Enhancing Interpretability in Factor Analysis by Means of Mathematical Optimization

Carrizosa, Emilio ; Guerrero, Vanesa; Romero Morales, Dolores ; Satorra, Albert

Enhancing Interpretability in Factor Analysis by Means of Mathematical Optimization

Emilio Carrizosa, Vanesa Guerrero, Dolores Romero Morales, and Albert Satorra

Journal article (Accepted manuscript*)

Uploaded to CBS Research Portal: November 2019

Enhancing interpretability in Factor Analysis by means of Mathematical Optimization *

Emilio Carrizosa[1], Vanesa Guerrero [†] [2], Dolores Romero Morales[3],

and Albert Satorra[4]

[1]Instituto de Matemáticas de la Universidad de Sevilla (IMUS), Seville, Spain

ecarrizosa@us.es

[2]Department of Statistics, Universidad Carlos III de Madrid, Getafe, Spain

vanesa.guerrero@uc3m.es

[3]Department of Economics, Copenhagen Business School, Frederiksberg, Denmark

drm.eco@cbs.dk

[4]Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain

albert.satorra@upf.edu

†Corresponding Author

Abstract

Exploratory Factor Analysis (EFA) is a widely used statistical technique to discover the structure of latent unobserved variables, called factors, from a set of observed variables. EFA exploits the property of rotation invariance of the factor model to enhance factors' interpretability by building a sparse loading matrix. In this paper, we propose an optimization-based procedure to give meaning to the factors arising in EFA by means of an additional set of variables, called *explanatory variables*, which may include in particular the set of observed variables. A goodness-of-fit criterion is introduced which quantifies the quality of the interpretation given this way. Our methodology also exploits the rotational invariance of EFA to obtain the best orthogonal rotation of the factors, in terms of the goodness-of-fit, but making them match to some of the explanatory variables, thus going beyond traditional rotation methods. Therefore, our approach allows the analyst to interpret the factors not only in terms of the observed variables, but in terms of a broader set of variables. Our experimental results demonstrate how our approach enhances interpretability in EFA, first in an empirical dataset, concerning volumes of reservoirs in California, and second in a synthetic data example.

Keywords: Exploratory Factor Analysis, Interpretability, Factor rotation, Explanatory variables, Mathematical Optimization

1 Introduction

Extracting knowledge from data, such as global underlying patterns or unusual behaviors, has become a crucial task for analysts to build models and improve decision making in many areas, such as Health Care, Bertsimas et al. (2016), Risk Management, Van Vlasselaer et al. (2017), Data Mining, Carrizosa and Romero Morales (2013); Olafsson et al. (2008), or Information Visualization, Carrizosa et al. (2018a,b); von Landesberger et al. (2017).

In this paper, we focus on latent variable models to uncover the hidden information enclosed in raw data, Anderson (2003); Bartholomew et al. (2011); Blei (2014). Specifically, we address the problem of interpreting latent variables arising in Exploratory Factor Analysis (EFA) via a Mathematical Optimization perspective. EFA is a multivariate statistical method that brings into light the effects of (a small number of) latent variables, called factors, on a set of observed variables, Mulaik (2009); Spearman (1904); Thurstone (1931). Given a set of p observed variables $\boldsymbol{y} \in \mathbb{R}^p$, the standard exploratory factor analysis model postulates that $\boldsymbol{y}$ is a linear function of random vectors, $\boldsymbol{f} \in \mathbb{R}^r$, $r << p$, the factors, and an error term $\boldsymbol{\epsilon} \in \mathbb{R}^p$ satisfying

$$\boldsymbol{y} = \boldsymbol{\Lambda} \boldsymbol{f} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times r}$ is the so-called loading matrix. The components of $\boldsymbol{f}$ are assumed to be standardized and $\boldsymbol{f}$ and $\boldsymbol{\epsilon}$ uncorrelated vectors.

The analysts regularly have to deal with the non-straightforward task of providing

1

meaningful explanations, based on their expertise, to what the factors define, namely, they have to interpret the factors. Therefore, many domains, such as Medicine, Economics or Social Sciences, would benefit from approaches, as the one presented in this work, that provide new tools to improve on the interpretation of the factors. For the relevance of factor interpretation, see Fabrigar et al. (1999); Martínez et al. (1998); Taeb and Chandrasekaran (2018); Yamamoto et al. (2017). In this paper, an empirical data set of water reservoirs of California (see Taeb et al. (2017)) that requires factor interpretation will be used for illustration of the methods we advocate.

It is well known that the factor model in (1) is not identifiable, Shapiro (1985). In other words, model (1) also holds for $\boldsymbol{\Lambda^*} = \boldsymbol{\Lambda M^\top}$ and $\boldsymbol{f^*} = \boldsymbol{Mf}$, where $\boldsymbol{M} \in \mathbb{R}^{r \times r}$ is any orthogonal matrix, i.e. $\boldsymbol{MM^\top} = \boldsymbol{M^\top M} = \boldsymbol{\mathcal{I}_r}$, $\boldsymbol{\mathcal{I}_r}$ being the r-dimensional identity matrix, and thus $\boldsymbol{f^*}$ is a rotation of $\boldsymbol{f}$. This property is known as the rotational invariance of the factor model, which turns out to be very useful for interpretation purposes, although might fail in some situations, Browne (2001); Hattori et al. (2017); Kaiser (1958); Ročková and George (2016).

The goal of classic rotation methods is to find a rotation matrix that produces an interpretable factor loading matrix. Interpretability is explored through sparsity, which is used as a proxy. A sparse loading matrix has *few* non-zero elements, and then, each factor in $\boldsymbol{f}$ can be interpreted in terms of few of the components of $\boldsymbol{y}$. Thus, rotation matrices $\boldsymbol{M}$ that yield a sparse $\boldsymbol{\Lambda^*} = \boldsymbol{\Lambda M^\top}$ are sought. Nevertheless, the sparsity of a loading matrix does not necessarily give interpretable factors, as discussed in Yamamoto

2

et al. (2017). On top of that, different rotation criteria may yield different conclusions, in terms of interpretability, on the very same data, see Sass and Schmitt (2010).

After obtaining the rotated loading matrix $\mathbf{\Lambda}^*$, the user has to analyze it, and interpret the rotated factors $\boldsymbol{f}^*$ matching them with the variables $\boldsymbol{y}$ involved in the factor model (1) based on his/her expertise. We refer the reader to the work of Browne (2001), and references therein, for an overview of rotation methods in EFA. Other approaches in the literature seeking sparsity in $\mathbf{\Lambda}$, but at the expense of a poorer fitting of the model to the observed data, involve, for instance, regularization techniques, (Choi et al., 2010; Trendafilov et al., 2017; Witten et al., 2016; Yamamoto et al., 2017), Bayesian frameworks, (Kaufmann and Schumacher, 2017; Zhao et al., 2016), clustering approaches, (Adachi and Trendafilov, 2018; Uno et al., 2016) or regression models, (Carvalho et al., 2008).

Instead of using sparsity as a proxy of interpretability, alternative approaches have recently been proposed in the literature, e.g. Fan et al. (2016); Li and Jung (2017); Taeb et al. (2017); Taeb and Chandrasekaran (2018). Such methods make use of an additional set of observed variables, called *explanatory variables*, assumed to be meaningful to the user and related to the latent and unobserved phenomena hidden in the dataset under study. Either one can identify many of these explanatory variables in the hope of some of them being related to the factors (exploratory approach), or the user has high expertise in the data and knows what explanatory variables to use (confirmatory approach). For instance, in economic applications there is usually available data about social, geographical or political indicators that may explain the factors and do not necessarily take part in the

3

construction of the factors.

The methodologies proposed in Fan et al. (2016); Li and Jung (2017); Taeb et al. (2017); Taeb and Chandrasekaran (2018) fit a model in which the explanatory variables are integrated together with the factors. Thus, the variability in $\boldsymbol{y}$ is split into what can be explained (interpreted) through the explanatory variables; the latent phenomena which remain unexplained after adding those additional variables; and an error term. Therefore, the user has more difficulties in monitoring the interpretation procedure, since there might be latent phenomena which still remain unexplained after incorporating the exaplanatory variables in the model.

In order to overcome the difficulties of existing methodologies, in our approach we give more freedom to the users to seek an assignment of the explanatory variables to the factors. We propose an optimization-based approach to obtain a one-to-one correspondence between the factors and the explanatory variables, which thus allows an straightforward interpretation of the factors arising in model (1). Our aim is to optimally match (groups of) the explanatory variables to an orthogonal rotation of the factors. This way the interpretability of the factors is enhanced: the user knows the meaning of the explanatory variables involved and to which factors they are matched. In order to obtain this rotation matrix and assess the quality of the matching, the correlations between the rotated factors and the explanatory variables are maximized, ensuring that the weakest explanation (the smaller correlation) is the strongest possible (maximum). In our empirical example, we show how the factors that affect the volumes of major water reservoirs in California are

4

matched, and thus interpreted, with physical and economic explanatory variables, as done in Taeb et al. (2017) for a graphical model.

Therefore, the methodology proposed in this paper is not a purely rotation method but a one-step approach to interpret the factors arising in a factor model using simultaneously a rotation and a matching with a set of explanatory variables giving meaning to them. In fact, it differs from the traditional rotation methods: whereas traditional rotation methods aim to find a simple structure in the loading matrix to interpret the factors by means of the variables used to extract them, our methodology seeks a rotation that makes them match with an additional set of variables. Our approach does not take care of the loading matrix structure, neither the relation with the variables used to obtain the factors, unless they are considered as explanatory as well.

