Sustainable progress indicators based on confirmatory factorial analysis

Adrià Prior
Tutoress: Mercè Farré
Bellaterra

July 2020

Abstract

This document reviews the theory of factorial analysis, giving importance to the fundamental results that validate the use of the technique. Some of the results shown in this work, although understood as necessary in the construction of the statistical method, couldn't be found proved in other references, and their demonstrations are included. Once the theory has been explained, it is shown trough an example, a procedure to obtain probability based indices built over latent factors fitting a confirmatory factorial model. Indices are intended to evaluate the evolution of economic, as well as social, ecological and urban aspects in metropolitan regions, and so they can be considered sustainable progress indicators. The procedure was proposed by a research team of the IERMB[1] and the MCS[2], in a study published in 2019. The example in this report is still in the discussion phase. The aim is to initiate the updating of that methodology, to apply it in a future study with new and extended data.

1 Introduction

Factorial analysis is a statistical theory that allows, under certain conditions, expressing approximately the variables of a random vector of which we have observations, as a linear combination of a few new variables called factors, through a stochastic model. The so called factor model is built in such a way that allows to search for interpretable factors in the context of the investigation. The objective is explaining the individuals or observations of the initial random vector in terms of this new factors. One use of factorial analysis is to try to quantitatively model as factors qualitative aspects of interest (such as intelligence or social class, for example), considering as initial observable variables ones that is thought could describe the aspect that is studied. This document explains the fundamental results of factorial analysis as well as the necessary ones to apply the theory, giving importance to the elemental theorems that validate the use of the technique. Once the theory is known, it is shown a procedure to evaluate sustainable progress in metropolitan regions, using sustainable progress indicators based on confirmatory factorial analysis, trough a concrete analysis, relying on socioeconomic data of Europe metropolitan regions provided by IERMB (*Barcelona Institute of Regional and Metropolitan Studies*). The procedure was proposed and used in Marull et al. (*11*) to evaluate sustainable progress in Europe metropolitan regions and megaregions between 1995 and 2010.

The first part of this document, consisting on sections 2 and 3, explains the theoretical foundations

[1] *Barcelona Institute of Regional and Metropolitan Studies* in UAB campus.
[2] *Mathematical Consulting Service*, Math Department, UAB.

of factorial analysis. The objective is adjusting adequately a factor model to a given set of variables, which we shall call initial variables. There are two principal variants of factorial analysis: confirmatory and exploratory. Exploratory factorial analysis is commonly used to find a factor model that fits the initial variables, whereas confirmatory factorial analysis is often used after an exploratory analysis, with the aim of fitting a specific factor model with the values of some of its parameters fixed in advance by the researcher, much of them usually forced to be zero. Section 2 is devoted to exploratory factorial analysis and section 3 is dedicated to confirmatory factorial analysis. In the second section we define the orthogonal factor model and we present its basic properties. Sections 2.2 2.3 and 2.4 discuss the way to find, in practice, an estimated solution to the orthogonal factor model for an initial random vector, using a set of observations of it summarized in a data matrix, we finally present the principal factors method of estimation. Section 2.5 shows how to obtain the values that may take the factors for a given observation of the initial random vector, once the model has been fitted, this values are called factor scores. Section 2.6 discuss the interpretation of the factors, and overviews the varimax rotation method, that aims to provide an interpretable factors, taking advantage of the non uniqueness of solution to the model. Section 3 introduces confirmatory factorial analysis highlighting the theoretical and practical differences with the exploratory version, the factor model is defined, the existence and uniqueness of solution to it is discussed, and its basic properties are shown, the ways to find its parameters estimates are overviewed and the factors interpretation is addressed.

In the second part of this document, consisting on section 4, we perform a confirmatory factorial analysis to the data set provided by IERMB to show the procedure to obtain sustainable progress indicators. Sustainable progress in city networks accounts an increasing level of economic competitiveness, urban complexity and social and environmental well-being (Marull et al. (*11*)), this concept arises to fulfil the limitations of GDP and per capita income as measures of overall human well-being, among other objectives. The object of study are the metropolitan regions (cities and their respective metropolitan areas). To evaluate sustainable progress in metropolitan regions it is necessary to measure economic, ecological, social and urban aspects, as well as the urban complexity, and confirmatory factorial analysis can help in this task. The procedure proposed in Marull et al.(*11*) is based on fitting a factor model to a vector of socioeconomic, ecological and urban variables, using observations of this vector for different metropolitan regions, and having observations of different years for each region. The factor model is fitted using confirmatory factorial analysis, in such a way that the initial variables are explained with a few new factors and this are interpreted, if possible, as economic, social, ecological and urban aspects, from this factors are derived simple indices measuring such aspects, and this simple indices are finally integrated to a compound indicator to evaluate sustainable progress. The indicator is evaluated in the initial region-year observations, and by observing the evolution on the values of it for a region, it is possible to tell if the level of the aspects measured has seen or not an increase over the years, and this way evaluate if the progress of a metropolitan region has been sustainable.

Early this year, IERMB started working in a project to update Marull et al.'s(*11*) study, with new data from 2012 to 2019, and this work was proposed to initialize the new analysis. Unfortunately, due the Covid-19 crisis, the necessary data was not prepared until later in June, and we could only dispose from data of a single year; 2016, in consequence, the applied part in this document was reduced to an analysis of the 2016 data set, and it is intended only to show the procedure to obtain sustainable progress indicators trough confirmatory factorial analysis, but it can't be taken as a meaningful analysis, in one hand, because of the lack of data, on the other hand, because we hadn't discussed the interpretation of the factors with the experts on the matter of sustainable progress and it's derived social, economic, urban and ecological dimensions in IERMB; the interpretation was based on our intuition about this matters and in the considerations in Marull et al.'s(*11*) study, hence, it is important to remark that the analysis has to be seen only as an explanation of the statistical procedure to obtain the indicators, but never as valid to draw conclusions about the (miss) evaluated aspects in the analyzed metropolitan regions using its results. This will be clear during the explanation of the performed analysis. We would have liked to discuss the interpretation of the factors with the people of IERMB, but we haven't had the time to do it properly. The reader may realize that with data of a single year it is not possible to evaluate progress, since the progress is seen comparing an indicator values trough different years, but the procedure to obtain the indicator is the same given a data set of a single year or a data set of various years, thus, we will see the procedure to obtain compound indicators to evaluate economic, social, ecological, and urban aspects, but we will only be able to give the value of this indicators for the year 2016 of each metropolitan region. Disposing of data of more years, it is possible to see the evolution in the values of a compound indicator and evaluate the progress of metropolitan regions in different aspects. Whereas the applied part was reduced, we went deeper in the theory of factorial analysis. We have not seen this as a problem, on the contrary, it is though important to have a certain knowledge of the theory before applying it, in fact, some of the results shown in the first part of this document couldn't be found proved by the author, although they were seen as fundamental and necessary, concretely, this results are the theorems 2.1.5 and 3.2.4.

Thus, the applied part consist on the analysis of a 2016 data set, section 4.1 explains the variables used in the analysis, consisting on socioeconomic, environmental and urban ones, section 4.2 shows the 4 factor model adjusted to the data. We interpret the factors as socioeconomic, environmental and urban aspects related with the initial variables, this labelling is done taking into account their mathematical relationships with the initial variables, it can not be taken as valid and it must be taken as an example, since it was not discussed with the experts in IERMB. In the section 4.3 various indicators are derived from the factors, including simple indicators, measuring the aspect corresponding to each factor, and compound indexes, which take into account all ecological, socioeconomic and urban factors, one of this compound indexes is taken as a sustainable progress indicator, and all the regions observed are evaluated.

For the interest of the reader, Marull et al.(*11*) study is of public access and it can be found in the IERMB's website.

2 Exploratory factorial analysis

Exploratory factorial analysis is the main version of factorial analysis, it will allow us to understand the procedure, and it will serve as a basis for confirmatory factorial analysis. The objective will be to find an adequate solution (at least approximately) to the orthogonal factor model for a given set of observed variables. Let's define what does it mean.

2.1 The model and the fundamental results

Definition 2.1.1. Let $X^t = (X_1, \dots, X_p)$ be a $p \times 1$ random vector with $E[X] = 0_{p\times 1}$. We say that the orthogonal factor model holds for X if there exist two random vectors $f^t = (f_1, \dots, f_m)$ with $m < p$ and $u^t = (u_1, \dots, u_p)$ and a matrix $Q = (q_{ij}) \in M_{p\times m}(\mathbb{R})$ such that

$$\begin{aligned}
X_1 &= q_{11}f_1 + q_{12}f_2 + \cdots + q_{1m}f_m + u_1 \\
X_2 &= q_{21}f_1 + q_{22}f_2 + \cdots + q_{2m}f_m + u_2 \\
&\vdots \\
X_p &= q_{p1}f_1 + q_{p2}f_2 + \cdots + q_{pm}f_m + u_p
\end{aligned}$$

In short: $X = Qf + u$, and satisfying:

i) $E[f] = 0_{m\times 1}$, $Cov(f) = I_m$, with I_m the identity matrix on $\mathbb{R}^m$.
ii) $E[u] = 0_{p\times 1}$, $Cov(u) = \Psi$, with $\Psi \in M_{p\times p}(\mathbb{R})$ and diagonal.
iii) $Cov(f,u) = 0_{m\times p}$, where $Cov(f,u)$ denotes the cross-covariance matrix between f and u.

In this case we say that the triplet (Q, f, u) is a solution to the orthogonal factor model for X, $(f_1, \dots, f_m)$ are called common factors of the model, $(u_1, \dots, u_p)$ are called specific factors and the matrix Q is called the loadings matrix. We consider the model with $m < p$ because one of the objectives is explaining the initial variables in a simplified way with a few factors.

In practice X will be the initial random vector of which we will have observations, the assumption $E[X] = 0_{p\times 1}$ is not restrictive since data can be centered to get the model and translated to the original center at the end, if necessary. We will be interested in the common factors while the specific ones could be understood as stochastic errors to hold the model. With respect to the conditions $i), ii)$ and $iii)$, the condition $i)$ asks the common factors to be uncorrelated and have unit variance, this condition is why we call the model orthogonal, considering the covariance as a scalar product. The condition $ii)$ ask the specific factors to be uncorrelated and $iii)$ ask the common factors to be uncorrelated with the specific factors. We could consider more general assumptions as allowing the common factors to be correlated, but is convenient for our current purpose of introducing factorial analysis to leave them for the confirmatory version.

To clarify notation, in this document we will use Σ_X to denote the covariance matrix of a random vector X, as well as $Cov(X)$ or $Cov(X, X)$ using the cross-covariance matrix notation, depending on the situation, that is $\Sigma_X = Cov(X) = Cov(X, X)$. Let's see the basic properties of the model:

Proposition 2.1.2. Let X be a random vector with $E[X] = 0_{p\times 1}$. If the orthogonal factor model holds for X and (Q, f, u) is a solution, then $Cov(X, f) = Q$.

Proof. Using that (Q, f, u) is a solution to the model, that is: $X = Qf + u$ satisfying $i), ii)$ and $iii)$, the properties $i)(Cov(f, f) = I_m)$ and $iii)(Cov(f, u) = 0_{m\times p})$ and basic properties of the cross-covariance matrix we have:

$$Cov(X, f) = Cov(Qf + u, f) = Cov(Qf, f) + Cov(u, f) = QCov(f, f) = QI_m = Q. \blacksquare$$

Thus, the variances between the initial variables and the common factors are given by the loadings; $Cov(X_i, f_j) = q_{ij}$. This result will help us to interpret the factors, that is, understand what the factors represent in the context of the investigation, concretely if the initial data is standardized, is valid to interpret a factor in terms of the variables more correlated with it, and so those that more contribute to it, although the interpretation will not always be possible or clear. We will discuss this point further.

