf{w}^{const}$. Starting from a set of F different portfolios observed along T subsequent times, it is possible to obtain F different compositions at a given time. The composition evolution can also be estimated by exploiting a subsampling procedure (10). Starting from such available data, we propose two classes of graphical tools:

1. tools for comparing the composition of several portfolios at a fixed time,
2. tools for visualizing the evolution of portfolio composition along time.

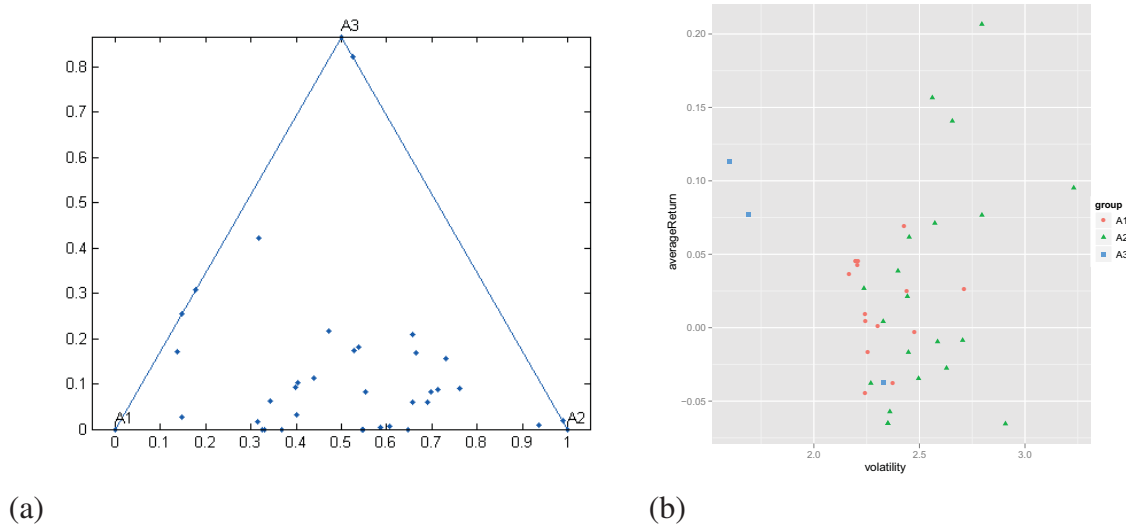
For the sake of space, we show in the following only some of the proposed graphical representations using a set of Italian mutual funds operating in the European equity sector. Original data were collected from Bloomberg on a daily basis from January 4, 1999 to September 30, 2002 (976 observations). In particular, we consider the time series of the returns of: (i) a set of 39 Italian equity funds; (ii) the Morgan Stanley Europe equity index (benchmark); (iii) the benchmark sub-indexes concerning 10 activity sectors, namely: Energy (ENR), Materials (MAT), Industrials (IND), Consumer Discretionary (CDIS), Consumer Staples (CSTA), Health Care (HC), Financial (FNCL), Information Technology (IT), Telecommunication Services (TEL), and Utilities (UTI). Each series is priced in Euro and is re-scaled on weekly basis (the Wednesday observation is considered as representative of the whole week, as common practice in financial data analysis). As a result, the final weekly time series are composed of 195 observations.

The final aim of an effective visualization tool for summarizing compositions is to obtain a proper and decipherable synthesis for the composition of a given portfolio at a fixed time. Through a simultaneous representation of a set of portfolios, it is also possible to compare differences in compositions and to cluster similar portfolios. Among the others, we propose to exploit archetypal analysis (6) in order to obtain a reduced set of artificial portfolios that can be used to best represent all the observed portfolios. Once some benchmark (archetypal) portfolios have been identified, the next step is to interpret such

portfolios through iconic/parallel coordinate plots and then to compare the observed portfolios with the archetypes. For a detailed treatment of the use of archetypal analysis in a benchmarking framework see (11).