The remainder of the paper is structured as follows. In Section 2, we introduce the key ingredients of our approach. Section 3 presents our model, in which the rotation and the matching are optimized. Section 4 illustrates with both empirical and synthetic data the methodology proposed. Finally, Section 5 concludes the paper with some comments and future research lines.

2   Preliminaries

In this work, it is assumed that the researcher has already fitted the model (1) and the factor scores have been computed using any of the many procedures in the literature, e.g.

DiStefano et al. (2009); Gorsuch (1983); Grice (2001). Let $\boldsymbol{F} \in \mathbb{R}^{n \times r}$ be a matrix of factor scores, where n is the number of cases and r the number of factors. Thus, the j-th column of $\boldsymbol{F}$, referred as $\boldsymbol{F}_j$, contains the scores of j-th factor for each individual in the sample, $j \in J = \{1, \ldots, r\}$.

Establishing a one-to-one correspondence between factors and explanatory variables might be unnatural from the practical point of view, since it is unlikely that unobserved phenomena (factors) are similar to observed events (explanatory variables). Therefore, in order to be able to interpret the factors by collections of phenomena, we assume that explanatory variables are grouped according to, for instance, their subject or nature. For instance, in the empirical example in Section 4 concerning the volumes of water reservoirs in California, Taeb et al. (2017) considered groups of both physical and economic variables, which involve measurements of specific variables within each group, temperatures or snow melting in the first case and a consumer price index or number of agricultural workers. Thus, each factor could be interpreted (matched) either as a combination of physical or economic variables. Without loss of generality, these groups of variables could be made up of just one explanatory variable as well. For instance, each factor in the reservoirs example could be interpreted as temperatures or a consumer index as singletons. Thus, let $\mathcal{C}_i$ be a group of $|\mathcal{C}_i|$ explanatory variables, where $|\cdot|$ denotes the cardinality (number of elements) of a set, and $|\mathcal{C}_i| \geq 1$, $i \in I = \{1, \ldots, q\}$, namely each group contains at least one explanatory variable. These variables have been measured over the same cases used in EFA to obtain the factors' scores. Therefore, we consider that there are q data matrices

6

available $\boldsymbol{X}_i \in \mathbb{R}^{n \times |\mathcal{C}_i|}$. We will refer to $\mathcal{C}_i$, $i \in I$, as the *interpretability clusters*, which are thus defined as groups of the so-called explanatory variables.

We seek a matching between the interpretability clusters and the rotated factors through an orthogonal matrix $\boldsymbol{M}$, i.e. $\boldsymbol{F}^* = \boldsymbol{M}\boldsymbol{F}$, where $\boldsymbol{M}\boldsymbol{M}^\top = \boldsymbol{M}^\top \boldsymbol{M} = \boldsymbol{\mathcal{I}}_r$, $\boldsymbol{\mathcal{I}}_r$ being the r-dimensional identity matrix, optimizing a goodness of fit criterion based on the coefficient of determination. See Guadagnoli and Velicer (1991) and Kaufman and Rousseeuw (1990) for alternative criteria. Let us now introduce some notation, including the one for the matching matrix which will give us the recipe to interpret the factors.

Let $\boldsymbol{H}$ be a 0–1 matrix of size $q \times r$ representing the matching, which is composed by elements h_{ij}, $i \in I$, $j \in J$, such that

$$h_{ij} = \begin{cases} 1 & \text{if the interpretability cluster } \mathcal{C}_i \text{ is matched with the rotated factor } \boldsymbol{F}_j^* \\ 0 & \text{otherwise.} \end{cases}$$

In other words, when $h_{ij} = 1$, the j-th rotated factor is interpreted as the i-th interpretability cluster. For instance, if $\mathcal{C}_i$ is made up of meteorological conditions and $h_{ij} = 1$, then the rotated factor $\boldsymbol{F}_j^*$ is interpreted (or named) as meteorological conditions.

It is assumed that there is at least one match in $\boldsymbol{H}$, this is $h_{ij} = 1$ for at least one $i \in I$ and $j \in J$. Thus, the space of possible matchings between rotated factors and interpretability clusters, $\mathcal{H}$, is defined as the set of all 0–1 non-zero matrices of size $q \times r$.

Moreover, we need to define a measure to assess the quality of the matching of the interpretability cluster $\mathcal{C}_i$ with the rotated factor $\boldsymbol{F}_j^*$. In this paper, we use the coefficient of determination. Let $R_{ij}^2(\boldsymbol{M})$ denote the coefficient of determination between the interpretability

cluster $\mathcal{C}_i$ and the rotated factor $\boldsymbol{F}_j^*$. Observe that $R_{ij}^2(\boldsymbol{M})$ depends on the rotation matrix $\boldsymbol{M}$ used to rotate the factors, thus for different $\boldsymbol{M}$ the value of $R_{ij}^2(\boldsymbol{M})$ changes. In the best case scenario, this is a perfect linear relationship between the j-th rotated factor and the i-th interpretability cluster, $R_{ij}^2(\boldsymbol{M}) = 1$ for all $h_{ij} = 1$, and thus we would like to have $R_{ij}^2(\boldsymbol{M})$ as large as possible for those i and j with $h_{ij} = 1$. This is achieved by making large the minimum of those $R_{ij}^2(\boldsymbol{M})$, this is maximizing the minimum coefficient of determination between the matched interpretablity clusters and the rotated factors. Therefore, we define the goodness of fit criterion $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$, which measures the quality of the interpretation as

$$\mathcal{S}(\boldsymbol{M}, \boldsymbol{H}) = \min_{\substack{h_{ij}=1 \\ i \in I, j \in J}} R_{ij}^2(\boldsymbol{M}). \tag{2}$$

Clearly, $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H}) \in [0, 1]$, and the case $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H}) = 1$ corresponds to the best case scenario, namely a perfect interpretation of the factors according to this criterion. Observe that, in this case, all the matched factors and interpretability clusters (those for which $h_{ij} = 1$, $i \in I$, $j \in J$) are strongly related ($R_{ij}^2(\boldsymbol{M}) = 1$, for all $i \in I$, $j \in J$ such that $h_{ij} = 1$). However, if there is a $i \in I$ and $j \in J$ such that $h_{ij} = 1$ and $R_{ij}^2(\boldsymbol{M}) < 1$, then $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H}) < 1$ and the quality of the interpretation would be, in general, poorer. The bounds in $\mathcal{S}$ defining good or bad interpretations depend on the empirical application, and therefore cannot be defined in advance.

In order to illustrate the criterion $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ in (2) and how it depends on the choice of the rotation matrix $\boldsymbol{M}$ and the matching $\boldsymbol{H}$, its value is analyzed for all the combinations of

$$\boldsymbol{M}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \boldsymbol{M}_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \boldsymbol{M}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

$$\boldsymbol{H}_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \boldsymbol{H}_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \boldsymbol{H}_3 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} \quad \boldsymbol{H}_4 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Figure 1: Different rotation and matching matrices for a simulated example with $r = 2$ factors and $q = 4$ interpretability clusters.

three different rotations and four matchings through a simulated example. Let us consider $r = 2$ factors $\boldsymbol{F} = (\boldsymbol{F}_1, \boldsymbol{F}_2)$, arising from model (1), and $q = 4$ interpretability clusters $\mathcal{C}_1, \ldots, \mathcal{C}_4$, which contain one interpretability variable each. Let us assume that the space of possible solutions to obtain the best $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ is made up of the three rotation matrices $\boldsymbol{M}_1$, $\boldsymbol{M}_2$, $\boldsymbol{M}_3$ and the four matchings $\boldsymbol{H}_1, \boldsymbol{H}_2, \boldsymbol{H}_3, \boldsymbol{H}_4$, namely $\mathcal{H} = \{\boldsymbol{H}_1, \boldsymbol{H}_2, \boldsymbol{H}_3, \boldsymbol{H}_4\}$, shown in Figure 1. Observe that the matching matrix $\boldsymbol{H}_1$ would interpret the first rotated factor as the three interpretability clusters $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_4$ and the second rotated factor as $\mathcal{C}_3$; matching $\boldsymbol{H}_2$ would interpret the first rotated factor as $\mathcal{C}_4$ and the second as $\mathcal{C}_3$; in $\boldsymbol{H}_3$ second factor remains unexplained while the first rotated factor is interpreted with $\mathcal{C}_4$; finally, in $\boldsymbol{H}_4$ the first rotated factor is interpreted as the two interpretability clusters $\mathcal{C}_1$ and $\mathcal{C}_2$ and the second rotated factor as $\mathcal{C}_1$ and $\mathcal{C}_3$.