Proposition 2.1.3. Let X be a random vector with $E[X] = 0_{p\times 1}$. If the orthogonal factor model holds for X and (Q, f, u) is a solution, then $\Sigma_X = QQ^t + \Psi$

Proof.

$$\begin{aligned}\Sigma_X = Cov(X) = Cov(Qf + u) &= Cov(Qf) + Cov(Qf, u) + Cov(u, Qf) + Cov(u)\\ &= QCov(f)Q^t + QCov(f, u) + Cov(u, f)Q^t + Cov(u)\\ &= QI_mQ^t + \Psi\\ &= QQ^t + \Psi\end{aligned}$$

Where we have used the properties $i), ii)$ and $iii)$ of the solution. ■

Observation 2.1.4. In particular, if the orthogonal factor model with m factors holds for $X = (X_1, \ldots, X_p)^t$, denoting as ψ_i the ith element on the diagonal of Ψ, such that we can write Ψ as $\Psi = diag(\psi_1, \ldots, \psi_p)$, and defining $h_i^2 := \sum_{j=1}^m q_{ij}^2$, for $i \in \{1, \ldots, p\}$, we have

$$Var(X_i) = \sum_{j=1}^{m} q_{ij}^2 + \psi_i = h_i^2 + \psi_i, \text{ for } i \in \{1, \ldots, p\} \tag{1}$$

The value h_i^2 is called the ith communality, while ψ_i is called the ith specific variance.

Once we have seen the first two basic properties of the orthogonal factor model, assuming it holds, it's time to ask if it's possible to find conditions for our initial variables, that ensure existence of solution to the model, since we want to obtain the factors from a data set performed by observations of the initial variables. We will state and prove that the necessary condition given in the last proposition is sufficient, if we suppose Ψ to be positive definite, following the hint given in Mardia et al. (1979, p. 276) (*10*). We observe that if the model holds, Ψ is necessarily positive semi definite, since it's a covariance matrix, and therefore the condition of positive definiteness does not seem very restrictive, we will discuss this point further. Concretely the statement of the result is given in the next theorem:

Theorem 2.1.5. (Existence of solution to the orthogonal factor model)
Let $X^t = (X_1, \ldots, X_p)$ be a $p \times 1$ random vector with $E[X] = 0_{p \times 1}$. If there exist two matrices $Q \in M_{p \times m}(\mathbb{R})$, with $m < p$, and $\Psi \in M_{p \times p}(\mathbb{R})$, with Ψ diagonal and positive definite, such that $\Sigma_X = QQ^t + \Psi$, then there exist two random vectors $f^t = (f_1, \ldots, f_m)$ and $u^t = (u_1, \ldots, u_p)$ that satisfy the orthogonal factor model with loadings matrix Q and $Cov(u) = \Psi$, that is; satisfying $X = Qf + u, Cov(u) = \Psi$ and $i), ii), iii)$.

Proof. Following the hint given in Mardia et al. (1979, p. 276) (*10*), we will show first that there exist a multivariate normal random vector $Y^t = (Y_1, \ldots, Y_m)$ with $Y \sim N_m(0_{m \times 1}, I_m + Q^t \Psi^{-1} Q)$, and secondly we will show that the pair of random vectors defined by:

$$\begin{pmatrix} u \\ f \end{pmatrix} := \underbrace{\begin{pmatrix} I_p & Q \\ -Q^t \Psi^{-1} & I_m \end{pmatrix}^{-1}}_{A^{-1}} \begin{pmatrix} X \\ Y \end{pmatrix} \tag{2}$$

give a solution to the orthogonal factor model.
Denote $W = I_m + Q^t \Psi^{-1} Q$, Ψ is invertible since it's symmetric and positive definite by hypothesis, thus we can take W. First of all, let's see that W is symmetric and positive semi definite, and therefore we can consider a multivariate normal vector Y with covariance matrix W:

$$(Q^t \Psi^{-1} Q)^t = Q^t (\Psi^1)^t (Q^t)^t = Q^t \Psi^{-1} Q$$

since Ψ is diagonal, and therefore $Q^t \Psi^{-1} Q$ is symmetric, hence $W = I_m + Q^t \Psi^{-1} Q$ is also symmetric since the sum only affects the diagonal of $Q^t \Psi^{-1} Q$. Let $v \in M_{m \times 1} M(\mathbb{R})$ be any vector, let $y := Qv$, we have $v^t\, Q^t \Psi^{-1} Q v = y^t \Psi^{-1} y \geq 0$ since Ψ^{-1} is positive definite, hence, by definition, $Q^t \Psi^{-1} Q$ is positive semi definite, and since it is also symmetric we have $det(Q^t \Psi^{-1} Q) \geq 0$. Now using that if B and C are positive semidefinite matrices then $det(B + C) \geq det(B) + det(C)$ (Lin and Sra)(*9*), we have:

$$det(I_m + Q^t \Psi^1 Q) \geq det(I_m) + det(Q^t \Psi^{-1} Q) \geq 1$$

Thus, $det(W) > 0$, and since W is also symmetric, is positive definite (Cedó and Reventós)(*3*), as we wanted to see. Since $det(W) > 0$, W is invertible, which we will use later.
Now let's see that the matrix A defining the vector of factors in (2), is invertible. The identity:

$$det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = det(D) det(A - BD^{-1}C)$$

holds when D is square and invertible (Schur, 1917)(*15*), using it for A we have

$$det(A) = \begin{pmatrix} I_p & Q \\ -Q^t\Psi^{-1} & I_m \end{pmatrix} = det(I_m)det(I_p + QI_mQ^t\Psi^{-1}) = det(I_p + QQ^t\Psi^{-1})$$

Now,

$$det(I_p + QQ^t\Psi^{-1}) = det(\Psi\Psi^{-1} + QQ^t\Psi^{-1}) = det((\Psi + QQ^t)\Psi^{-1}) = det(\Psi + QQ^t)det(\Psi^{-1})$$

QQ^t is symmetric and positive semidefinite, hence $det(QQ^t) \geq 0$, Ψ^{-1} is positive definite thus $det(\Psi^{-1}) > 0$, hence, $det(\Psi + QQ^t) \geq det(\Psi) + det(QQ^t) > 0$ and therefore $det(A) = det(\Psi + QQ^t)det(\Psi^{-1}) > 0$, thus A is invertible, and $\Sigma_X = \Psi + QQ^t$ is also invertible.
After this technical details, we are ready to see that the factors in (2) give a solution to the model. First let's see that $X = Qf + u$ holds. Since A is invertible, we have:

$$\begin{pmatrix} u \\ f \end{pmatrix} = \begin{pmatrix} I_p & Q \\ -Q^t\Psi^{-1} & I_m \end{pmatrix}^{-1} \begin{pmatrix} X \\ Y \end{pmatrix} \iff \begin{pmatrix} X \\ Y \end{pmatrix} = \underbrace{\begin{pmatrix} I_p & Q \\ -Q^t\Psi^{-1} & I_m \end{pmatrix}}_{A} \begin{pmatrix} u \\ f \end{pmatrix}$$

Hence, $X = I_pu + Qf = Qf + u$. We need to see that f and u satisfy $i)$, $ii)$ and $iii)$. First, we will see that $ii)$ holds showing that $Cov(u) = \Psi$ and $E(u) = 0_{p\times 1}$. As

$$\begin{pmatrix} u \\ f \end{pmatrix} = A^{-1} \begin{pmatrix} X \\ Y \end{pmatrix}$$

then

$$E\begin{pmatrix} u \\ f \end{pmatrix} = EA^{-1}\begin{pmatrix} X \\ Y \end{pmatrix} = A^{-1}E\begin{pmatrix} X \\ Y \end{pmatrix} = 0_{(p+m)\times 1}$$

And therefore $E[u] = 0_{p\times 1}$ and $E[f] = 0_{m\times 1}$. We will use the identity:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{pmatrix}$$

that holds for any block matrix with D and $(A - BD^{-1}C)$ invertible (Banachiewicz, 1937)(*1*), to obtain the inverse of A. In our case; $A - BD^{-1}C = I_p + QQ^t\Psi^{-1}$, we will denote this matrix M, we have already seen that $det(M) = det(A) > 0$, and thus M is invertible, also in our case $D = I_m$, and so we can apply the identity, we obtain:

$$A^{-1} = \begin{pmatrix} I_p & Q \\ -Q^t\Psi^{-1} & I_m \end{pmatrix}^{-1} = \begin{pmatrix} M^{-1} & -M^{-1}Q \\ Q^t\Psi^{-1}M^{-1} & I_m - Q^t\Psi^{-1}M^{-1}Q \end{pmatrix}$$

Thus,

$$\begin{pmatrix} u \\ f \end{pmatrix} = A^{-1}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} M^{-1} & -M^{-1}Q \\ Q^t\Psi^{-1}M^{-1} & I_m - Q^t\Psi^{-1}M^{-1}Q \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}$$

and therefore:

$$u = M^{-1}X - M^{-1}QY = M^{-1}(X - QY) \tag{3}$$

Now, $Cov(u) = Cov(M^{-1}(X - QY)) = M^{-1}Cov(X - QY)(M^{-1})^t$, and developing the covariance:

$$\begin{aligned}
Cov(X - QY) &= Cov(X) + Cov(X, -QY) + Cov(-QY, X) + Cov(-QY) \\
&= \Sigma_X - Cov(X,Y)Q^t - QCov(Y,X) - QCov(Y)(-Q)^t \\
&= \Sigma_X + 0 + QCov(Y)Q^t \\
&= (\Psi + QQ^t) + Q(I_m + Q^t\Psi^{-1}Q)Q^t \\
&= (I_p + QQ^t\Psi^{-1})\Psi + QQ^t + QQ^t\Psi^{-1}QQ^t \\
&= (I_p + QQ^t\Psi^{-1})\Psi + (I_p + QQ^t\Psi^{-1})QQ^t \\
&= (I_p + QQ^t\Psi^{-1})(\Psi + QQ^t) \\
&= (I_p + QQ^t\Psi^{-1})\Psi(I_p + \Psi^{-1}QQ^t) \\
&= M\Psi M^t
\end{aligned}$$

where we have used $\Sigma_X = QQ^t + \Psi$ by hypothesis, and $Cov(X,Y) = 0$ since Y is taken independently of X, therefore we obtain:

$$Cov(u) = M^{-1}Cov(X - QY)(M^{-1})^t = M^{-1}Cov(X - QY)(M^t)^{-1} = M^{-1}M\Psi M^t(M^t)^{-1} = \Psi$$

as we wanted to see.
Using similar arguments we will prove that $Cov(f) = I_m$, and so $i)$ will be done since we have already seen $E[f] = 0$. Concretely, we will now use the identity:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}$$

provided that A and $(D - CA^{-1}B)$ are invertible (Banachiewicz, 1937)(*1*). In our case: $D - CA^{-1}B = I_m - (-Q^t\Psi^{-1})I_pQ = I_m + Q^t\Psi^{-1}Q = W$ and we know that W is invertible, and $A = I_p$ invertible, so we can apply the identity to obtain:

$$A^{-1} = \begin{pmatrix} I_p & Q \\ -Q^t\Psi^{-1} & I_m \end{pmatrix}^{-1} = \begin{pmatrix} I_p - QW^{-1}Q^t\Psi^{-1} & -QW^{-1} \\ W^{-1}Q^t\Psi^{-1} & W^{-1} \end{pmatrix}$$

Thus,

$$\begin{pmatrix} u \\ f \end{pmatrix} = A^{-1}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} I_p - QW^{-1}Q^t\Psi^{-1} & -QW^{-1} \\ W^{-1}Q^t\Psi^{-1} & W^{-1} \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}$$

and we obtain

$$f = W^{-1}Q^t\Psi^{-1}X + W^{-1}Y = W^{-1}(Q^t\Psi^{-1}X + Y) \tag{4}$$

Then, $Cov(f) = Cov(W^{-1}(Q^t\Psi^{-1}X + Y)) = W^{-1}Cov(Q^t\Psi^{-1}X + Y)(W^{-1})^t$

and developing the covariance:

$$\begin{aligned}
Cov(Q^t\Psi^{-1}X+Y) &= Cov(Q^t\Psi^{-1}X)+Cov(Q^t\Psi^{-1}X,Y)+Cov(Y,Q^t\Psi^{-1}X)+Cov(Y)\\
&= Q^t\Psi^{-1}Cov(X)\Psi^{-1}Q+Q^t\Psi^{-1}Cov(X,Y)+Cov(Y,X)\Psi^{-1}Q+Cov(Y)\\
&= Q^t\Psi^{-1}\Sigma_X\Psi^{-1}Q+\Sigma_Y\\
&= Q^t\Psi^{-1}(QQ^t+\Psi)\Psi^{-1}Q+W\\
&= Q^t\Psi^{-1}(QQ^t\Psi^{1}Q+\Psi\Psi^{-1}Q)+W\\
&= Q^t\Psi^{-1}QQ^t\Psi^{-1}Q+Q^t\Psi^{-1}Q+W\\
&= (Q^t\Psi^{-1}Q+I_m)(Q^t\Psi^{-1}Q)+W\\
&= W(Q^t\Psi^{-1}Q)+W\\
&= W(Q^t\Psi^{-1}Q+I_m)\\
&= WW
\end{aligned}$$

and using that W is symmetric we get the desired result:

$$\begin{aligned}
Cov(f) &= W^{-1}Cov(Q^t\Psi^{-1}X+Y)(W^{-1})^t\\
&= W^{-1}Cov(Q^t\Psi^{-1}X+Y)(W^t)^{-1}\\
&= W^{-1}WW(W^t)^{-1}=W^{-1}WWW^{-1}\\
&= I_m
\end{aligned}$$