Figure 1(a) depicts the triangular plot of the estimated compositions for the 39 Italian mutual funds. The triangular vertices represent the archetypal portfolios: portfolios falling near to a given archetype share with it a similar composition. Figure 1(b) shows the risk (x-axis)/return (y-axis) plot for the 39 portfolios. Point colours and shapes refer to the archetype each portfolio is nearest to. When more than three archetypes are used, the tri-

Figure 1: (a) *Triangular plot of the estimated compositions for the 39 Italian mutual funds in the archetype space.* (b) *Risk/Return plot for the 39 portfolios with reference to the archetypal classification.*



angular plot can be replaced using a parallel coordinate plot or an asymmetric map. The parallel coordinate plot preserve the original dimension of the data whereas the asymmetric map (8) is based on a dimension reduction technique, in particular correspondence analysis. The second tool allows a better view of the data if the percentage of explained variability of the reduced subspace is sufficient. For both the representations, the use of dynamic graphic tools such as selection, deletion and coloring, allow users to better understand portfolio composition and to classify different portfolios by considering their closeness to the archetypes.

When subsequent blocks of data are used for estimating the evolution of the portfolio composition, the aim of a proper visualization tool is to have a proper synthesis for the evolution of the composition of a portfolio. Also in this case, it is possible to compare different portfolios through a simultaneous representation. The time dimension can be added to archetypal analysis exploiting the moving archetype approach proposed by Cutler and Stone (7). Visualization tools extending the asymmetric map have been explored and results appears interesting.

3. Concluding remarks

From a practical viewpoint, the results of the style analysis model show which index or indexes a given portfolio actually acted most like. A post analysis of style models can be a valuable tool to compare and cluster portfolios sharing the same investment styles. Meanwhile, the use of graphical tools allow a better interpretation of style models and a prompt use of estimated coefficients to compare several portfolios at a fixed time and/or to analyze the evolution in terms of compositions. For these purposes, in this paper we propose an analysis strategy based on archetypes and on asymmetric maps.

Finally, we note that the style coefficients may be considered compositional data as well. Hence, suitable methods (1) developed for their analysis can be exploited to gain further insights into the data structure. In this direction, it is worth to investigate alternative visualization tools to take into account peculiar features of such a kind of data (3).

References

- [1] Aitchison J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44, 2, 139-177.
- [2] Aitchison J. (1983). Principal component analysis of compositional data. *Biometrika*, 70, 1, 57-65.
- [3] Aitchison, J., Greenacre, G. (2002) Biplots of compositional data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51, 4, 375-392.
- [4] Conversano, C., Vistocco, D. (2004). Model Based Visualization of Portfolio Style Analysis. In Antock J. (eds), *Proceedings of the International Conference "COMP-STAT 2004"*, Physica-Verlag, 815–822.
- [5] Conversano, C., Vistocco, D. (2009). Analysis of Mutual Fund Management Styles: A Modeling, Ranking and Visualizing Approach. *Journal of Applied Statistics*, in press.
- [6] Cutler A., Breiman L. Archetypal Analysis. *Technometrics*, 36:338–347, 1994.
- [7] Cutler A., Breiman L. Moving archetypes. *Physica D*, 1:1–16, 1997.
- [8] Greenacre M. *Correspondence analysis in practice*, Second edition, Chapman and Hall, 2007.
- [9] Horst, J.K. Ter, Nijman, T.H, De Roon, F.A. (2004). Evaluating Style Analysis. *Journal of Empirical Finance*, (11):29–51.
- [10] Politis, D.N., Romano, J.P. and Wolf, M. (1999). *Subsampling*. Springer-Verlag, New York.
- [11] G.C. Porzio, G. Ragozini, and D. Vistocco. On the use of archetypes as benchmarks. *Applied Stochastic Models in Business and Industry*, 24:419–437, 2008.
- [12] Sharpe, W. (1992). Asset Allocation: Management Styles and Performance Measurement. *The Journal of Portfolio Management*.
- [13] Sharpe, W. (1998). Determining a Fund's Effective Asset Mix. *Investment Management Review*, vol. 2, no. 6, (December):59–69.