9

The value of $\mathcal{S}$ in (2) is obtained for the twelve combinations of rotation matrices and matchings in Figure 1. Different rotations induce different coefficients of determination between the factors and the interpretability clusters. For visualization purposes these coefficients are depicted as heatmaps. The darker the color is, the larger the coefficient. Figure 2 contains the the heatmaps representing $R_{ij}^2(\boldsymbol{M}_k)$, $k = 1, 2, 3$ for the rotations considered in our example. Observe that each column in Figure 2 reports the same heatmap (same rotation matrix) for the four different matchings. Ideally, all pairs with $h_{ij} = 1$ should have a dark entry in the heatmap. A *one* sign ($\boldsymbol{1}$) is added to those cells in the heatmaps with $h_{ij} = 1$. The star sign ($*$) indicates in which of those cells the minimum coefficient of determination is achieved, yielding the value of $\mathcal{S}$. This cell is expected to have darkest color. Moreover, the values of $\mathcal{S}$ in (2) for the three rotations and four matchings are reported in Figure 2. We observe the effects on changing rotations and matchings. The value of $\mathcal{S}$ is close to 0 for rotation $\boldsymbol{M}_1$ and matchings $\boldsymbol{H}_1$, $\boldsymbol{H}_2$ and $\boldsymbol{H}_4$, and rotation $\boldsymbol{M}_2$ and matchings $\boldsymbol{H}_1$ and $\boldsymbol{H}_4$ (Figure 2 (a), (b), (d), (j) and (k)), increases to 0.20, 0.38 0.41 and 0.67 for $\boldsymbol{M}_1$ and matching $\boldsymbol{H}_3$, and $\boldsymbol{M}_3$ and matchings $\boldsymbol{H}_4$, $\boldsymbol{H}_1$, and $\boldsymbol{H}_2$, respectively (Figures 2 (g), (l), (c) and (f)). Finally, $\mathcal{S}$ achieves 0.80 for rotation $\boldsymbol{M}_2$ and matchings $\boldsymbol{H}_2$ and $\boldsymbol{H}_3$ (Figure 2 (e) and (h)) and 0.90 for $\boldsymbol{M}_3$ and $\boldsymbol{H}_3$ (Figure 2 (i)). Therefore, according to (2), the best interpretation of factors $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ by means of the four interpretability clusters considered in this example is given by the rotation matrix $\boldsymbol{M}_2$ and the matching in $\boldsymbol{H}_2$ (Figure 2 (e)), as pointed out in the previous paragraph, since the largest value of $\mathcal{S}$ is achieved for that combination of rotation and matching among the different choices

10

available. Therefore, the best meaning to $\boldsymbol{F}_1$ would be provided by interpretability cluster $\mathcal{C}_4$, and for $\boldsymbol{F}_2$ by $\mathcal{C}_3$. However, if just factor $\boldsymbol{F}_1$ is to be interpreted, namely matching $\boldsymbol{H}_3$ is considered, a better value of $\mathcal{S}$ is obtained, 0.90, using the rotation $\boldsymbol{M}_3$ instead (Figure 2 (i)).

The example described above illustrates the strong effect that the choice of the rotation matrix $\boldsymbol{M}$ and matching $\boldsymbol{H}$ have in the criterion $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$. In what follows we propose an optimization-based methodology that automatically obtains the best $\boldsymbol{M}$ and $\boldsymbol{H}$ for the interpretability clusters at hand.

3   The model

If the matching $\boldsymbol{H} \in \mathcal{H}$ that makes the interpretation of the factors in terms of the interpretability clusters were given, i.e. known in advance, our aim would be to obtain the orthogonal rotation of the factors, $\boldsymbol{M}$, such that (2) is maximized. In other words, if the user is able to match the interpetability clusters with the factors because of his/her expertise, our aim would be to obtain the best rotation of the factors which maximizes the minimum coefficient of determination between the matched rotated factors and clusters. This yields the nonlinear continuous optimization problem $(InRot)$:

$$\max_{\boldsymbol{M}} \quad \mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$$

$$\text{s.t.} \quad \boldsymbol{M} \cdot \boldsymbol{M}^\top = \mathcal{I}_r. \qquad\qquad (InRot)$$

Obtaining the globally optimal solution to problem $(InRot)$ is challenging due to its nonlinearities and non-convex constraints, Wen and Yin (2013). Nevertheless, even locally optimal solutions to this problem might outperform the traditional rotation techniques in EFA, e.g. the rotation methods described in Browne (2001) and Hattori et al. (2017), in the sense that it has indeed the benefit of providing a straightforward interpretation to the factors through the matching $\boldsymbol{H}$ with (groups of) explanatory variables.

However, the approach stated above requires to guess the matching between the interpretability clusters and the factors. In practice, assuming $\boldsymbol{H}$ to be known is unrealistic, although some structural properties may by available. Some examples of this structure consist of constraining the total number of factors which are interpreted; the maximum number of interpretability clusters than can be used to interpret a factor or the maximum number of factors than can be interpreted with each of the interpretability clusters. Such information can be incorporated into the search of the best $\boldsymbol{H}$ by considering a subset $\tilde{\mathcal{H}}$ of the set of all possible matchings $\mathcal{H}$, namely $\tilde{\mathcal{H}} \subset \mathcal{H}$. Thus, the matching $\boldsymbol{H} \in \tilde{\mathcal{H}}$ is, in general, unknown and should be decided together with the rotation matrix $\boldsymbol{M}$. Therefore, the mixed 0–1 nonlinear optimization problem $(InRotMat)$, which together decides the best rotation matrix $\boldsymbol{M}$ and the best matching $\boldsymbol{H} \in \tilde{\mathcal{H}}$, is stated as:

$$\max_{\boldsymbol{M}, \boldsymbol{H}} \quad \mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$$

12
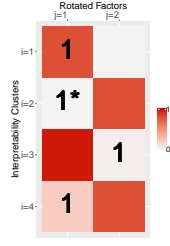
(a) $\mathcal{S}(\boldsymbol{M}_1, \boldsymbol{H}_1) = 0.00$
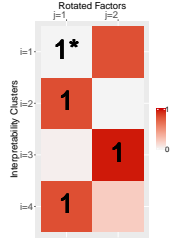
(b) $\mathcal{S}(\boldsymbol{M}_2, \boldsymbol{H}_1) = 0.00$
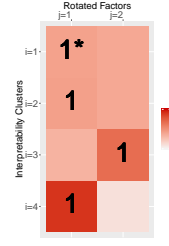
(c) $\mathcal{S}(\boldsymbol{M}_3, \boldsymbol{H}_1) = 0.41$
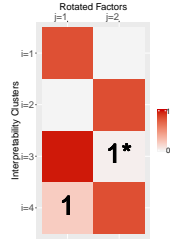
(d) $\mathcal{S}(\boldsymbol{M}_1, \boldsymbol{H}_2) = 0.03$
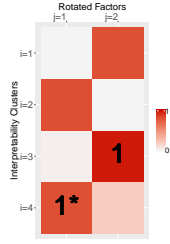
(e) $\mathcal{S}(\boldsymbol{M}_2, \boldsymbol{H}_2) = 0.80$
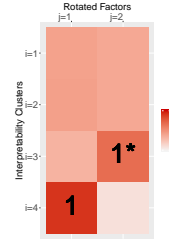
(f) $\mathcal{S}(\boldsymbol{M}_3, \boldsymbol{H}_2) = 0.67$
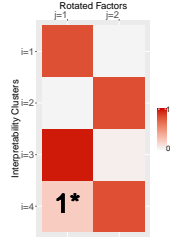
(g) $\mathcal{S}(\boldsymbol{M}_1, \boldsymbol{H}_3) = 0.20$
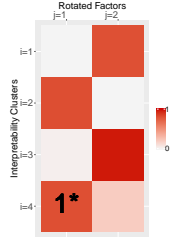
(h) $\mathcal{S}(\boldsymbol{M}_2, \boldsymbol{H}_3) = 0.80$
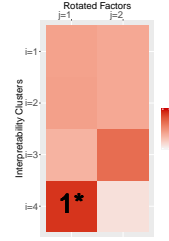
(i) $\mathcal{S}(\boldsymbol{M}_3, \boldsymbol{H}_3) = 0.90$
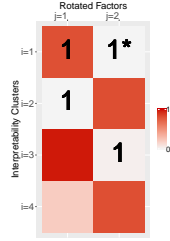
(j) $\mathcal{S}(\boldsymbol{M}_1, \boldsymbol{H}_3) = 0.00$
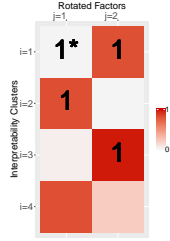
(k) $\mathcal{S}(\boldsymbol{M}_2, \boldsymbol{H}_3) = 0.00$
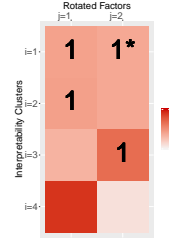
(l) $\mathcal{S}(\boldsymbol{M}_3, \boldsymbol{H}_3) = 0.38$

Figure 2: Illustration of $\mathcal{S}$ in (2) for different rotations and matchings.

$$\text{s.t.} \qquad \boldsymbol{M} \cdot \boldsymbol{M}^\top = \boldsymbol{\mathcal{I}}_r,$$

$$\boldsymbol{H} \in \tilde{\mathcal{H}}. \qquad\qquad (InRotMat)$$

Depending on the structure for the matching impose in $\tilde{\mathcal{H}}$, the value of $\mathcal{S}$ can be affected. For instance, if it is required to interpret just one of the factors, then the optimal value of $\mathcal{S}$ will be the largest coefficient of determination between the interpretability clusters and the rotated factors. In general, when the number of factors to be interpreted increases, the optimal value of $\mathcal{S}$ decreases, since its value is such that the worst of those coefficients is maximized. Based on this decrease, the user can decide by looking at the optimal objective function when to stop, i.e., how many factors to interpret in order to prevent serious erosion of the solution to the point where it might affect substantive interpretation of the solution.