Finally, let's see *iii*), that is, $Cov(u,f)=0_{p\times m}$.
Using the expressions (3) and (4) we have:

$$\begin{aligned}
Cov(u,f) &= Cov(M^{-1}(X-QY),W^{-1}(Q^t\Psi^{-1}X+Y))\\
&= M^{-1}Cov(X-QY,Q^t\Psi^{-1}X+Y)(W^{-1})^t\\
&= M^{-1}[Cov(X,Q^t\Psi^{-1}X+Y)+Cov(-QY,Q^t\Psi^{-1}X+Y)](W^{-1})^t\\
&= M^{-1}[Cov(X,Q^t\psi^{-1}X)+Cov(X,Y)+Cov(-QY,Q^t\Psi^{-1}X)+Cov(-QY,Y)]W^{-1}\\
&= M^{-1}[Cov(X,Q^t\psi^{-1}X)+Cov(X,Y)-QCov(Y,X)(Q^t\Psi^{-1})^t-QCov(Y,Y)]W^{-1}\\
&= M^{-1}[Cov(X,X)\Psi^{-1}Q-QCov(Y,Y)]W^{-1}\\
&= M^{-1}[\Sigma_X\Psi^{-1}Q-Q\Sigma_Y]W^{-1}
\end{aligned}$$

We observe

$$\Sigma_X\Psi^{-1}Q-Q\Sigma_Y=(QQ^t+\Psi)\Psi^{-1}Q-QW=QQ^t\Psi^{-1}Q+Q-QW=Q(Q^t\Psi^{-1}Q+I_m)-QW=QW-QW=0_{p\times m}$$

so, $Cov(u,f)=0_{p\times m}$, and we are done. ■

Theorem 2.1.5 says that the orthogonal factor model has at least one solution, but the solution is not unique, in fact, every rotation of the factors will give another solution to the model:

Proposition 2.1.6. Let X be a random vector with $E[X]=0_{p\times1}$, let $m<p$ and let $G\in M_m(\mathbb{R})$ be an orthogonal matrix, that is $GG^t=I_m$, if (Q,f,u) is a solution to the orthogonal factor model for X, with m factors, then (QG,G^tf,u) is also a solution to the orthogonal factor model for X, with m factors.

Proof.

If (Q,f,u) is a solution to the orthogonal factor model for X with m factors and $G\in M_m(\mathbb{R})$ is orthogonal, we have $X=Qf+u=Q(GG^t)f+u=QG(G^tf)+u$, and it also holds
i) $E[G^tf]=G^tE[f]=0_{m\times1}$, $Cov(G^tf)=G^tCov(f)G=G^tI_mG=G^tG=I_m$

ii) $E[u] = 0_{p\times 1}$, $Cov(u) = \Psi$
iii) $Cov(G^t f, u) = G^t Cov(f, u) = 0_{m\times p}$

And therefore $(QG, G^t f, u)$ is a solution to the orthogonal factor model for X, with m factors. ■

In particular, the result holds when $G \in M_m(\mathbb{R})$ is orthogonal and $det(G) = 1$, that is, when G is a rotation matrix in $\mathbb{R}^m$. This fact will allow us to find interpretable factors, if we find a solution (Q, f, u) but the factors f are not interpretable in the context of investigation. There exists methods that try to find an adequate rotation G such that the factors $G^t f$ of the solution $(QG, G^t f, u)$ might be interpretable.

The theorem 2.1.5 indicates us how to proceed to find a solution to the model for an initial vector $X^t = (X_1, \ldots, X_p)$. If we find $Q \in M_{p\times m}(\mathbb{R})$ and $\Psi \in M_{p\times p}(\mathbb{R})$ diagonal and positive definite such that $\Sigma_X = QQ^t + \Psi$, the factors given by the expressions (3) and (4) will give a solution to the orthogonal factor model for X with matrix of loadings Q. In practice we have a data matrix $\widetilde{X} \in M_{n\times p}(\mathbb{R})$, where each row of $\widetilde{X}$ is an observation of the random vector $X^t = (X_1, \ldots, X_p)$ for which we want to fit the model, and we estimate Σ_X using the sample covariance matrix S, then, our objective will be to find matrices $\widehat{Q} \in M_{p\times m}(\mathbb{R})$ and $\widehat{\Psi} \in M_{p\times p}(\mathbb{R})$, with $\widehat{\Psi}$ diagonal and positive definite, such that the equality:

$$S = \widehat{Q}\widehat{Q}^t + \widehat{\Psi}$$

holds, at least approximately. If we find such matrices $\widehat{Q}$ and $\widehat{\Psi}$, they can be taken as loadings and specific variances estimates and so, they give rise to an estimate solution for the factor model. Finding $\widehat{Q}$ and $\widehat{\Psi}$ is our next objective.

2.2 The equation $S = QQ^t + \Psi$

We now consider $Q \in M_{p\times m}(\mathbb{R})$ and $\Psi \in M_{p\times p}(\mathbb{R})$ as unknown matrices, being the second a diagonal matrix with positive entries, whereas S is a positive definite known matrix, satisfying the equation

$$S = QQ^t + \Psi \tag{5}$$

We will use the notation $\widehat{Q}$ and $\widehat{\Psi}$ to refer to a known, adequate solution to (5), with known meaning that $\widehat{Q}$ and $\widehat{\Psi}$ only depend on the data in $\widetilde{X}$. $\widehat{Q}$ and $\widehat{\Psi}$ will be found using numerical methods.

We first observe that if we find $\widehat{Q} \in M_{p\times m}(\mathbb{R})$ and $\widehat{\Psi} \in M_{p\times p}(\mathbb{R})$, with $\widehat{\Psi}$ diagonal and positive definite, such that $S = \widehat{Q}\widehat{Q}^t + \widehat{\Psi}$, then, given any orthogonal matrix $G \in M_m(\mathbb{R})$:

$$(\widehat{Q}G)(\widehat{Q}G)^t + \widehat{\Psi} = (\widehat{Q}G)(G^t\widehat{Q}^t) + \widehat{\Psi} = \widehat{Q}\widehat{Q}^t + \widehat{\Psi} = S.$$

In practice, this is not a problem, since it allows to search for an interpretable factors (proposition 2.1.6), but from a numerical point of view the non uniqueness of $\widehat{Q}$ is a drawback. An usual technique is to impose additional restrictions on the matrices Q and Ψ to resolve this indetermination, then we will estimate Q and Ψ under the restrictions, and we will do rotations later, if necessary. Two usual restrictions are

$$Q^t Q \text{ is diagonal} \tag{6}$$

$$Q^t \Psi^{-1} Q \text{ is diagonal} \tag{7}$$

We will discuss the restriction (6), (7) can be discussed similarly (Peña, 2002, p. 361)(*13*). Let $m < p$ and let $\widehat{Q} \in M_{p\times m}(\mathbb{R})$ and $\widehat{\Psi} \in M_{p\times p}(\mathbb{R})$ satisfying $S = \widehat{Q}\widehat{Q}^t + \widehat{\Psi}$, let $\widehat{Q}^t\widehat{Q} = V\Lambda V^t$ be the spectral decomposition of $\widehat{Q}^t\widehat{Q}$, and let $Q_r = \widehat{Q}V$, then

$$Q_r^t Q_r = V^t\widehat{Q}^t\widehat{Q}V = V^t V\Lambda V^t V = \Lambda$$

since V is orthogonal. Hence Q_r satisfies the restriction (6), now let G be any orthogonal matrix and let $\widetilde{Q} = \widehat{Q}G$, then:

$$\widetilde{Q}^t\widetilde{Q} = G^t\widehat{Q}^t\widehat{Q}G$$

Thus, $\widetilde{Q}^t\widetilde{Q}$ will be diagonal if the columns of G form a basis of eigenvectors of $\widehat{Q}^t\widehat{Q}$, therefore, the only matrix $\widehat{Q}G$ satisfying (6) will be $Q_r = \widehat{Q}V$ except for the chose of the eigenvectors V, in fact, it can be shown that if another matrix $\widehat{Q}^*$ satisfies $S = \widehat{Q}^*(\widehat{Q}^*)^t + \widehat{\Psi}$, then $\widehat{Q}^* = \widehat{Q}G$, with G orthogonal, and therefore the only matrix satisfying (6) will be $Q_r = \widehat{Q}V$, except of V.

We want then to solve the equation (5) under one of the restrictions (6) or (7). The system may have -in the best case- an unique solution, depending on the number of initial variables p and the number of factors m. If the system has an infinite number of solutions, we will say that the factor model is undetermined or not well defined.

Concretely, $Q \in M_{p\times m}(\mathbb{R})$, so it has pm unknown parameters, whereas $\Psi \in M_{p\times p}(\mathbb{R})$, and we restrict the problem to Ψ diagonal, so Ψ has p unknown parameters an so we have $pm + p$ unknown parameters to estimate in the factor model. On the other hand, the matrix equation (5) define $\frac{1}{2}p(p+1)$ equations involving the unknown parameters, to see this, let $q_i = (q_{i1}, \ldots, q_{im})$ be the ith row of Q, for $i = 1, \ldots, p$, let ψ_i be the ith element on the diagonal of Ψ and let $S = (s_{ij})_{ij}$, then we can write the equation $QQ^t + \Psi = S$ as

$$\begin{pmatrix} \langle q_1, q_1\rangle + \psi_1 & \langle q_1, q_2\rangle & & \cdots & \langle q_1, q_p\rangle \\ & \langle q_2, q_2\rangle + \psi_2 & \langle q_2, q_3\rangle & \cdots & \langle q_2, q_p\rangle \\ & & \ddots & \ddots & \vdots \\ & & & & \langle q_{p-1}, q_p\rangle \\ & & & & \langle q_p, q_p\rangle + \psi_p \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & & \cdots & s_{1p} \\ & s_{22} & s_{23} & \cdots & s_{2p} \\ & & \ddots & \ddots & \vdots \\ & & & & s_{p-1p} \\ & & & & s_{pp} \end{pmatrix}$$

For example, the equation given by the elements in the position $(1,1)$ of the matrices is $\langle q_1, q_1\rangle + \psi_1 = s_{11}$, where the parameters in $\langle q_1, q_1\rangle + \psi_1$ are unknown and s_{11} is known. Thus, the number of different equations is the number of elements of S above the diagonal, plus the number of elements in the diagonal, that gives a total of $\frac{1}{2}p(p+1)$ different equations. Similarly, It can be shown that conditions (6) and (7) introduce $\frac{1}{2}m(m-1)$ equations involving the parameters of Q and Ψ. Hence, the number of different equations minus the number of parameters to estimate in the factor model is given by:

$$d = \frac{1}{2}p(p+1) + \frac{1}{2}m(m-1) - (pm + p). \tag{8}$$

We can have three situations:

$\mathbf{d < 0}$: In this case we have more (unknown) parameters than equations and therefore there is an infinite number of solutions to the system, and the model is undetermined.

$\mathbf{d = 0}$: In this case the number of equations is the same as the number of unknown parameters and hence there exist an unique and exact solution to the system, this situation is not possible in general (i.e. for any given p and $m < p$).

$\mathbf{d > 0}$: In this case there is no exact solution to the system; there are more equations that unknown parameters, and we will have approximate solutions to (5), we will search for $\widehat{Q}$ and $\widehat{\Psi}$ that minimize the errors of estimation, for example, in the least squares sense.

Evaluating d (8) we can know the maximum number of factors we can identify for a given set of initial variables, without the model being undetermined.