To solve the mixed 0–1 nonlinear optimization problem $(InRotMat)$, we first rewrite it as

$$\max_{\boldsymbol{M},\boldsymbol{H},z} \quad z$$

$$\text{s.t.} \qquad z \leq R_{ij}^2(\boldsymbol{M})h_{ij} + (1 - h_{ij}) \quad i \in I, j \in J,$$

$$\boldsymbol{M} \cdot \boldsymbol{M}^\top = \boldsymbol{\mathcal{I}}_r, \qquad\qquad (3)$$

$$\boldsymbol{H} \in \tilde{\mathcal{H}}.$$

To optimally solve problem (3) is difficult due to its combinatorial structure and its non-convex constraints. Unfortunately, the software available to handle mixed integer

14

nonlinear optimization problems are much less powerful than the linear integer or nonlinear continuous counterparts. Therefore, we propose an alternating algorithm to obtain good solutions to problem (3) which consists of, in turn, fixing the rotation matrix M to a certain value and find the best possible matching H for that M by solving the resulting mixed integer linear problem from (3) in which M is data and H is a decision variable; then, H is fixed to the so-obtained value and the best possible rotation M for that H is found by solving now the nonlinear continuous optimization problem from (3) in which H is data and M is a decision variable. Observe that this second step corresponds to solving problem ($InRot$). These two alternating steps are repeated until a stopping criterion is satisfied. In order to avoid getting stuck in bad local optima, a multistart strategy is used, i.e., the process above is repeated for several randomly selected initial solutions.

4   Simulation and empirical results

In order to illustrate the usefulness of the methodology described in Section 3, we show how factors can be interpreted in terms of the explanatory variables (measured through the goodness of fit criterion stated in (2)) using both an empirical dataset as well as a synthetic one.

4.1   Implementation

The computational experiments have been carried out using R, R Development Core Team (2008), on a PC Intel® Core™ i7-7700, 16GB of RAM. First, the scores for r factors have been generated using the `fa` function from package 'psych' on the observed variables $\boldsymbol{Y} \in \mathbb{R}^{n \times p}$. Therefore, the so-obtained matrix $\boldsymbol{F} \in \mathbb{R}^{n \times r}$ contains (columnwise) the r standardized and uncorrelated factors. In order to give meaning to these factors, $q \geq r$ interpretability clusters, $\mathcal{C}_1, \ldots, \mathcal{C}_q$, are used. Each of these clusters is made up of a single explanatory variable, namely $\mathcal{C}_1 = \{\boldsymbol{X}_1\}, \ldots, \mathcal{C}_q = \{\boldsymbol{X}_q\}$ and, therefore, $|\mathcal{C}_i| = 1$, for $i = 1, \ldots, q$. Second, problem $(InRotMat)$ is solved to obtain the best rotation and matching matrices, $\boldsymbol{M}$ and $\boldsymbol{H}$ respectively, that maximize the goodness of fit criterion stated in (2), $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$. The set $\tilde{\mathcal{H}}$ in which the matching is sought requires that the r factors are interpreted, namely there must be at least one 1 in each column of $\boldsymbol{H}$; each factor is interpreted with at most one interpretability cluster, namely there is at most one 1 in each column of $\boldsymbol{H}$; and finally, each interpretability cluster is used to interpret at most one factor, namely there is at most one 1 in each row. As described at the end of Section 3, an alternating strategy has been performed setting as a stopping criterion a maximum number of iterations equal to 5. The `nloptr` optimization routine, included in the R package with the same name, Johnson (2008), and solver GLPK in the 'ROI' package, Theußl et al. (2017), have been used to solve the optimization problems. This procedure is embedded into a multistart routine with 10 different starting solutions, namely rotation and matching matrices to start the alternating algorithm. Finally, the goodness of our methodology is illustrated using same heatmaps as in Section 2 and the value of $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ for the best

16

rotation $\boldsymbol{M}$ and matching $\boldsymbol{H}$ obtained.

For the sake of completeness, we show in the Appendix the loading matrix $\boldsymbol{\Lambda}$ obtained from estimating model (1) in our empirical and synthetic datasets, and its orthogonal rotated counterparts $\boldsymbol{\Lambda}^*$ using some of the methodologies in the `fa` function in R, namely `varimax`, `quartimax`, `bentlerT`, `equamax`, `varimin` and `geominT`. Besides these matrices, we include the rotated loading matrix obtained using the rotation matrix obtained with our methodology. We point out that our methodology does not seek a simple structure in the loading matrix as the other methodologies do.

4.2 An empirical dataset

Our methodology is illustrated in the empirical dataset called `reservoir` in Taeb et al. (2017) concerning volumes of major reservoirs in California. We point out that our approach is different from the one by Taeb et al. (2017), since they develop a latent graphical model `reservoir` dataset. The data has been downloaded from https://github.com/armeentaeb/WRR-Reservoir. There are $p = 55$ observed variables $\boldsymbol{Y}$, each of them corresponding to a reservoir. Two factors ($r = 2$) are extracted, $\boldsymbol{F} = (\boldsymbol{F}_1, \boldsymbol{F}_2)$, and interpreted, using $q = 7$ physical and economic explanatory variables, namely `Palmer Drought Severity Index` $(\boldsymbol{X}_1)$, `Colorado river discharge` $(\boldsymbol{X}_2)$, `state-wide consumer price index` $(\boldsymbol{X}_3)$, `hydroelectric power` $(\boldsymbol{X}_4)$, `state-wide number of agricultural workers` $(\boldsymbol{X}_5)$, `Sierra Nevada snow pack` $(\boldsymbol{X}_6)$ and `temperature` $(\boldsymbol{X}_7)$. As mentioned in Section 4.1, the interpretability clusters are defined as these explanatory variables, namely

17

$\mathcal{C}_i = \{\boldsymbol{X}_i\}$, $i = 1, \ldots, 7$. The data has been cleaned following the guidelines in Taeb et al. (2017), yielding $n = 118$ observations. Each observation corresponds to monthly measurements from January 2004 to October 2013.

Following the same terminology as the example stated in Section 2, Figure 3 depicts the heatmap of the solution obtained by our methodology and the value of $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ for the best rotation $\boldsymbol{M}$ and matching $\boldsymbol{H}$ obtained. Recall that this heatmap contains the coefficients of determination between the explanatory variables and the rotated factors according to the best rotation obtained in which a *one* ($\boldsymbol{1}$) is added to those cells in the heatmaps with $h_{ij} = 1$, and a star sign (*) indicates in which of those cells the minimum coefficient of determination is achieved.

Our methodology is able to provide the major water reservoirs in California latent factors with an interpretation: the first one is interpreted as the `hydroelectric power` and the second as the `temperature`. The quality of the interpretation is supported by a goodness of fit criterion $\mathcal{S}$ equal to 0.649. Note that although the coefficient of determination between the second rotated factor and explanatory variable `Sierra Nevada snow pack` is rather high (dark cell), the best interpretation of the second rotated factor is given by explanatory variable `temperature`, i.e., the one in which the coefficient of determination between the rotated factor and the explanatory variable is the highest.

18
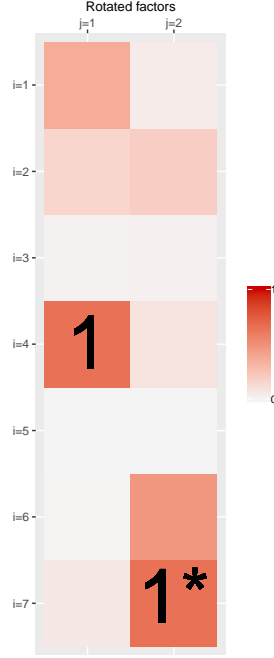
Figure 3: Heatmap for `reservoir` dataset in Taeb et al. (2017) where $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H}) = 0.649$.

4.3   A synthetic dataset

In this section, using a synthetic dataset, we illustrate how the quality of the interpretation decreases when the relationship between the explanatory variables and the factors weakens. In what follows, we describe how the data have been generated and detail the experiments which have been carried out.

Regarding the data generation, the scores for three factors ($r = 3$) have been generated for a sample of $n = 100$ using the `fa` function from package 'psych' in R on $p = 15$ multivariate normally distributed random variables $\boldsymbol{Y}$. Therefore, the so-obtained matrix of factor scores $\boldsymbol{F} \in \mathbb{R}^{100 \times 3}$ contains (columnwise) three standardized and uncorrelated

factors. In order to give meaning to these factors, $q = 30$ explanatory variables, namely $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{30}$, have been randomly generated. As mentioned in Section 4.1, the interpretability clusters are defined as these explanatory variables, namely $\mathcal{C}_i = \{\boldsymbol{X}_i\}$, $i = 1, \ldots, 30$. Whereas $\boldsymbol{X}_i$, $i = 4, \ldots, 30$ are multivariate normally distributed random variables with zero mean and a randomly generated covariance matrix, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ and $\boldsymbol{X}_3$ are produced as $\boldsymbol{X}_i = \boldsymbol{F}\mathcal{M} + \sigma\varepsilon$, $i = 1, \ldots, 3$, where $\mathcal{M}$ is a random 3-dimensional orthonormal matrix and $\varepsilon \sim \mathcal{N}(0, 1)$. Varying σ from 0 to 1, with a step of 0.25, 5 cases of explanatory variables are obtained. For $\sigma = 0$, the first three explanatory variables have been generated to be the rotation of the factors according to matrix $\mathcal{M}$. Our methodology is able to successfully detect this interpretation (except for a permutation of the factors), as it is shown below. For the remaining cases of σ, the relationship between the explanatory variables and the factors weakens, and therefore, the interpretation of the factors in terms of these variables gets worse, yielding smaller values of the goodness of fit criterion $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$, as shown below.