After discussing the equation $S = QQ^t + \Psi$ we are almost ready to see the method of Principal Factors, which, among other methods, try to find adequate estimates $\widehat{Q}$ and $\widehat{\Psi}$ satisfying the above equation, at least approximately. The method of Principal Factors will find solutions satisfying the restriction (6), and we can use (8) to choose a number of factors such that $d \geq 0$. Before seeing this method we must explain why, in practice, is usual to standardize the initial variables.

2.3 Use of the correlation matrix

We recall that in practice we will have the data summarized in a matrix $\widetilde{X} \in M_{n\times p}(\mathbb{R})$, where the rows in $\widetilde{X}$ will a sample of the random vector $X^t = (X_1, \dots, X_p)$.

In practice, it is habitual to standardize the data. Assuming the data is centred, by default, the standardization consist in applying the transformation:

$$\widetilde{X}_z = \widetilde{X} D^{-1/2}$$

where $D = diag(s_{11}, \dots, s_{pp})$ and s_{ii} is the sample variance of the variable X_i, we suppose $s_{ii} > 0$, $\forall i = 1, \dots, p$[3]. $\widetilde{X}_z$ is the standardized data matrix and their rows are a sample of the standardized initial random vector $X_z^t = (X_1/\sigma_1, \dots, X_p/\sigma_p)$, where $\sigma_i^2 = Var(X_i)$, being $\sigma_i > 0$, $\forall i = 1, \dots, p$.
Then, we search for a solution to the orthogonal factor model for X_z, using the data in $\widetilde{X}_z$. It holds that $\Sigma_{X_z} = R$, where R is the correlation matrix of X; $R = (Corr(X_i, X_j))_{ij}$, and it also holds $S_z = \widehat{R}$, where S_z is the sample covariance matrix of X_z and $\widehat{R}$ is the sample correlation matrix of X. Hence, to estimate a solution to the orthogonal factor model for the standardized vector X_z, we will have to find $\widehat{Q}$ and $\widehat{\Psi}$ such that

$$S_z = \widehat{R} = \widehat{Q}\widehat{Q}^t + \widehat{\Psi} \tag{9}$$

at least, approximately.

The reason why it is common to work with standardized data is illustrated in the remark 2.3.1.

Observation 2.3.1. If (Q, f, u) is a solution to the orthogonal factor model for X_z, then, using proposition 2.1.2, and denoting $C = diag(\sigma_1^2, \dots, \sigma_p^2)$ it holds:

$$\begin{aligned} Q &= Cov(X_z, f) = Cov(C^{-1/2}X, f) = C^{-1/2}Cov(X, f) = \\ &= (Cov(X_i, f_j)/\sigma_i)_{ij} = (Corr(X_i, f_j))_{ij} = Corr(X, f) \\ &= R_{Xf}. \end{aligned}$$

Where R_{Xf} denotes the matrix of correlations between X and the factors f.

Therefore, the matrix of loadings Q, that will be estimated by $\widehat{Q}$, is the matrix of correlations between the initial variables and the factors, making the loadings easier to interpret. Concretely, correlations, unlike covariances, don't depend on the units with which the variables are measured, and their absolute value is bounded by one, this make them comparable and a better indicator of relationship. Hence, if the loadings are the correlations between the factors and the initial variables, we might be able to interpret a factor in terms of the variables in which the factor have a loading near ±1, in other words, in terms of the variables that the factor explains the most. We will discuss the interpretation of the factors more precisely later on.

The last observation is useful because the hiddent factors f do not change when variables are rescaled, as it is stated in the next proposition

[3] Let $A = diag(a_{11}, \dots, a_{pp})$ be a diagonal matrix with $a_{ii} > 0$, $\forall i = 1, \dots, p$, $A^{-1/2}$ denote the matrix $(A^{-1})^{1/2} = diag((\frac{1}{a_{11}})^{1/2}, \dots, (\frac{1}{a_{pp}})^{1/2})$.

Proposition 2.3.2. Let X be the random vector of initial variables, let $X_z = C^{-1/2}X$ be the standardized vector, and let (Q_z, f, u_z) be a solution to the orthogonal factor model for the standardized variables in X_z, with m factors and with $Cov(u_z) = \Psi$. Then $(C^{1/2}Q_z, f, C^{1/2}u_z)$ is a solution to the orthogonal factor model for X, with m factors and with $Cov(C^{1/2}u_z) = C\Psi$

Proof. If (Q_z, f, u_z) is a solution to the orthogonal factor model for X_z then

$$X = C^{1/2}X_z = C^{1/2}(Q_z f + u_z) = C^{1/2}Q_z f + C^{1/2}u_z$$

and it also holds
i) $E[f] = 0_{m\times 1}$, $Cov(f) = I_m$
ii) $E[C^{1/2}u_z] = C^{1/2}E[u_z] = 0_{p\times 1}$, $Cov(C^{1/2}u_z) = C^{1/2}Cov(u_z)C^{1/2} = C^{1/2}\Psi C^{1/2} = C\Psi$, and $C\Psi$ is also diagonal.
iii) $Cov(f, C^{1/2}u_z) = Cov(f, u_z)C^{1/2} = 0_{m\times p}$. ■

Therefore, we might find interpretable factors searching a solution for the standardized variables, and then, if desired, we can use this factors to explain the unstandardized data X. In this case, if $\widehat{Q_z}$ is the estimated loadings matrix of X_z, $C^{1/2}$ will be estimated by $D^{1/2}$, and the loadings matrix of X will be estimated by $D^{1/2}\widehat{Q_z}$.

We remind that our objective is to find matrices $\widehat{Q} \in M_{p\times m}(\mathbb{R})$, with $m < p$, and $\widehat{\Psi} \in M_{p\times p}(\mathbb{R})$, with $\widehat{\Psi}$ diagonal and positive definite, such that the equality $S = \widehat{Q}\widehat{Q}^t + \widehat{\Psi}$ holds, at least approximately, for some m such that d, given by (8), satisfy $d \geq 0$. We're now ready to look at the principal factors method to find such matrices $\widehat{Q}$ and $\widehat{\Psi}$ for the standardized case $S = \widehat{R}$, which is justified by what we have seen above.

2.4 The principal factors method

Let $\widetilde{X}$ be the data matrix of observations of $X^t = (X_1, \dots, X_p)$ and X_z be the standardized vector. Let $m < p$ such that $d \geq 0$, let $\widehat{R}$ be the sample correlation matrix. We want to find $\widehat{Q} \in M_{p\times m}(\mathbb{R})$ and $\widehat{\Psi} \in M_{p\times p}(\mathbb{R})$, with $\widehat{\Psi}$ diagonal and positive definite, such that $\widehat{R} = \widehat{Q}\widehat{Q}^t + \widehat{\Psi}$, at least approximately, to fit the orthogonal factor model for X_z. Solutions with $\widehat{\Psi}$ positive semidefinite will also be permissible. We recall (observation 2.1.4) that if the factor model holds for $X_z = (X_{1z}, \dots, X_{pz})$, then

$$1 = Var(X_{iz}) = \sum_{j=1}^{m} q_{ij}^2 + \psi_i = h_i^2 + \psi_i \qquad (10)$$

since the variables are standardized.

The principal factors method is an iterative method based on the spectral decomposition, that needs initial estimates of the communalities h_i^2, the method follows the next steps:

1. Compute the sample correlation matrix $\widehat{R}$ using $\widetilde{X}$.
2. Compute initial estimates $\widehat{h_i^2}$ of the communalities. Let $\widehat{R} = (r(x_i, x_j))_{ij}$, two common estimates are:
   (a) $\widehat{h_i^2} = \max_{i\neq j} |r(x_i, x_j)|$
   (b) $\widehat{h_i^2} = R^2_{i.others}$, where $R^2_{i.others}$ is the multiple correlation coefficient of X_i with the other variables in X.

3. Compute the initial specific variances $\widehat{\psi_i} = 1 - \widehat{h_i^2}$, and set $\widehat{\Psi} = diag(\widehat{\psi_1}, \dots, \widehat{\psi_p})$.

4. The matrix $\widehat{R} - \widehat{\Psi}$ is symmetric and therefore we can consider its spectral decomposition $\widehat{R} - \widehat{\Psi} = V \Lambda V^t$, where $\Lambda = diag(\lambda_1, \dots, \lambda_p)$ with $\lambda_1 \geq \dots \geq \lambda_p$. Suppose that the first $m < p$ eigenvalues $\lambda_1 \geq \dots \geq \lambda_m$ are positive[4]. Let V_m be the matrix whose columns are the first m columns of V, in the same order, and let $\Lambda_m = diag(\lambda_1, \dots, \lambda_m)$, then set $\widehat{Q} = V_m \Lambda_m^{1/2}$, it holds:

$$\widehat{Q}\widehat{Q}^t = V_m \Lambda_m^{1/2} \Lambda_m^{1/2} V_m^t = V_m \Lambda_m V_m^t \approx \widehat{R} - \widehat{\Psi} \tag{11}$$

   It is known that this is the best approximation of rank m of $\widehat{R} - \widehat{\Psi}$ according to the Frobenius norm (in the least squares sense), in fact, if $\lambda_i = 0$, $\forall i > m$, $\widehat{Q}\widehat{Q}^t = \widehat{R} - \widehat{\Psi}$. Moreover, it holds:

$$\widehat{Q}^t \widehat{Q} = (\Lambda_m^{1/2})^t V_m^t V_m \Lambda_m^{1/2} = (\Lambda_m^{1/2})^t I_m \Lambda_m^{1/2} = \Lambda_m$$

   because the columns of V are orthogonal. Hence, $\widehat{Q}^t \widehat{Q}$ is diagonal and the restriction (6) is satisfied.

5. Redefine the specific variances in terms of $\widehat{Q}$: $\widehat{\psi_i} = 1 - \sum_{j=1}^m q_{ij}^2$, and set $\widehat{\Psi} = diag(\widehat{\psi_1}, \dots, \widehat{\psi_p})$, here $(q_{ij})_{ij} = \widehat{Q}$. Then, the equations in (10) hold exactly: $Var(X_{iz}) = 1 = \sum_{j=1}^m q_{ij}^2 - \widehat{\psi_i}$.

6. Repeat steps 4 and 5 until some convergence criterion is reached, for example until specific variances $\widehat{\psi_i}$ have converged to a stable value. If $\widehat{\psi_i} \geq 0, \forall i = 1, \dots, p$, the solution given by the estimates $\widehat{Q}$ and $\widehat{\Psi}$ is permissible. Values $\widehat{\psi_i}$ out of $[0, 1]$ may appear during the iteration, in this case they are forced to be 0 or 1, in step 4, some of the first m eigenvalues of $\widehat{R} - \widehat{\Psi}$ may be negative, this is not a serious problem if they are small and we can suppose them to be zero (Peña, p. 363)(*13*).

Observation 2.4.1. The approximation in (11) will be good if the eigenvalues $\lambda_{m+1}, \dots, \lambda_p$ are close to zero, that is the desired situation, but it may not be the case in general. The quality of the approximation can be evaluated directly comparing $\widehat{R}$ with the estimation $\widehat{Q}\widehat{Q}^t + \widehat{\Psi}$ given by the method, if it's not a good approximation, one option is to consider a model with more factors, without the model being undetermined. In fact a solution with $m = p$ factors will always exist, this solution is given by the principal components of X_z, concretely, let $\widehat{R} = V \Lambda V^t$ be the spectral decomposition of $\widehat{R}$, all the eigenvalues in Λ are non negative, since the sample correlation matrix is positive semidefinite, then, $X_z \approx (V \Lambda^{1/2}) \Lambda^{-1/2} Y$, where Y is the vector of principal components of X_z, hence, $((V\Lambda^{1/2}), \Lambda^{-1/2}Y, 0_{p\times 1})$ is an estimate solution of p factors to the orthogonal factor model for X_z, concretely $Var(\Lambda^{-1/2}Y) = \Lambda^{-1/2} Var(Y) \Lambda^{-1/2} \approx \Lambda^{-1/2} \Lambda \Lambda^{-1/2} = I_p$. This solution, however, is not desirable, the reason is that we want to explain the initial variables with a few common factors, and if we not allow the small errors given by the specific factors, we will need more common factors to hold the model.