Figure 4 depicts, for each value of σ, the heatmap of the solution obtained by our methodology, namely the solution of $(InRotMat)$, and the value of $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ for the best rotation $\boldsymbol{M}$ and matching $\boldsymbol{H}$ obtained. Recall that the largest plus sign indicates the cell where the minimum is achieved. Observe that, in general, the larger σ is, the smaller $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ (the lighter the pixels where $h_{ij} = 1$). The first purpose of the synthetic experiment was to find the perfect fit between the rotated factors and the explanatory variables for $\sigma = 0$. This is shown in Figure 4 (a), in which $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H}) = 1$, since the

20

rotated factors are exactly the first three explanatory variables. The second purpose of the experiment was to illustrate how the quality of the interpretation worsen when σ grows. We can see that the value of the goodness of fit criterion $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ decreases from 0.939 in Figure 4 (b) to 0.460 in Figure 4 (e) whilst σ increases up to 1. This fact agrees with the way in which data was generated and illustrates the influence of the (non)existence of a relationship between the factors and the explanatory variables for interpretation purposes. In summary, our results show that the rotated factors are matched with the explanatory variables $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ and $\boldsymbol{X}_3$, which are indeed the ones that most resemble the factors, and therefore, the ones that are the most useful to interpret them.

The previous experiment has been repeated for 100 instances in which σ varies in a range from 0 to 5 with a step of 0.25, yielding 21 cases. Table 1 contains a statistical summary of the results obtained for each value of σ. The first column in Table 1 contains the values of σ, the second reports the averages $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ and the last one the standard deviations. Observe that, as expected, the larger σ is, the smaller $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$. This makes sense due to the way the explanatory variables have been generated: for small values of σ there are three explanatory variables which are practically equal to a rotation of the factors to be interpreted. Figure 5 illustrates graphically the results presented in Table 1, in which the black line represents the average value and the red band represents the maximum and minimum optimal values of $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ in the 100 instances for each σ.

(a) $\mathcal{S}(\boldsymbol{H}, \boldsymbol{M}) = 1$ (b) $\mathcal{S}(\boldsymbol{H}, \boldsymbol{M}) = 0.939$ (c) $\mathcal{S}(\boldsymbol{H}, \boldsymbol{M}) = 0.787$ (d) $\mathcal{S}(\boldsymbol{H}, \boldsymbol{M}) = 0.612$ (e) $\mathcal{S}(\boldsymbol{H}, \boldsymbol{M}) = 0.460$
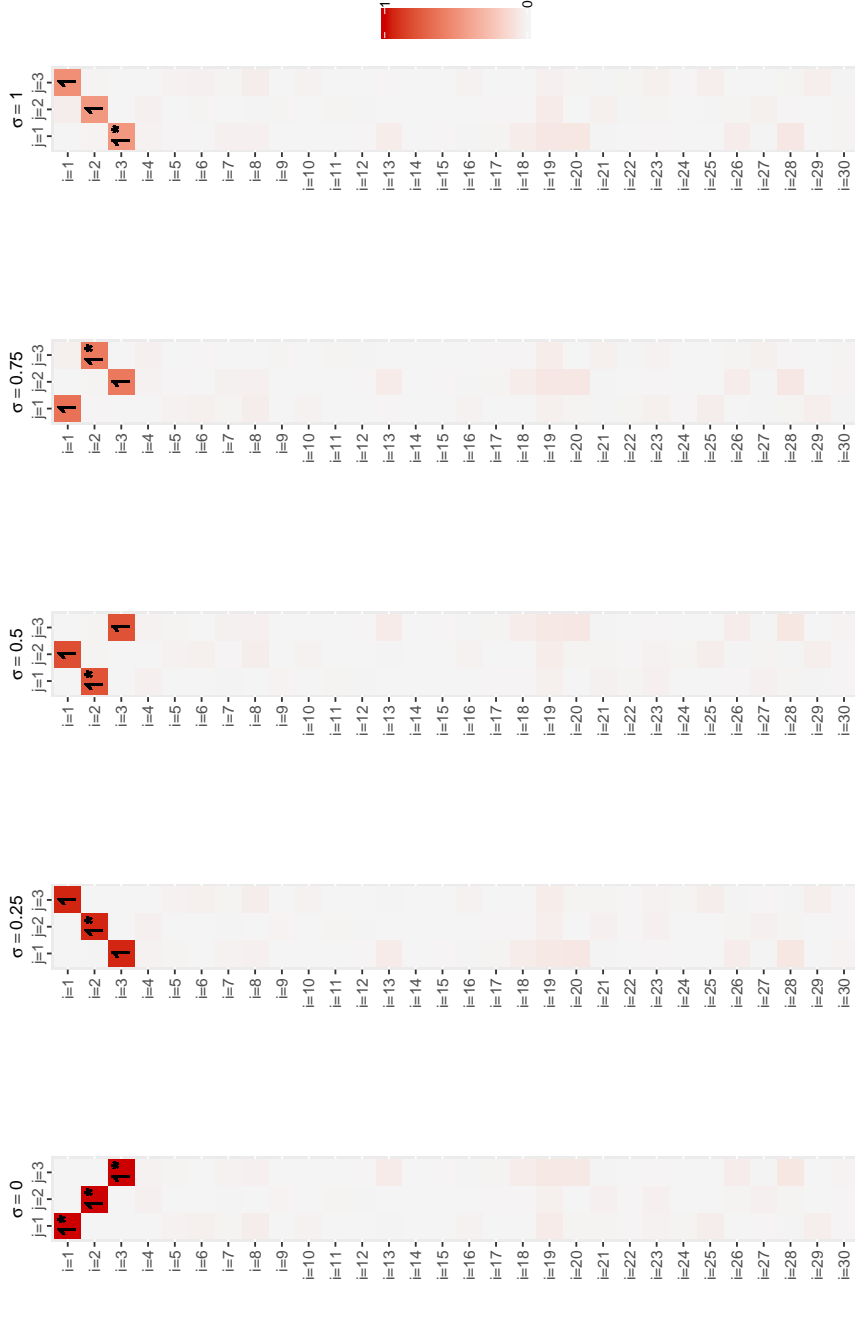
Figure 4: Heatmaps and $\mathcal{S}(\boldsymbol{H}, \boldsymbol{M})$ for synthetic dataset with different values of σ.

σ	Mean	Std
0	1.000	0.000
0.25	0.935	0.007
0.5	0.780	0.026
0.75	0.608	0.036
1	0.451	0.049
1.25	0.352	0.051
1.5	0.272	0.048
1.75	0.212	0.048
2	0.168	0.038
2.25	0.149	0.030
2.5	0.125	0.025
2.75	0.115	0.021
3	0.105	0.024
3.25	0.096	0.019
3.5	0.095	0.021
3.75	0.089	0.020
4	0.085	0.018
4.25	0.082	0.016
4.5	0.077	0.012
4.75	0.076	0.012
5	0.075	0.010

Table 1: Statistical summary of the performance of the model (optimal value of $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$) obtained in the 100 instances generated for each value of σ.
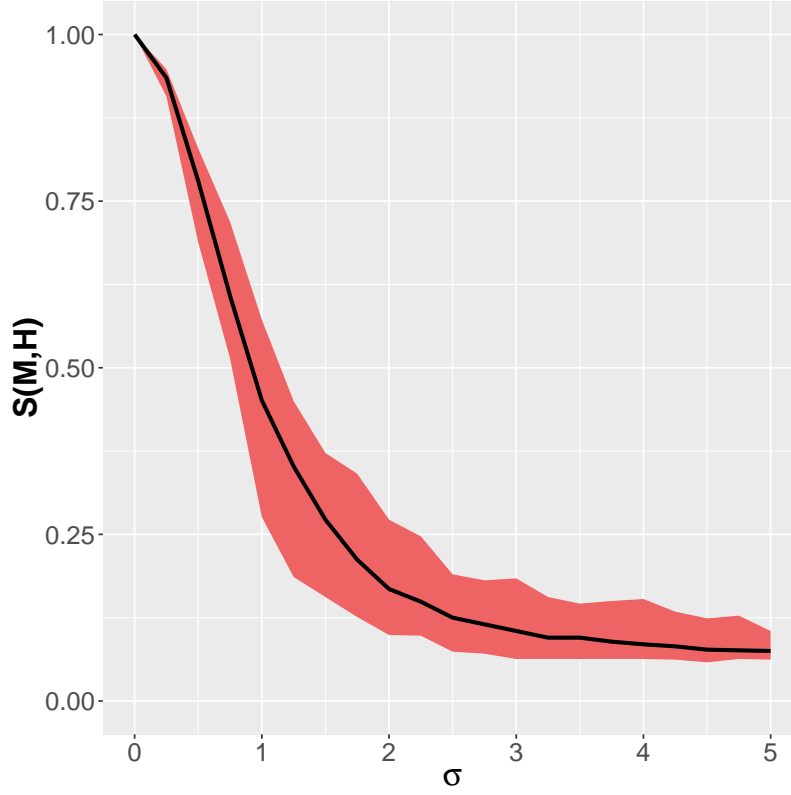
Figure 5: Minimum and maximum optimal values of $\mathcal{S}(\boldsymbol{M}, \boldsymbol{H})$ in the 100 instances (red) and the mean (black) for different values of σ.

5 Conclusions

In this paper we presented an optimization-based approach to interpret factors arising from Exploratory Factor Analysis (EFA). Besides the factors, we assume that a set of explanatory variables has been measured over the same cases. Contrary to classic methods, which interpret the factors in terms of the variables involved in EFA, our approach allows

the analyst to interpret the factors by means of the so-called set of explanatory variables, which might contain the observed ones as well. Our proposal is to match these variables to the factors in order to give meaning to them. A quality criterion which assesses the goodness of the matching according to an orthogonal rotation of the factors has been introduced. Our methodology has been shown to work successfully on both an empirical dataset and a synthetic one.

A potential virtue of our approach with respect to other alternatives is the possibility of modulating the criterion for rotation, i.e. the capacity to specify sets of variables associated (defining) the final factors. This may be of interest, for instance, in the case of large surveys, where there are many variables that can be used to interpret factors, but we do not want to involve all of them in the actual factor model analysis of equation (1). In addition, it is worth noting the flexibility of our approach to guide the interpretation of the factors with additional available information trough the structure of the matching matrix.

There are several extensions to the methodology proposed in this paper yielding interesting lines of future research. First, we can easily accommodate other goodness of fit criteria by changing the expression of $\mathcal{S}$ in (2), replacing either the min operator or the coefficient of determination, R^2, by different rules, Guadagnoli and Velicer (1991); Kaufman and Rousseeuw (1990). This is particularly necessary if categorical explanatory variables are to be used. Second, it is worth noting that, although we have used orthogonal rotations, our methodology extends to oblique rotations. However, uncorrelatedness of the factors would be lost and the optimization problems to solve would be more cumbersome. Third,

the extension to Dynamic Factor Models is a non-trivial challenge which deserves further analysis, Geweke (1977); Stock and Watson (2011). Fourth, developing a methodology which does not require the estimation of the factor scores would be interesting and desirable. It is well known that estimating factor scores is a controversial in factor analysis, known as factor score indeterminacy in the literature, Gorsuch (1983). However, this perspective requires new ideas which are out of the scope of this work. Finally, the idea of enhancing interpretability through auxiliary explanatory variables can be applied to other Data Science contexts, such as Principal Components, Correspondence or Canonical Correlation Analysis.

References

K. Adachi and N. T. Trendafilov. Sparsest factor analysis for clustering variables: a matrix decomposition approach. *Advances in Data Analysis and Classification*, 12(3):559–585, 2018. doi: 10.1007/s11634-017-0284-z.

T. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 3rd edition, 2003.

D. J. Bartholomew, M. Knott, and I. Moustaki. *Latent Variable Models and Factor Analysis: A Unified Approach*. Wiley, 2011.

D. Bertsimas, A. O'Hair, S. Relyea, and J. Silberholz. An analytics approach to designing

combination chemotherapy regimens for cancer. *Management Science*, 62(5):1511–1531, 2016. doi: 10.1287/mnsc.2015.2363.

D. M. Blei. Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application*, 1(1):203–232, 2014. doi: 10.1146/ annurev-statistics-022513-115657.

M. W. Browne. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36(1):111–150, 2001. doi: 10.1207/S15327906MBR3601_05.

E. Carrizosa and D. Romero Morales. Supervised classification and mathematical optimization. *Computers & Operations Research*, 40(1):150–165, 2013. doi: 10.1016/ j.cor.2012.05.015.

E. Carrizosa, V. Guerrero, and D. Romero Morales. Visualizing data as objects by DC (difference of convex) optimization. *Mathematical Programming*, 169(1):119–140, 2018a. doi: 10.1007/s10107-017-1156-1.

E. Carrizosa, V. Guerrero, and D. Romero Morales. On mathematical optimization for the visualization of frequencies and adjacencies as rectangular maps. *European Journal of Operational Research*, 265(1):290–302, 2018b. doi: 10.1016/j.ejor.2017.07.023.

C. M. Carvalho, J. Chang, J. E. Lucas, J. R. Nevins, Q. Wang, and M. West. High-dimensional sparse factor modeling: applications in gene expression genomics. *Journal*

of the American Statistical Association, 103(484):1438–1456, 2008. doi: 10.1198/016214508000000869.

J. Choi, H. Zou, and G. Oehlert. A penalized maximum likelihood approach to sparse factor analysis. *Statistics and its Interface*, 3(4):429–436, 2010. doi: 10.4310/SII.2010.v3.n4.a1.

C. DiStefano, M. Zhu, and D. Mîndrilă. Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14(20):1–11, 2009. URL http://pareonline.net/getvn.asp?v=14&n=20.

L.R. Fabrigar, D.T. Wegener, R.C. MacCallum, and E.J. Strahan. Evaluating the use of Exploratory Factor Analysis in psychological research. *Psychological Methods*, 4(3):272–299, 1999. doi: 10.1037/1082-989X.4.3.272.

J. Fan, Y. Liao, and W. Wang. Projected principal component analysis in factor models. *Annals of Statistics*, 44(1):219–254, 2016. doi: 10.1214/15-AOS1364.

J. Geweke. The Dynamic Factor Analysis of Economic Time Series. In D. J. Aigner and A. S. Goldberge, editors, *Latent Variables in Socio-Economic Models*. North-Holland Publishing Company, Amsterdam, North-Holland, 1977.

R.L. Gorsuch. *Factor Analysis*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 2 edition, 1983.

J. W. Grice. Computing and evaluating factor scores. *Psychological Methods*, 6(4):430–450, 2001. doi: 10.1037/1082-989X.6.4.430.

E. Guadagnoli and W. Velicer. A comparison of pattern matching indices. *Multivariate Behavioral Research*, 26(2):323–343, 1991. doi: 10.1207/s15327906mbr2602_7.

M. Hattori, G. Zhang, and K. J. Preacher. Multiple local solutions and Geomin rotation. *Multivariate Behavioral Research*, 52(6):720–731, 2017. doi: 10.1207/S15327906MBR3601_05.

S. G. Johnson. The NLopt nonlinear-optimization package. 2008. URL http://ab-initio.mit.edu/nlopt.

H. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, 1958. doi: 10.1007/BF02289233.

L. Kaufman and P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. John Wiley, New York, 1990.

S. Kaufmann and C. Schumacher. Identifying relevant and irrelevant variables in sparse factor models. *Journal of Applied Econometrics*, 32(6):1123–1144, 2017. doi: 10.1002/jae.2566.

G. Li and S. Jung. Incorporating covariates into integrated factor analysis of multi-view data. *Biometrics*, 73:1433–1442, 2017. doi: 10.1111/biom.12698.

M. E. Martínez, J. R. Marshall, and L. Sechrest. Invited commentary: Factor Analysis and the search of objectivity. *American Journal of Epidemiology*, 148:17–19, 1998.

S. A. Mulaik. *Foundations of Factor Analysis*. CRC press, 2nd edition, 2009.

S. Olafsson, X. Li, and S. Wu. Operations research and data mining. *European Journal of Operational Research*, 187(3):1429–1448, 2008. doi: 10.1016/j.ejor.2006.09.023.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL http://www.R-project.org. ISBN 3-900051-07-0.

V. Ročková and E. I. George. Fast bayesian factor analysis via automatic rotations to sparsity. *Journal of the American Statistical Association*, 111(516):1608–1622, 2016. doi: 10.1080/01621459.2015.1100620.

D. A. Sass and T. A. Schmitt. A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, 45(1):73–103, 2010. doi: 10.1080/00273170903504810.

A. Shapiro. Identifiability of factor analysis: Some results and open problems. *Linear Algebra and its Applications*, 70:1–7, 1985. doi: 10.1016/0024-3795(85)90038-2.

C. Spearman. "General Intelligence", objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292, 1904. doi: 10.2307/1412107.

J. H. Stock and M. Watson. Dynamic Factor Models. In M. P. Clement and D. F. Hendry, editors, *Oxford Handbook on Economic Forecasting*. Oxford University Press, Oxford, 2011. doi: 10.1093/oxfordhb/9780195398649.013.0003.

A. Taeb and V. Chandrasekaran. Interpreting latent variables in factor models via convex optimization. *Mathematical Programming*, 167(1):129–154, 2018. doi: 10.1007/s10107-017-1187-7.

A. Taeb, J.T. Reager, M. Turmon, and V. Chandrasekaran. A statistical graphical model of the California reservoir system. *Water Resources Research*, 53:9721–9739, 2017. doi: 10.1002/2017WR020412.

S. Theußl, F. Schwendinger, and K. Hornik. Roi: The R optimization infrastructure package. 2017. URL http://epub.wu.ac.at/id/eprint/5858.

L. L. Thurstone. Multiple factor analysis. *Psychological Review*, 38(5):406–427, 1931. doi: 10.1002/1097-4679(194804)4:2⟨224::AID-JCLP2270040225⟩3.0.CO;2-7.

N. T. Trendafilov, S. Fontanella, and K. Adachi. Sparse exploratory factor analysis. *Psychometrika*, 82(3):778–794, 2017. doi: 10.1007/s11336-017-9575-8.

K. Uno, H. Satomura, and K. Adachi. Fixed factor analysis with clustered factor score constraint. *Computational Statistics and Data Analysis*, 94:265–274, 2016. doi: 10.1016/j.csda.2015.08.010.

V. Van Vlasselaer, T. Eliassi-Rad, L. Akoglu, M. Snoeck, and B. Baesens. GOTCHA! Network-based fraud detection for social security fraud. *Management Science*, 63(9): 3090–3110, 2017. doi: 10.1287/mnsc.2016.2489.