2.5 Factor scores

Once the orthogonal factor model has been fitted, it may be of interest to have an estimation of the values that may take the factors for a fixed observation of the initial vector, this values are called factor scores. There is more than one option to choose for the factor scores, in our case, we will take as factor scores the expected value of the factors, conditioned to a given observation of the initial vector.

Let $X^t = (X_1, \dots, X_p)$ be the initial variables, and suppose $\Sigma_X = QQ^t + \Psi$ with $Q \in M_{p\times m}(\mathbb{R})$ with $m < p$ and $\Psi \in M_{p\times p}(\mathbb{R})$ diagonal an positive definite, then we have seen that the orthogonal factor model with m factors holds for X, with matrix of loadings Q, and concertely, the factors satisfying the model, given by the theorem 2.1.5 are ((3), (4)):

$$f = W^{-1}(Q^t \Psi^{-1} X + Y) \tag{12}$$

[4] The matrix $\widehat{R} - \widehat{\Psi}$, unlike $\widehat{R}$, can have negative eigenvalues.

$$u = M^{-1}(X - QY) \tag{13}$$

where: $W = I_m + Q^t\Psi^{-1}Q$, $M = I_p + QQ^t\Psi^{-1}$, and $Y \sim N_m(0_{m\times 1}, W)$. Our interest is focused only on common factors. Let $x_0 \in M_{p\times 1}(\mathbb{R})$ be an observation of the initial random vector X, It holds:

$$\begin{aligned} E[f \,|\, X = x_0] &= E[W^{-1}(Q^t\Psi^{-1}x_0 + Y)] = E[W^{-1}Q^t\Psi^{-1}x_0] + E[W^{-1}Y] = E[W^{-1}Q^t\Psi^{-1}x_0] + 0_{m\times 1} \\ &= W^{-1}Q^t\Psi^{-1}x_0 \end{aligned}$$

This will be taken as the factor scores corresponding to the observation x_0, and the following notation can be used:

$$f_{x_0} := E[f \,|\, X = x_0] = W^{-1}Q^t\Psi^{-1}x_0 \tag{14}$$

In general, we use f_x for a general x.

In practice, the unknown matrices Q and Ψ are replaced by it's respective estimates $\widehat{Q}$ and $\widehat{\Psi}$, also when computing W^{-1}. Let's see an equivalent expression for the factor scores:

Proposition 2.5.1. If $\Sigma_X = QQ^t + \Psi$ with $Q \in M_{p\times m}(\mathbb{R})$, and $\Psi \in M_{p\times p}(\mathbb{R})$ diagonal an positive definite, then $W^{-1}Q^t\Psi^{-1} = Q^t\Sigma_X^{-1}$

Proof.

$$\begin{aligned} W^{-1}Q^t\Psi^{-1} = Q^t\Sigma_X^{-1} &\iff (I_m + Q^t\Psi^{-1}Q)^{-1}Q^t\Psi^{-1} = Q^t(QQ^t + \Psi)^{-1} \\ &\iff (I_m + Q^t\Psi^{-1}Q)^{-1}Q^t\Psi^{-1}(QQ^t + \Psi) = Q^t \\ &\iff (I_m + Q^t\Psi^{-1}Q)^{-1}(Q^t\Psi^{-1}QQ^t + Q^t) = Q^t \\ &\iff (I_m + Q^t\Psi^{-1}Q)^{-1}(Q^t\Psi^{-1}Q + I_m)Q^t = Q^t \\ &\iff Q^t = Q^t \end{aligned}$$

The equality $Q^t = Q^t$ is true, therefore the equality $W^{-1}Q^t\Psi^{-1} = Q^t\Sigma_X^{-1}$ holds. ■

Hence, an equivalent expression for the factor scores (14) is:

$$f_x = Q^t\Sigma_X^{-1}x \tag{15}$$

for an observation $x \in M_{p\times 1}(\mathbb{R})$ of X.

The last expression (15) for the scores is known as *Thompson's factor scores* (Thompson, 1935)(*16*). An alternative approach to obtain this expression is by means of a regression argument and assuming that the initial random vector X has multivariate normal distribution (see Hardle and Simar, p. 322) (*4*). Using the conditional expectation argument, we haven't needed this last assumption.

As in (14), in practice, Q is replaced by $\widehat{Q}$, and Σ_X is replaced by $\widehat{\Sigma}_X = \widehat{Q}\widehat{Q}^t + \widehat{\Psi}$, although it can also by replaced by the sample covariance matrix S (Hardle and Simar, p. 323) (*4*), then we have estimates

$$\widehat{f}_x = \widehat{Q}^t\widehat{\Sigma}_X^{-1}x \tag{16}$$

We will also refer to the estimates as factor scores.

We can give an expression for the factor scores of the initial observations of X summarized in $\widetilde{X}$.
Let x_i^t be the ith row of $\widetilde{X}$; the ith observation of X, let $\widehat{f}_i := \widehat{f}_{x_i}$ be the factor scores of the ith observation, then $\widehat{f}_i = \widehat{Q}^t\widehat{\Sigma}_X^{-1}x_i$, and therefore:

$$(\widehat{f}_i)^t = (\widehat{Q}^t\widehat{\Sigma}_X^{-1}x_i)^t = x_i^t(\widehat{Q}^t\widehat{\Sigma}_X^{-1})^t = x_i^t\widehat{\Sigma}_X^{-1}\widehat{Q}$$

Therefore, denoting by F the matrix whose ith row are the factor scores of the ith observation: $(\widehat{f}_i)^t$, we have:

$$F = \widetilde{X}\widehat{\Sigma}_X^{-1}\widehat{Q} \tag{17}$$

This is an expression for the factor scores matrix for the whole set of the observations in $\widetilde{X}$.

2.6 Factors interpretation, and rotations

Interpreting the factors is understanding what they represent in the context of investigation. A factor is interpreted in terms of the variables with which it is more correlated (positively or negatively) and, therefore, the variables that the factor explains the most. Interpretation will not always be clear, in this case, there are methods to rotate the factors in a such way that the rotated ones may be easier to interpret.

Suppose that a m-factorial model was found to be reasonable for the standardized variables X_z, i.e. we have found adequate matrices $\widehat{Q}$ and $\widehat{\Psi}$ such that $\widehat{Q}\widehat{Q}^t + \widehat{\Psi}$ is a good approximation of $\widehat{R}$, in this case $\widehat{Q} \approx Corr(X, f)$, as we pointed in section 2.3, and we will use $\widehat{Q}$ to interpret the factors.

Let $X^t = (X_1, \dots, X_p)$ be the initial vector and let $f^t = (f_1, \dots, f_m)$ be the factors in that model. Denote $\widehat{Q} = (q_{ij})_{ij}$, and let $q_j = (q_{1j}, \dots, q_{pj})^t$ be the jth column of $\widehat{Q}$. The column q_j gives the correlations between the factor f_j and the initial variables. Suppose that a column q_j have values either close ± 1, or close to zero. A value q_{ij} close to 1 indicates positive relationship between the variable X_i and the factor f_j, i.e. X_i will be large when f_j is large, a value q_{ij} close to -1 indicates negative relationship between X_i and f_j, i.e. X_i will be large for large negative values of f_j, finally a value q_{ij} close to zero indicates no linear relationship between X_j and f_j. Thus, if a column q_j has values either close to ± 1 or close to zero, the factor f_j may be interpretable. On the other hand, if the columns have intermediate values, the factor f_j may be difficult to interpret.

We want then the columns of $\widehat{Q}$ to have values either close to ± 1 or close to zero, it is also desirable that every pair of columns of $\widehat{Q}$ have the loadings close to ± 1 on different rows, that is, each variable should be loaded highly on at most one factor. If all the columns of $\widehat{Q}$ have a few values close to ± 1 and the remaining loadings are close to zero, and each variable is loaded highly on at most one factor, we will say that the matrix of loadings $\widehat{Q}$ has a "*simple structure*". In this situation each variable is mainly explained by one single factor, each factor can be interpreted in terms of the variables that it explains the most, and all the factors might be interpretable. On the other hand, if the columns of $\widehat{Q}$ have too many intermediate values, the factors may not be interpretable.

For example, suppose that the initial variables $X^t = (X_1, \dots, X_p)$ are the qualifications in p different mental ability tests, and we have this qualifications for n individuals in a data matrix $\widetilde{X}$, suppose that we find a factor model fitting the data with only one factor f_1 that is positively correlated with all the test qualifications X_i, with correlations close to 1, that is, f_1 explains the qualifications of all the tests, and the larger is f_1, the larger will be this qualifications, then, the factor f_1 could be interpreted as the "overall level of intelligence" of an individual, ratifying this interpretation with the criteria of the experts in the matter. This last example was one of the first uses of factorial analysis. We refer to the "overall level of intelligence" as a qualitative aspect because in principle one would label the level of intelligence of an individual using qualitative values such as "high" or "low". If the factor f_1 is interpreted as the "overall level of intelligence", we then can measure this aspect quantitatively, as the score of f_1 for a given qualifications on the tests in X, hence, under this interpretation, high values of f_1 indicate a "high overall level of intelligence", whereas low values indicate a "low overall level of intelligence", since the correlation of f_1 with the tests qualifications is positive.

We recall (proposition 2.1.6) that, if (Q, f, u) is a solution to the orthogonal factor model for X_z with m

factors, and $G \in M_m(\mathbb{R})$ is an orthogonal matrix, then $(QG, G^t f, u)$ is also a solution to the orthogonal factor model for X_z, with m factors. If we fit the model for X_z but $\widehat{Q}$ has not a simple structure, there are methods that aim to provide a rotation matrix G such that $\widehat{Q}G$ has values either close to ±1 or close to zero, to make the rotated factors $G^t f$ interpretable. We will overview the method of the Varimax rotation, proposed by Kaiser (1958)(*8*), which is one of the most popular methods to get an adequate matrix of loadings.

2.6.1 The Varimax rotation

Let $\widehat{Q} \in M_{p\times m}(\mathbb{R})$ be the unrotated matrix of loadings of the factor model, let $\Delta = \widehat{Q}G$, with $G \in M_m(\mathbb{R})$ an unknown orthogonal matrix. The Varimax rotation is an iterative method that try to provide an adequate rotation G such that Δ has either values close to ±1 or close to zero, by maximizing a function of the rotated loadings Δ. We denote $\Delta = (\delta_{ij})_{ij}$, then the simplest version of the varimax method would consist on maximizing the function:

$$\Phi = \sum_{j=1}^{m}\sum_{i=1}^{p}(\delta_{ij}^2 - \bar{\delta_j})^2$$

where $\bar{\delta_j} = \frac{1}{p}\sum_{i=1}^{p}\delta_{ij}^2$, we observe that $\sum_{i=1}^{p}(\delta_{ij}^2 - \bar{\delta_j})^2$ is the sample variance of the jth column of Δ^2 (except a constant). Thus, this optimization was proposed expecting that if the column variances were maximized then the elements δ_{ij}^2 in the columns will be either close to 1 or close to 0 as desired. An improvement of the method was found by weighting the rows of Δ, concretely, let

$$d_{ij} = \frac{\delta_{ij}}{hi} \ , \bar{d_j} = \frac{1}{p}\sum_{i=1}^{p} d_{ij}^2$$

where $h_i = \sqrt{\langle q_i, q_i\rangle}$ is the square root of the ith communality, and q_i is the ith row of $\widehat{Q}$. The norm of the rows of Δ is equal to the norm of the rows of $\widehat{Q}$ since Δ is a rotation of $\widehat{Q}$, so the transformation makes the matrix $\Delta^* = (d_{ij})_{ij}$ to have unitary rows. Then, the function to maximize is:

$$\Phi = \sum_{j=1}^{m}\sum_{i=1}^{p}(d_{ij}^2 - \bar{d_j})^2$$

The maximization of this function is done numerically, under the restriction $\Delta = \widehat{Q}G$, with $G \in M_m(\mathbb{R})$ a rotation matrix.

3 Confirmatory factorial analysis

3.1 Introduction

In confirmatory factorial analysis we will search for a solution that generalizes the orthogonal factor model 2.1.1, now allowing correlations between the common factors. In this version the values of some parameters of the model are fixed in advance, and only the non fixed parameters are estimated.

Confirmatory factorial analysis, unlike the exploratory version, is used to test if the data fits a factor model with a prefixed structure. An use of confirmatory factorial analysis is to try to quantitatively model as factors qualitative aspects or hidden features that can't be directly measured, choosing as initial variables those indicators that are believed to be able to indirectly describe the aspects studied. In this case, the researcher has a certain amount of knowledge of the initial variables, and is in position to formulate hypothesis involving the factors of the model, for example fixing some loadings to be zero, and therefore choosing the variables that each factor can explain. If the model under this hypothesis fits the data, the factors might be interpreted as the aspects or hidden features of study. Is recommended to do an exploratory analysis before the confirmatory one, to choose a model that is not in contradiction with the observed data. When some values of the parameters of the model are fixed, we say that we formulate a hypothesis about the model, in the sense that we make a supposition of the structure of the model for our variables, that can be rejected if the imposed model doesn't fit the data. Usually, we will fix the value of some parameters of the model in such a way that the factors can be interpreted as desired, therefore, if the model holds, in the sense that it is well adjusted, rotations won't be necessary.

3.2 The model and the fundamental results

In this case we will allow correlations between the common factors, this is a more realistic assumption if we want to use them to model different aspects of interest, than ask them to be uncorrelated. We define:

Definition 3.2.1. Let $X^t = (X_1, \dots, X_p)$ be a $p \times 1$ random vector with $E[X] = 0_{p\times 1}$. We say that the factor model holds for X if there exist two random vectors $f^t = (f_1, \dots, f_m)$ with $m < p$ and $u^t = (u_1, \dots, u_p)$ and a matrix $Q = (q_{ij})_{ij} \in M_{p\times m}(\mathbb{R})$ such that

$$\begin{aligned} X_1 &= q_{11}f_1 + q_{12}f_2 + \cdots + q_{1m}f_m + u_1 \\ X_2 &= q_{21}f_1 + q_{22}f_2 + \cdots + q_{2m}f_m + u_2 \\ &\vdots \\ X_p &= q_{p1}f_1 + q_{p2}f_2 + \cdots + q_{pm}f_m + u_p \end{aligned}$$

In short: $X = Qf + u$, and satisfying:

i) $E[f] = 0_{m\times 1}$, $Cov(f) = \Theta$, with $\Theta \in M_m(\mathbb{R})$ symmetric and positive semidefinite.
ii) $E[u] = 0_{p\times 1}$, $Cov(u) = \Psi$, with $\Psi \in M_p(\mathbb{R})$ and diagonal.
iii) $Cov(f, u) = 0_{m\times p}$.

In this case we say that the triplet (Q, f, u) is a solution to the factor model for X. It is also usual to reefer to the factors as "latent" or "hidden" factors for X, in the sense that if the factor model holds for X, the known variables in X are explained by the factors, which are unknown before adjusting the model.

We observe that in this case, we don't restrict the common factors to be uncorrelated, concretely, we now allow $Cov(f) = \Theta$, with Θ any covariance matrix, unlike the orthogonal case, were we asked $Cov(f) = I_m$. The basic properties of the model are:

Proposition 3.2.2. Let X be a random vector with $E[X] = 0_{p\times 1}$. If the factor model holds for X and (Q, f, u) is a solution, then $Cov(X, f) = Q\Theta$.

Proof. Using that (Q, f, u) is a solution to the factor model 3.2.1, that is: $X = Qf + u$ satisfying $i), ii)$ and $iii)$, we have:

$$Cov(X, f) = Cov(Qf + u, f) = Cov(Qf, f) + Cov(u, f) = QCov(f, f) = Q\Theta. \ ■$$

Proposition 3.2.3. Let X be a random vector with $E[X] = 0_{p\times 1}$. If the factor model holds for X and (Q, f, u) is a solution, then $\Sigma_X = Q\Theta Q^t + \Psi$

Proof.

$$\begin{aligned} \Sigma_X = Cov(X) = Cov(Qf + u) &= Cov(Qf) + Cov(Qf, u) + Cov(u, Qf) + Cov(u) \\ &= QCov(f)Q^t + QCov(f, u) + Cov(u, f)Q^t + Cov(u) \\ &= Q\Theta Q^t + \Psi \end{aligned}$$

Where we have used the properties $i), ii)$ and $iii)$ of the solution. ■

Therefore, the necessary condition for the model to have solution is now $\Sigma_X = Q\Theta Q^t + \Psi$, with Θ and Ψ covariance matrices, with the second being diagonal. This condition will also be sufficient analogously to the orthogonal case, if we suppose Ψ positive definite. The result is given by the next theorem, which is a corollary of the theorem 2.1.5 of existence of solution to the orthogonal factor model.

Theorem 3.2.4. (Existence of solution to the factor model)
Let $X^t = (X_1, \dots, X_p)$ be a $p \times 1$ random vector with $E[X] = 0_{p\times 1}$. If there exist three matrices $Q \in M_{p\times m}(\mathbb{R})$, with $m < p$, $\Psi \in M_p(\mathbb{R})$, with Ψ diagonal and positive definite, and $\Theta \in M_m(\mathbb{R})$, with Θ symmetric and positive semidefinite, such that $\Sigma_X = Q\Theta Q^t + \Psi$, then there exist two random vectors $f^t = (f_1, \dots, f_m)$ and $u^t = (u_1, \dots, u_p)$ that satisfy the factor model 3.2.1 with loadings matrix $Q, Cov(f) = \Theta$ and $Cov(u) = \Psi$, that is; satisfying $X = Qf + u, Cov(f) = \Theta, Cov(u) = \Psi$ and $i), ii), iii)$.

Proof. Suppose $\Sigma_X = Q\Theta Q^t + \Psi$, with $Q \in M_{p\times m}(\mathbb{R})$, $\Psi \in M_p(\mathbb{R})$, with Ψ diagonal and positive definite, and $\Theta \in M_m(\mathbb{R})$, with Θ symmetric and positive semidefinite. Since Θ is symmetric, we can consider it's spectral decomposition $\Theta = V\Lambda V^t$, $\Lambda = diag(\lambda_1, \dots, \lambda_m)$, and since Θ is positive semidefinite $\lambda_1 \geq \dots \geq \lambda_m \geq 0$, therefore we can take $\Lambda^{1/2} = diag(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m})$ and write $\Theta = V\Lambda^{1/2}\Lambda^{1/2}V^t$. Now denote $Q_0 = QV\Lambda^{1/2}$, then:

$$\begin{aligned}\Sigma_X &= Q\Theta Q^t + \Psi = QV\Lambda V^t Q^t + \Psi = QV\Lambda^{1/2}\Lambda^{1/2}V^t Q^t + \Psi \\ &= (QV\Lambda^{1/2})(QV\Lambda^{1/2})^t + \Psi \\ &= Q_0 Q_0^t + \Psi\end{aligned}$$

Therefore $\Sigma_X = Q_0 Q_0^t + \Psi$, with $Q_0 \in M_{p\times m}(\mathbb{R})$ and $\Psi \in M_p(\mathbb{R})$, with Ψ diagonal and positive definite, so we are on the hypothesis of the theorem 2.1.5, hence, there exists two random vectors $f_0 = (f_{01}, \dots, f_{0m})^t$ and $u_0 = (u_{01}, \dots, u_{0p})^t$ that satisfy the orthogonal factor model for X with matrix of loadings Q_0 and $Cov(u_0) = \Psi$, that is, satisfying $X = Q_0 f_0 + u_0$, $Cov(u_0) = \Psi$, $E[u_0] = 0_{p\times 1}$, $Cov(f_0) = I_m$, $E[f_0] = 0_{m\times 1}$ and $Cov(f_0, u_0) = 0_{m\times p}$, now define the random vectors $f := V\Lambda^{1/2} f_0$ and $u := u_0$, let's see that this vectors give a solution to the factor model 3.2.1 for X with matrix of loadings Q, $Cov(u) = \Psi$, and $Cov(f) = \Theta$. It holds:

$$X = Q_0 f_0 + u_0 = QV\Lambda^{1/2} f_0 + u_0 = Qf + u$$

and it also holds:

$$Cov(u) = Cov(u_0) = \Psi \ , \ E[u] = E[u_0] = 0_{p\times 1}$$

$$\begin{aligned}Cov(f) &= Cov(V\Lambda^{1/2} f_0) = V\Lambda^{1/2} Cov(f_0)(V\Lambda^{1/2})^t = V\Lambda^{1/2} Cov(f_0)\Lambda^{1/2}V^t \\ &= V\Lambda^{1/2} I_m \Lambda^{1/2} V^t = V\Lambda^{1/2}\Lambda^{1/2}V^t = V\Lambda V^t \\ &= \Theta\end{aligned}$$

$$E[f] = E[V\Lambda^{1/2} f_0] = V\Lambda^{1/2} E[f_0] = 0_{m\times 1}$$

$$Cov(f, u) = Cov(V\Lambda^{1/2} f_0, u_0) = V\Lambda^{1/2} Cov(f_0, u_0) = 0_{m\times p}$$

Therefore, $X = Qf + u$ with f and u satisfying $i), ii)$ and $iii)$ with $Cov(f) = \Theta$ and $Cov(u) = \Psi$, so we are done. ■

3.3 Determination of the model

In view of the last theorem 3.2.4, to fit the model to the data in $\widetilde{X}$, we will estimate Σ_X using the sample covariance matrix S, and we will now have to find three matrices $\widehat{Q} \in M_{p\times m}(\mathbb{R})$, with $m < p$, $\widehat{\Theta} \in M_m(\mathbb{R})$, with $\widehat{\Theta}$ symmetric and positive semidefinite, and $\widehat{\Psi} \in M_p(\mathbb{R})$, with $\widehat{\Psi}$ diagonal and positive definite such that the equality

$$S = \widehat{Q}\widehat{\Theta}\widehat{Q}^t + \widehat{\Psi} \tag{18}$$

holds, at least approximately, if we find such matrices $\widehat{Q}$, $\widehat{\Theta}$, and $\widehat{\Psi}$, they will give raise to an estimate solution to the factor model. Hence, we will now try to satisfy, at least approximately, the equation:

$$S = Q\Theta Q^t + \Psi \tag{19}$$

where we think of $Q \in M_{p\times m}(\mathbb{R})$, with $m < p$, $\Theta \in M_m(\mathbb{R})$, with Θ symmetric and positive semidefinite, and $\Psi \in M_p(\mathbb{R})$, with Ψ diagonal and with positive entries, as unknown matrices, whereas S is a known symmetric positive semidefinte matrix.

As we mentioned in the introduction, now the objective is not finding an appropriate initial solution to the model and then do rotations if necessary, but try to fit from the beginning the model with the desired structure. We will impose the structure we want the model to have by fixing some values of the parameters of Q, Θ, and Ψ, and then we will search a solution to (19) under this restrictions, no further restrictions will we added if it's not necessary. It will only be necessary to estimate the non fixed parameters. We denote by t the number of free, non fixed parameters of Q, Θ, and Ψ, then, the degrees of freedom of the model are in this case given by:

$$d = \frac{1}{2}p(p+1) - t \tag{20}$$

this is, the number of different equations defined by (19) minus the number of free, unknown parameters to estimate. As in the orthogonal case, the factor model will be determined if $d \geq 0$ (Peña, p. 387)(*13*).

3.4 The maximum likelihood method

The Maximum Likelihood Estimation (MLE) method is a wellknown technique to find adequate estimates $\widehat{Q}$, $\widehat{\Theta}$ and $\widehat{\Psi}$ satisfying (19), at least approximately, and with the desired fixed values. The application of the method here is based on the assumption that the initial random vector $X^t = (X_1, \ldots, X_p)$ is multinormally distributed and hence, if the data deviates from this hypothesis, the found estimates may be spurious. On the other hand, if the data can be supposed multinormal, we may obtain good estimates and we will be able to properly test if the imposed model fits the data.

The MLE in the factor model case was successfully developed by Jöreskog (1967)(*5*) and (1969)(*6*). The idea is to suppose that the true covariance matrix of X, Σ_X, can be decomposed as $\Sigma_X = Q\Theta Q^t + \Psi$, with Q, Θ and Ψ with the desired fixed values, the estimates $\widehat{Q}$, $\widehat{\Theta}$ and $\widehat{\Psi}$ given by the method will be the maximum likelihood estimates under this hypothesis.