T. von Landesberger, D. Fellner, and R. Ruddle. Visualization system requirements for data processing pipeline design and optimization. *IEEE Transactions on Visualization and Computer Graphics*, 23(8):2028–2041, 2017. doi: 10.1109/TVCG.2016.2603178.

Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(1-2):397–434, 2013. doi: 10.1007/s10107-012-0584-1.

D. M. Witten, R. Tibshirani, and T. Hastie. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10(3):515–534, 2016. doi: 10.1093/biostatistics/kxp008.

M. Yamamoto, K. Hirose, and H. Nagata. Graphical tool of sparse factor analysis. *Behaviormetrika*, 44(1):229–250, 2017. doi: 10.1007/s41237-016-0007-3.

S. Zhao, C. Gao, S. Mukherjee, and B. E. Engelhardt. Bayesian group factor analysis with structured sparsity. *Journal of Machine Learning Research*, 17(196):1–47, 2016.

Appendix

This Appendix contains the loading matrix $\mathbf{\Lambda}$ obtained from estimating model (1) for the empirical and synthetic datasets studied in Section 4, first column of Tables 2 and 3 respectively. Columns 2-7 show examples of rotated loading matrices for both datasets using different methods available in the `fa` function in R: `varimax` ($\mathbf{\Lambda}^*_{\mathtt{varimax}}$), `quartimax` ($\mathbf{\Lambda}^*_{\mathtt{quratimax}}$), `bentlerT` ($\mathbf{\Lambda}^*_{\mathtt{bentlerT}}$), `equamax` ($\mathbf{\Lambda}^*_{\mathtt{equamax}}$), `varimin` ($\mathbf{\Lambda}^*_{\mathtt{varimin}}$) and `geominT`

$(\mathbf{\Lambda}^*_{\texttt{geomin}})$. Besides the outputs of these rotation methods, we include the rotated loading matrix obtained using the rotation matrix obtained with our methodology ($InRotMat$) ($\mathbf{\Lambda}^*_{(InRotMat)}$) for the empirical dataset in last column of Table 2. Finally, the rotated loading matrices using our methodology for the synthetic dataset ($\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=0}}$, $\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=0.25}}$, $\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=0.5}}$, $\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=0.75}}$ and $\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=1}}$) are depicted in Table 4.

Λ		$\Lambda^*_{varimax}$		$\Lambda^*_{quartimax}$		$\Lambda^*_{bentlerT}$		$\Lambda^*_{equamax}$		$\Lambda^*_{varimin}$		$\Lambda^*_{geominT}$		$\Lambda^*_{(InRotMat)}$	
0.84	0.08	0.56	0.63	0.65	0.54	0.62	0.56	0.65	0.54	0.84	−0.00	0.62	0.57	0.42	0.73
0.71	0.11	0.45	0.57	0.52	0.50	0.50	0.52	0.52	0.50	0.72	−0.05	0.49	0.53	0.40	0.60
0.55	−0.62	0.82	−0.07	0.80	−0.19	0.81	−0.16	0.80	−0.19	0.49	0.66	0.81	−0.15	−0.33	0.76
0.76	−0.10	0.63	0.44	0.69	0.35	0.67	0.38	0.69	0.35	0.75	0.17	0.67	0.39	0.22	0.74
0.79	0.32	0.36	0.77	0.47	0.71	0.44	0.73	0.47	0.71	0.81	−0.24	0.43	0.73	0.62	0.59
0.32	0.08	0.18	0.28	0.22	0.25	0.21	0.26	0.22	0.25	0.33	−0.05	0.21	0.26	0.21	0.26
0.66	−0.49	0.81	0.09	0.82	−0.03	0.82	0.00	0.82	−0.03	0.61	0.55	0.82	0.01	−0.17	0.80
0.91	−0.17	0.78	0.50	0.84	0.38	0.83	0.41	0.84	0.38	0.89	0.25	0.82	0.42	0.22	0.90
0.59	−0.03	0.45	0.38	0.50	0.31	0.48	0.33	0.50	0.31	0.58	0.08	0.48	0.34	0.22	0.54
0.63	−0.47	0.78	0.08	0.79	−0.03	0.79	0.00	0.79	−0.03	0.59	0.53	0.79	0.01	−0.17	0.77
0.76	−0.47	0.87	0.17	0.89	0.04	0.89	0.08	0.89	0.04	0.71	0.54	0.89	0.09	−0.11	0.88
0.71	−0.37	0.77	0.21	0.79	0.10	0.79	0.13	0.79	0.10	0.67	0.43	0.78	0.14	−0.04	0.80
0.76	−0.31	0.77	0.29	0.80	0.17	0.79	0.21	0.80	0.17	0.73	0.38	0.79	0.22	0.03	0.82
0.57	0.37	0.16	0.66	0.26	0.63	0.23	0.64	0.26	0.63	0.60	−0.32	0.22	0.64	0.57	0.36
0.72	0.38	0.27	0.77	0.37	0.72	0.34	0.74	0.37	0.72	0.75	−0.31	0.33	0.74	0.64	0.49
0.68	−0.23	0.65	0.30	0.69	0.20	0.68	0.23	0.69	0.20	0.65	0.29	0.67	0.24	0.08	0.71
0.68	0.57	0.11	0.88	0.23	0.86	0.20	0.87	0.23	0.86	0.73	−0.51	0.18	0.87	0.80	0.38
0.81	−0.37	0.84	0.28	0.87	0.16	0.87	0.19	0.87	0.16	0.77	0.44	0.86	0.21	0.00	0.89
0.59	0.65	−0.01	0.87	0.12	0.87	0.08	0.87	0.12	0.87	0.64	−0.59	0.07	0.87	0.83	0.27
0.20	−0.17	0.26	0.01	0.26	−0.03	0.26	−0.02	0.26	−0.03	0.18	0.18	0.26	−0.01	−0.07	0.25
0.85	−0.29	0.81	0.37	0.86	0.25	0.85	0.28	0.86	0.25	0.82	0.36	0.84	0.29	0.09	0.89
0.91	0.20	0.53	0.77	0.63	0.68	0.61	0.71	0.63	0.68	0.92	−0.12	0.59	0.72	0.56	0.74
0.89	−0.10	0.73	0.53	0.80	0.42	0.78	0.46	0.80	0.42	0.88	0.18	0.77	0.47	0.28	0.86
0.59	−0.42	0.72	0.09	0.73	−0.01	0.73	0.02	0.73	−0.01	0.55	0.48	0.73	0.03	−0.14	0.72
0.10	−0.18	0.20	−0.06	0.19	−0.09	0.19	−0.08	0.19	−0.09	0.09	0.19	0.19	−0.08	−0.12	0.17
0.89	0.05	0.62	0.64	0.70	0.55	0.68	0.57	0.70	0.55	0.89	0.03	0.67	0.58	0.41	0.79
0.74	−0.44	0.84	0.18	0.86	0.06	0.86	0.09	0.86	0.06	0.70	0.51	0.86	0.10	−0.10	0.86
0.56	0.63	−0.02	0.84	0.11	0.84	0.07	0.84	0.11	0.84	0.62	−0.58	0.06	0.84	0.81	0.25
0.58	0.68	−0.04	0.89	0.09	0.88	0.05	0.89	0.09	0.88	0.63	−0.62	0.04	0.89	0.85	0.24
0.38	0.46	−0.04	0.60	0.05	0.60	0.02	0.60	0.05	0.60	0.42	−0.43	0.01	0.60	0.58	0.15
0.60	−0.57	0.83	−0.01	0.82	−0.13	0.82	−0.10	0.82	−0.13	0.54	0.62	0.82	−0.08	−0.27	0.78
0.81	−0.27	0.78	0.35	0.82	0.24	0.81	0.27	0.82	0.24	0.78	0.34	0.81	0.28	0.09	0.85
0.27	0.22	0.04	0.35	0.09	0.34	0.08	0.34	0.09	0.34	0.29	−0.20	0.07	0.34	0.32	0.15
−0.02	0.10	−0.08	0.06	−0.07	0.07	−0.07	0.07	−0.07	0.07	−0.01	−0.10	−0.07	0.07	0.08	−0.05
0.84	0.27	0.43	0.77	0.54	0.70	0.51	0.72	0.54	0.70	0.86	−0.20	0.50	0.73	0.60	0.65
0.69	0.44	0.21	0.80	0.32	0.76	0.29	0.77	0.32	0.76	0.73	−0.38	0.28	0.78	0.69	0.45
0.71	0.44	0.22	0.81	0.33	0.77	0.30	0.78	0.33	0.77	0.74	−0.38	0.29	0.78	0.70	0.46
0.59	0.15	0.33	0.51	0.40	0.46	0.38	0.48	0.40	0.46	0.61	−0.10	0.38	0.48	0.38	0.48
0.76	−0.15	0.65	0.41	0.71	0.31	0.69	0.34	0.71	0.31	0.74	0.21	0.69	0.35	0.18	0.75
0.83	0.11	0.53	0.64	0.62	0.56	0.59	0.58	0.62	0.56	0.83	−0.03	0.59	0.59	0.44	0.71
0.18	0.27	−0.05	0.32	−0.01	0.32	−0.02	0.32	−0.01	0.32	0.20	−0.25	−0.02	0.32	0.32	0.05
0.72	−0.58	0.92	0.07	0.93	−0.06	0.93	−0.03	0.93	−0.06	0.67	0.64	0.93	−0.01	−0.23	0.90
0.60	−0.58	0.83	−0.02	0.82	−0.14	0.83	−0.11	0.82	−0.14	0.54	0.63	0.83	−0.09	−0.28	0.79
0.86	−0.34	0.86	0.34	0.90	0.21	0.89	0.25	0.90	0.21	0.82	0.41	0.88	0.26	0.05	0.92
0.73	0.04	0.51	0.53	0.58	0.45	0.56	0.47	0.58	0.45	0.73	0.02	0.55	0.48	0.34	0.65
0.85	−0.30	0.83	0.36	0.87	0.24	0.86	0.28	0.87	0.24	0.82	0.37	0.85	0.29	0.08	0.90
0.52	0.67	−0.07	0.84	0.05	0.85	0.02	0.85	0.05	0.85	0.58	−0.62	0.00	0.85	0.82	0.20
0.75	−0.10	0.62	0.44	0.68	0.35	0.66	0.37	0.68	0.35	0.74	0.17	0.65	0.38	0.22	0.73
0.82	−0.41	0.88	0.26	0.91	0.13	0.90	0.17	0.91	0.13	0.78	0.48	0.90	0.18	−0.03	0.92
0.63	0.64	0.03	0.89	0.15	0.88	0.12	0.89	0.15	0.88	0.68	−0.58	0.10	0.89	0.84	0.31
0.40	−0.29	0.49	0.06	0.50	−0.01	0.50	0.01	0.50	−0.01	0.37	0.33	0.50	0.02	−0.10	0.49
0.67	0.11	0.42	0.53	0.49	0.46	0.47	0.48	0.49	0.46	0.67	−0.05	0.46	0.49	0.37	0.56
0.74	0.52	0.19	0.89	0.31	0.85	0.28	0.86	0.31	0.85	0.79	−0.45	0.27	0.87	0.78	0.46
0.84	0.19	0.49	0.71	0.59	0.63	0.56	0.66	0.59	0.63	0.86	−0.11	0.55	0.67	0.52	0.69
0.64	0.63	0.04	0.90	0.17	0.88	0.13	0.89	0.17	0.88	0.69	−0.57	0.12	0.89	0.84	0.32
−0.07	0.24	−0.22	0.13	−0.20	0.16	−0.20	0.15	−0.20	0.16	−0.05	−0.25	−0.20	0.15	0.19	−0.17
0.76	0.10	0.49	0.59	0.57	0.51	0.55	0.54	0.57	0.51	0.77	−0.03	0.54	0.54	0.41	0.65
0.54	0.52	0.04	0.75	0.15	0.73	0.12	0.74	0.15	0.73	0.58	−0.47	0.10	0.74	0.70	0.27
0.63	−0.46	0.78	0.09	0.78	−0.02	0.78	0.01	0.78	−0.02	0.58	0.52	0.78	0.02	−0.16	0.76
0.55	0.33	0.18	0.62	0.27	0.59	0.25	0.60	0.27	0.59	0.58	−0.28	0.24	0.60	0.53	0.37