To simplfy notation, here we denote $\Sigma = \Sigma_X$. Suppose $X \sim N_p(0_{p\times 1}, \Sigma)$ and suppose $\Sigma = Q\Theta Q^t + \Psi$, with $Q \in M_{p\times m}(\mathbb{R})$, $\Theta \in M_m(\mathbb{R})$, with Θ symmetric and positive semidefinite, and $\Psi \in M_p(\mathbb{R})$, with Ψ diagonal and with positive entries, for a fixed $m < p$ and with some fixed values in Q, Θ and Ψ such that d in (20) satisfies $d \geq 0$. Σ is invertible as Ψ is positive definite. Let $\widetilde{X}$ be the data matrix of observations of X, where we denote by x_i^t the ith row of $\widetilde{X}$, that is, the ith observation of X. The likelihood function for the observations $(x_1, \ldots, x_n)$ in $\widetilde{X}$ is given by:

$$L(\widetilde{X}; \mu, \Sigma) = \frac{1}{|2\pi\Sigma|^{n/2}} \exp\Big\{ -\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^t \Sigma^{-1}(x_i - \mu)\Big\}$$

where $|\cdot|$ denotes the determinant, and the log-likelihood function for $\widetilde{X}$ is given by:

$$l(\widetilde{X}; \mu, \Sigma) = \log(L(\widetilde{X}; \mu, \Sigma)) = -\frac{n}{2}\log(|2\pi\Sigma|) - \frac{n}{2}tr(\Sigma^{-1}S) - \frac{n}{2}(\bar{x} - \mu)^t\Sigma^{-1}(\bar{x} - \mu)$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \;\text{ and }\; S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^t \tag{21}$$

(see for example Mardia et al., 1979, pp. 96-97)(*10*). We suppose the data is centered, so $\bar{x} = 0_{p\times 1}$, and we suppose $E[X] = \mu = 0_{p\times 1}$, then the log-likelihood function becomes:

$$l(\widetilde{X}; 0, \Sigma) = -\frac{n}{2}(\log(|2\pi\Sigma|) + tr(\Sigma^{-1}S))$$

we will denote $l(\Sigma) := l(\widetilde{X}; 0, \Sigma)$, so we have

$$l(\Sigma) = -\frac{n}{2}(\log(|2\pi\Sigma|) + tr(\Sigma^{-1}S)) \tag{22}$$

Maximizing $l(\Sigma)$ is equivalent to maximizing $L(\widetilde{X}, 0, \Sigma)$, since the logarithm is a strictly increasing function. For convenience, Joreskog (1967, p. 5)(*5*), propose to minimize the function:

$$F(\Sigma) = -\frac{2}{n}l(\Sigma) - \log(2\pi S) - p$$

instead of maximizing $l(\Sigma)$, which is equivalent, since the maximization is done over Σ and the term $\log(2\pi S) - p$ is fixed by the observations and the number of initial variables. Developing $F(\Sigma)$ we have:

$$\begin{aligned}
F(\Sigma) &= -\frac{2}{n}l(\Sigma) - \log(2\pi S) - p = \log(|2\pi\Sigma|) + tr(\Sigma^{-1}S) - \log(2\pi S) - p \\
&= tr(\Sigma^{-1}S) - p + \log(|2\pi\Sigma|) - \log(|2\pi\Sigma\Sigma^{-1}S|) \\
&= tr(\Sigma^{-1}S) - p + \log(\frac{|2\pi\Sigma|}{|2\pi\Sigma\Sigma^{-1}S|}) \\
&= tr(\Sigma^{-1}S) - p + \log(\frac{|2\pi\Sigma|}{|2\pi\Sigma||\Sigma^{-1}S|}) \\
&= tr(\Sigma^{-1}S) - p + \log(\frac{1}{|\Sigma^{-1}S|}) \\
&= tr(\Sigma^{-1}S) - \log(|\Sigma^{-1}S|) - p
\end{aligned}$$

Thus,

$$F(\Sigma) = tr(\Sigma^{-1}S) - \log(|\Sigma^{-1}S|) - p \tag{23}$$

We observe that this is a discrepancy function between S and Σ; as closer $F(\Sigma)$ is to zero, we can expect S be a better estimation of Σ, we will use this function in the next section to give a statistic to test the goodness of fit of the model to the data. Finally, replacing Σ in (22) by it's supposed decomposition $\Sigma = Q\Theta Q^t + \Psi$, we obtain:

$$F(Q, \Theta, \Psi) = tr((Q\Theta Q^t + \Psi)^{-1}S) - \log(|(Q\Theta Q^t + \Psi)^{-1}S|) - p \tag{24}$$

This is the function to minimize, concretely, let Ω_0 be the set of all matrices $M \in M_p(\mathbb{R})$, such that $M = Q\Theta Q^t + \Psi$ with $Q \in M_{p\times m}(\mathbb{R})$, $\Theta \in M_m(\mathbb{R})$ symmetric and positive semidefinite, and $\Psi \in M_p(\mathbb{R})$ diagonal and with positive entries, and with the desired fixed values on Q, Θ and Ψ. We will denote by $\widehat{Q}$, $\widehat{\Theta}$ and $\widehat{\Psi}$ the values of Q, Θ and Ψ minimizing $F(Q, \Theta, \Psi)$ (24) in the region given by Ω_0, that is, $\widehat{Q}\widehat{\Theta}\widehat{Q}^t + \widehat{\Psi} \in \Omega_0$ and gives the minimum of $F(Q, \Theta, \Psi)$ over all matrices $M \in \Omega_0$. The objective of the method is finding $\widehat{Q}$, $\widehat{\Theta}$ and $\widehat{\Psi}$. These matrices are the maximum likelihood estimates of Q, Θ and Ψ under the hypothesis $\Sigma \in \Omega_0$. The maximum likelihood estimate of Σ under this hypothesis is

$$\widehat{\Sigma} = \widehat{Q}\widehat{\Theta}\widehat{Q}^t + \widehat{\Psi} \tag{25}$$

Jöreskog (1967)(*5*) and (1969)(*6*) developed a numerical algorithm to find $\widehat{Q}$, $\widehat{\Theta}$ and $\widehat{\Psi}$. We won't see the algorithm here, but as a point, to avoid non positive definite solutions of Ψ, Jöresekog restricts the minimization of Ψ to a region R_ϵ such that $\psi_i \geq \epsilon$, for all $i = 1, \ldots, p$, for a prefixed small $\epsilon > 0$, sometimes the minimizing value of Ψ may be found on the boundary of R_ϵ, in this case further decrease of $F(Q,\Theta,\Psi)$ might be done over Ψ being positive definite, and the solution given by the estimates is said to be improper.

In confirmatory factorial analysis is also usual to standardize the initial variables, in this case we search a solution to the factor model for the standardized initial vector X_z, in the discussion above, for the standardized case, the maximum likelihood method will use $\Sigma_{X_z} = R$, and $S_z = \frac{n-1}{n}\widehat{R}$, where Σ_{X_z} is the true covariance matrix of X_z, and S_z is the sample covariance matrix of X_z given in (21), and R and $\widehat{R}$ are the true and sample correlation matrices of X, respectively, the common factors of a solution for X_z are also valid for X analogously to the orthogonal case (proposition 2.3.2).

3.5 The goodness of fit test

The goodness of fit test is an advantage if we can assume the data to be multinormal and use the maximum likelihood method to fit the factor model, it allows to test if the fit of the imposed model to the data in $\widetilde{X}$ is good, or if by cons the model is in contradiction with the observed data.

Let Ω_0 be the set defined in the above section, of all the matrices with the desired structure for Σ, and let $\Omega_1 = \{M \in M_p(\mathbb{R})$; M is symmetric and positive definite$\}$, we have $\Omega_0 \subseteq \Omega_1$. The goodness of fit test is a hypothesis test to test $H_0 : \Sigma \in \Omega_0$ against the alternative hypothesis $H_1 : \Sigma \in \Omega_1$. The test uses the likelihood-ratio technique. Let:

$$l_0 = \max_{\Sigma \in \Omega_0} l(\Sigma) \text{ and } l_1 = \max_{\Sigma \in \Omega_1} l(\Sigma)$$

where $l(\Sigma)$ is the log-likelihood function (22), $l_0 = l(\widehat{\Sigma})$, where $\widehat{\Sigma}$ is the maximum likelihood estimate of Σ under H_0, that is, $\widehat{\Sigma} = \widehat{Q}\widehat{\Theta}\widehat{Q}^t + \widehat{\Psi}$ (25), in the other hand, supposing that S in (21) is positive definite, we have $l_1 = l(S)$, developing:

$$\begin{aligned} l_1 = l(S) &= -\frac{n}{2}(\log(|2\pi S|) + tr(S^{-1}S)) = -\frac{n}{2}(\log(|2\pi S|) + p) = -\frac{n}{2}(\log((2\pi)^p|S|) + p) \\ &= -\frac{n}{2}[\log(|S|) + p + \log((2\pi)^p)] \end{aligned}$$

$$\begin{aligned} l_0 = l(\widehat{\Sigma}) &= -\frac{n}{2}(\log(|2\pi\widehat{\Sigma}|) + tr(\widehat{\Sigma}^{-1}S)) = -\frac{n}{2}(\log((2\pi)^p|\widehat{\Sigma}|) + tr(\widehat{\Sigma}^{-1}S)) \\ &= -\frac{n}{2}[\log(|\widehat{\Sigma}|) + tr(\widehat{\Sigma}^{-1}S) + \log((2\pi)^p)] \end{aligned}$$

Let λ be the likelihood ratio to test H_0 against H_1 then:

$$\begin{aligned} -2\log(\lambda) = 2(l_1 - l_0) &= 2(-\frac{n}{2}[\log(|S|) + p + \log((2\pi)^p)] + \frac{n}{2}[\log(|\widehat{\Sigma}|) + tr(\widehat{\Sigma}^{-1}S) + \log((2\pi)^p)]) \\ &= -n\log(|S|) - np - n\log((2\pi)^p) + n\log(|\widehat{\Sigma}|) + n\text{tr}(\widehat{\Sigma}^{-1}S) + n\log((2\pi)^p) \\ &= -n\log(|S|) + n\log(|\widehat{\Sigma}|) + n\text{tr}(\widehat{\Sigma}^{-1}S) - np \\ &= n(\log(|\widehat{\Sigma}|) - \log(|S|) + \text{tr}(\widehat{\Sigma}^{-1}S) - p) \\ &= n(-\log(|S|/|\widehat{\Sigma}|) + \text{tr}(\widehat{\Sigma}^{-1}S) - p) \\ &= n(\text{tr}(\widehat{\Sigma}^{-1}S) - \log(|\widehat{\Sigma}|^{-1}|S|) - p) \end{aligned}$$

We observe that $-2\log(\lambda) = nF(\Sigma)$, this is the statistic we will use to test $H_0 : \Sigma \in \Omega_0$ against $H_1 : \Sigma \in \Omega_1$. Let

$$U := -2\log(\lambda) = n(\text{tr}(\widehat{\Sigma}^{-1}S) - \log(|\widehat{\Sigma}|^{-1}|S|) - p) \tag{26}$$

It is known that, in the Gaussian case, U has an asymptotic χ^2_d distribution as $n \to \infty$, where d are the degrees of freedom of the model under H_0 (see for example Mardia et al. pp. 123-124)(*10*), we have seen

$d = \frac{1}{2}p(p+1) - t$ (20), where t is the number of free (non fixed) parameters of the model. As usual, if U_o is the observation of the statistic for the data in $\widetilde{X}$, we reject H_0 with a confidence level of $1-\alpha$ if $U_o > \chi^2_{d(1-\alpha)}$, where $\chi^2_{d(1-\alpha)}$ is the $(1-\alpha)$th percentile of χ^2_d, that is, $P\{U \leq \chi^2_{d(1-\alpha)}\} = 1-\alpha$. The imposed model will not contradict our data in $\widetilde{X}$ if H_0 is not rejected, in this case values of U_o close to zero will indicate a good fit of the model to the data.