Table 2: Loading matrix of `reservoir` dataset and its rotated counterparts using different rotation methods.

Λ			$\Lambda^*_{\text{varimax}}$			$\Lambda^*_{\text{quartimax}}$			$\Lambda^*_{\text{bentlerT}}$			$\Lambda^*_{\text{equamax}}$			$\Lambda^*_{\text{varimin}}$			$\Lambda^*_{\text{geominT}}$		
−0.09	0.03	0.45	0.05	−0.06	0.45	0.05	−0.06	0.45	0.05	−0.06	0.45	0.05	−0.06	0.45	0.05	−0.06	0.45	0.05	−0.06	0.45
−0.24	0.23	0.82	−0.05	0.02	0.88	−0.05	0.02	0.88	−0.05	0.02	0.88	−0.05	0.02	0.88	−0.05	0.02	0.88	−0.05	0.02	0.88
−0.36	−0.12	0.22	−0.18	−0.30	0.27	−0.18	−0.30	0.27	−0.18	−0.30	0.27	−0.18	−0.30	0.27	−0.18	−0.30	0.27	−0.18	−0.30	0.27
0.08	−0.04	−0.17	0.03	0.02	−0.19	0.03	0.02	−0.19	0.03	0.02	−0.19	0.03	0.02	−0.19	0.03	0.02	−0.19	0.03	0.02	−0.19
0.10	−0.12	−0.32	0.04	−0.03	−0.35	0.04	−0.03	−0.35	0.04	−0.03	−0.35	0.04	−0.03	−0.35	0.04	−0.03	−0.35	0.04	−0.03	−0.35
0.34	0.54	−0.11	0.01	0.64	−0.07	0.01	0.64	−0.07	0.01	0.64	−0.07	0.01	0.64	−0.07	0.01	0.64	−0.07	0.01	0.64	−0.07
0.16	0.63	−0.04	−0.15	0.63	0.06	−0.15	0.63	0.06	−0.15	0.63	0.06	−0.15	0.63	0.06	−0.15	0.63	0.06	−0.15	0.63	0.06
0.36	0.56	−0.00	0.05	0.66	0.03	0.05	0.66	0.03	0.05	0.66	0.03	0.05	0.66	0.03	0.05	0.66	0.03	0.05	0.66	0.03
0.28	0.28	0.06	0.12	0.37	0.05	0.12	0.37	0.05	0.12	0.37	0.05	0.12	0.37	0.05	0.12	0.37	0.05	0.12	0.37	0.05
0.39	0.31	−0.06	0.17	0.46	−0.08	0.17	0.46	−0.08	0.17	0.46	−0.08	0.17	0.46	−0.08	0.17	0.46	−0.08	0.17	0.46	−0.08
0.84	−0.35	0.30	0.95	0.08	0.01	0.95	0.08	0.01	0.95	0.08	0.01	0.95	0.08	0.01	0.95	0.08	0.01	0.95	0.08	0.01
0.72	−0.29	0.21	0.80	0.09	−0.03	0.80	0.09	−0.03	0.80	0.09	−0.03	0.80	0.09	−0.03	0.80	0.09	−0.03	0.80	0.09	−0.03
−0.40	−0.05	−0.03	−0.33	−0.24	0.05	−0.33	−0.24	0.05	−0.33	−0.24	0.05	−0.33	−0.24	0.05	−0.33	−0.24	0.05	−0.33	−0.24	0.05
−0.30	0.34	0.16	−0.36	0.14	0.29	−0.36	0.14	0.29	−0.36	0.14	0.29	−0.36	0.14	0.29	−0.36	0.14	0.29	−0.36	0.14	0.29
−0.43	0.17	−0.09	−0.46	−0.06	0.05	−0.46	−0.06	0.05	−0.46	−0.06	0.05	−0.46	−0.06	0.05	−0.46	−0.06	0.05	−0.46	−0.06	0.05

Table 3: Loading matrix of the synthetic dataset and its rotated counterparts using different rotation methods.

$\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=0}}$			$\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=0.25}}$			$\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=0.5}}$			$\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=0.75}}$			$\mathbf{\Lambda}^*_{(InRotMat)_{\sigma=1}}$		
−0.20	0.41	−0.01	−0.01	−0.41	−0.20	0.42	0.18	0.02	0.27	0.03	0.36	−0.04	−0.34	−0.30
−0.24	0.84	−0.13	−0.14	−0.84	−0.24	0.84	0.20	0.15	0.39	0.16	0.78	−0.17	−0.74	−0.44
0.03	0.34	0.28	0.28	−0.34	0.03	0.34	−0.04	−0.27	0.04	−0.27	0.34	0.27	−0.34	−0.07
0.03	−0.19	0.00	0.00	0.19	0.03	−0.19	−0.02	−0.01	−0.07	−0.01	−0.18	0.01	0.17	0.08
0.07	−0.34	0.07	0.08	0.34	0.07	−0.34	−0.06	−0.08	−0.13	−0.09	−0.32	0.09	0.31	0.15
0.11	−0.12	−0.62	−0.62	0.13	0.12	−0.14	−0.12	0.62	−0.16	0.61	−0.11	−0.61	0.10	0.18
0.22	0.06	−0.61	−0.61	−0.05	0.23	0.03	−0.24	0.60	−0.24	0.60	0.08	−0.60	−0.10	0.24
0.04	−0.04	−0.66	−0.66	0.05	0.05	−0.06	−0.06	0.66	−0.08	0.66	−0.05	−0.66	0.05	0.10
−0.08	−0.03	−0.39	−0.39	0.03	−0.07	−0.03	0.07	0.39	0.05	0.39	−0.05	−0.39	0.06	−0.04
−0.06	−0.17	−0.46	−0.46	0.18	−0.06	−0.18	0.06	0.46	0.01	0.46	−0.19	−0.46	0.19	0.02
−0.88	−0.32	−0.19	−0.20	0.32	−0.88	−0.29	0.89	0.21	0.80	0.22	−0.48	−0.23	0.54	−0.75
−0.72	−0.31	−0.17	−0.18	0.31	−0.72	−0.28	0.73	0.19	0.65	0.20	−0.44	−0.21	0.49	−0.61
0.25	0.18	0.26	0.27	−0.18	0.25	0.17	−0.25	−0.27	−0.20	−0.27	0.23	0.27	−0.24	0.18
0.25	0.39	−0.13	−0.13	−0.38	0.26	0.37	−0.27	0.13	−0.19	0.13	0.42	−0.13	−0.43	0.16
0.41	0.22	0.10	0.11	−0.21	0.41	0.20	−0.41	−0.11	−0.35	−0.12	0.29	0.12	−0.31	0.33

Table 4: Rotated loading matrices of the synthetic dataset using $(InRotMat)$ for different values of σ.