3.6 Other methods of estimation

The assumption that the vector of initial variables is multinormal is restrictive. There exist other methods to fit the factor model to the data that doesn't need an assumption on the distribution of the initial random vector, one popular example is the method of the generalized least squares (Jöreskog, 1971)(*7*), which it's simplest version would consist on minimizing the discrepancy function:

$$F(\Sigma) = \mathrm{tr}((S-\Sigma)(S-\Sigma)^t)$$

Under the restriction $\Sigma = Q\Theta Q^t + \Psi$, with Q, Θ and Ψ with some fixed values. Denoting by $e_i = (e_{i1}, \ldots, e_{ip})$ the ith row of $S - \Sigma$, we have:

$$F(\Sigma) = \mathrm{tr}((S-\Sigma)(S-\Sigma)^t) = \sum_{i=1}^{p} \langle e_i, e_i \rangle = \sum_{i=1}^{p}\sum_{j=1}^{p} e_{ij}^2$$

Thus, the method search for the estimates $\widehat{Q}$, $\widehat{\Theta}$, and $\widehat{\Psi}$, minimizing the sum of the squared residuals between S and $\widehat{\Sigma} = \widehat{Q}\widehat{\Theta}\widehat{Q} + \widehat{\Psi}$.

3.7 Factor scores

Once the factor model has been fitted, we could be interested on the factor scores for a given observation of the initial variables. As in the orthogonal case, we will take as factor scores the expected value of the factors, conditioned to a given observation of the initial vector.

Let $X^t = (X_1, \ldots, X_p)$ be the random vector of initial variables, and suppose $\Sigma_X = Q\Theta Q^t + \Psi$ with $Q \in M_{p\times m}(\mathbb{R})$, with $m < p$, with $\Theta \in M_m(\mathbb{R})$ symmetric and positive semidefinite and with $\Psi \in M_p(\mathbb{R})$ diagonal and positive definite, then, we have seen that the factor model with m factors holds for X, with matrix of loadings Q, concretely, recalling theorem 3.2.4, let $\Theta = V\Lambda V^t$ be the spectral decomposition of Θ, the factors satisfying the factor model given by theorem 3.2.4 are:

$$f = V\Lambda^{1/2} f_0 \;\text{ and }\; u = u_0$$

Where f_0 and u_0 are the factors satisfying the orthogonal factor model for the decomposition $\Sigma_X = Q_0 Q_0^t + \Psi$, with $Q_0 = QV\Lambda^{1/2}$, given by the theorem 2.1.5. Hence, recalling (12) and (13), we have

$$f = V\Lambda^{1/2} f_0 = V\Lambda^{1/2} W^{-1}(Q_0^t \Psi^{-1} X + Y) \tag{27}$$

$$u = u_0 = M^{-1}(X - Q_0 Y) \tag{28}$$

where: $W = I_m + Q_0^t \Psi^{-1} Q_0$, $M = I_p + Q_0 Q_0^t \Psi^{-1}$, and $Y \sim N_m(0_{m\times 1}, W)$. Let $x \in M_{p\times 1}(\mathbb{R})$ be an observation of the initial random vector X, it holds:

$$\begin{aligned} E[f \,|\, X = x] &= E[V\Lambda^{1/2} W^{-1}(Q_0^t \Psi^{-1} x + Y)] \\ &= E[V\Lambda^{1/2} W^{-1} Q_0^t \Psi^{-1} x] + E[V\Lambda^{1/2} W^{-1} Y] \\ &= V\Lambda^{1/2} W^{-1} Q_0^t \Psi^{-1} x + V\Lambda^{1/2} W^{-1} E[Y] \\ &= V\Lambda^{1/2} W^{-1} Q_0^t \Psi^{-1} x + 0_{m\times 1} \\ &= V\Lambda^{1/2} W^{-1} Q_0^t \Psi^{-1} x \end{aligned}$$

we show (proposition 2.5.1) $W^{-1}Q_0^t\Psi^{-1} = Q_0^t\Sigma_X^{-1}$, therefore, we have:

$$\begin{aligned} E[f \mid X = x] &= V\Lambda^{1/2}W^{-1}Q_0^t\Psi^{-1}x = V\Lambda^{1/2}Q_0^t\Sigma_X^{-1}x \\ &= V\Lambda^{1/2}(QV\Lambda^{1/2})^t\Sigma_X^{-1}x = V\Lambda^{1/2}\Lambda^{1/2}V^tQ^t\Sigma_X^{-1}x \\ &= \Theta Q^t\Sigma_X^{-1}x \end{aligned}$$

thus, denoting $f_x := E[f \mid X = x]$, we have:

$$f_x = \Theta Q^t\Sigma_X^{-1}x \tag{29}$$

This will be taken as the factor scores corresponding to the observation x. In practice Θ, Q and Σ_X are replaced by it's respective estimates $\widehat{\Theta}$, $\widehat{Q}$ and $\widehat{\Sigma}_X = \widehat{Q}\widehat{\Theta}\widehat{Q}^t + \widehat{\Psi}$, Σ_X can also be replaced by the sample covariance matrix, we will denote

$$\widehat{f_x} = \widehat{\Theta}\widehat{Q}^t\widehat{\Sigma}_X^{-1}x \tag{30}$$

Let x_i^t be the ith row of the data matrix $\widetilde{X}$, let $\widehat{f_i} := \widehat{f_{x_i}}$ be the factor scores of the ith observation, then $\widehat{f_i} = \widehat{\Theta}\widehat{Q}^t\widehat{\Sigma}_X^{-1}x_i$, we have:

$$(\widehat{f_i})^t = (\widehat{\Theta}\widehat{Q}^t\widehat{\Sigma}_X^{-1}x_i)^t = x_i^t(\widehat{\Theta}\widehat{Q}^t\widehat{\Sigma}_X^{-1})^t = x_i^t\widehat{\Sigma}_X^{-1}\widehat{Q}\widehat{\Theta}$$

Therefore, denoting by F the matrix whose ith row are the factor scores of the ith observation: $(\widehat{f_i})^t$, we have:

$$F = \widetilde{X}\widehat{\Sigma}_X^{-1}\widehat{Q}\widehat{\Theta} \tag{31}$$

This is an expression for the factor scores matrix for the whole set of observations in $\widetilde{X}$.

3.8 Factors interpretation

In confirmatory factorial analysis, in order to obtain interpretable hidden or latent factors, the initial variables are also usually standardized, we will search for a solution to the factor model for the standardized initial vector X_z, and this factors will also be valid for the initial vector X, as in the orthogonal case. We will also ask another general condition to make the factors interpretable: we will ask the common factors to have unit variance, we remind that Θ is the covariance matrix of the common factors, hence, this last condition will be achieved by fixing the values on the diagonal of Θ to be 1, before fitting the model. Under these assumptions the factors can also be interpreted using the matrix of loadings Q, although in this case it don't give the covariances (or correlations in the standardized case), between the initial variables and the common factors. In practice we will use the estimate $\widehat{Q}$ of Q.

Concretely, suppose that (Q, f, u) is a solution to the factor model for $X_z^t = (X_{z1}, \dots, X_{zp})$, with m factors, then:

$$\begin{aligned} X_{z1} &= q_{11}f_1 + q_{12}f_2 + \cdots + q_{1m}f_m + u_1 \\ X_{z2} &= q_{21}f_1 + q_{22}f_2 + \cdots + q_{2m}f_m + u_2 \\ &\vdots \\ X_{zp} &= q_{p1}f_1 + q_{p2}f_2 + \cdots + q_{pm}f_m + u_p \end{aligned}$$

if all the common factors have unit variance, then a loading q_{ij} measure the part of the variability of X_{iz} uniquely explained by f_j, and the loadings are comparable, in the sense that if $|q_{ij}| = |q_{rk}|$, then f_j explains the same part of variability of X_{iz} that f_k explains of X_{rz}. Since standardization only changes the scale of the initial variables, the loadings also measure the part of variability of the initial variables explained by the factors. If the factors have different variances, the loadings aren't comparable, for example if $q_{11} > q_{12} > 0$ and $q_{13} = \cdots = q_{1p} = 0$, but $Var(f_2)$ is large and $Var(f_1)$ is close to zero, large values of X_{z1} may correspond to the large values of f_2, whereas f_1 will take small values and they will not affect the values of X_{z1}, thus a large loading will not necessarily mean a large part of variability explained, standardization of the initial variables is also necessary to use the loadings as a comparable measure between different variables of the variability explained by the factors, for example suppose now that (Q, f, u) is a solution for the initial vector X, $Q = (q_{ij})_{ij}$, suppose $q_{11} = q_{21} = 1$ and $Var(f_1) = 1$, if $Var(X_1) = 1$ but $Var(X_2) = 100$, then X_2 can't be uniquely explained by f_1, whereas is possible to have $X_1 = f_1$. It is also desirable that the specific factors u have small variances, so that the initial variables are mainly explained by the common factors.

Therefore we will search for a solution to the factor model for the standardized variables, asking the common factors to have unit variance, in this situation, the interpretation of the factors may be straightforward if each column of the loadings matrix $\widehat{Q}$ have a few large (positive or negative) loadings whereas the remaining loadings of the column are close to zero, and if every pair of columns of $\widehat{Q}$ have the large loadings on different rows, that is, if each variable is loaded highly on at most one factor, in this case we will say that the matrix of loadings $\widehat{Q}$ has a "*simple structure*", as in the orthogonal case. In this situation each variable is mainly explained by one single factor, and a factor can be interpreted in terms of the variables that it explains the most, the simple structure can be achieved, for example, by fixing in advance some loadings to zero. Setting certain weights to be zero, the factors might be interpreted as some hidden features of interest related with the initial variables, and therefore they could be used to model them.

Observation 3.8.1. We note that a factor f_j can be correlated with a variable X_i although $q_{ij} = 0$. In this case the factor f_j may contribute indirectly to explain X_i, trough the factors f_k correlated with f_j, and with a large loading q_{ik}. In fact, if (Q, f, u) is a solution to the factor model for X_z, and the factors in f have unit variance, then, using proposition 3.2.2, and denoting $C = diag(\sigma_1^2, \ldots, \sigma_p^2)$, where $\sigma_i^2 = Var(X_i)$, it holds:

$$\begin{aligned} Q\Theta &= Cov(X_z, f) = Cov(C^{-1/2}X, f) = C^{-1/2}Cov(X, f) = \\ &= (Cov(X_i, f_j)/\sigma_i)_{ij} = \big(\frac{Cov(X_i, f_j)}{\sigma_i \cdot 1}\big)_{ij} = \big(\frac{Cov(X_i, f_j)}{\sqrt{Var(X_i)}\sqrt{Var(f_j)}}\big)_{ij} = (Corr(X_i, f_j))_{ij} = Corr(X, f) \\ &= R_{Xf}. \end{aligned}$$

Thus, the correlations between the initial variables in X and the factors in f are given by $Q\Theta$, in practice $Q\Theta$ is estimated by $\widehat{Q}\widehat{\Theta}$, where $\widehat{Q}$ and $\widehat{\Theta}$ are the estimates of Q and Θ, respectively.

4 Sustainable progress indicators based on confirmatory factorial analysis

IERMB provided us from a data set of observations of a random vector of social, economic, ecological and urban variables, corresponding to metropolitan regions of Europe in the year 2016, the first step in the process to obtain sustainable progress indicators for this regions, is to adjust an adequate factor model model to this variables, using confirmatory factorial analysis, in such a way that the few new factors can, hopefully, be interpreted as some environmental, urban and socioeconomic aspects explaining the initial variables.

4.1 The 2016 data set

After an exploratory analysis of the data, we found reasonable a factor model with 4 factors for 10 of the initial 14 variables provided by IERMB, the 4 missing variables where discarded for various reasons; one variable was discarded because it was equal to a rescaled variable in the 10 chosen for the analysis, and hence, it won't give extra information, the others were discarded because of a poor explanation of them unless more factors were taken, this variables could be taken into account in a future analysis; as we said in the introduction, we built this example only to illustrate the process of obtaining sustainable progress indicators, and to show an example of confirmatory factorial analysis. Concretely, the 10 chosen variables are given in the following list.

- **PIC** Gross domestic product per capita.
- **NPT** Number of registered patents per capita.
- **TAR** Unemployment rate (% of unemployed workers over active worker population).
- **CEP** Primary energy consumption per capita (kilotonnes of oil equivalent).
- **GEH** Greenhouse gases per capita (thousands of tonnes of CO_2 equivalent).
- **GUR** Urbanization rate (% of urban surface over total surface).
- **DPB** Population density (population over urban surface).
- **PAI** Weight of industrial sector activity (% over total activity).
- **PAS** Weight of services sector activity (% over total activity).
- **DPR** Productive diversity (adimensional value).

The given observations of the variable NPT do not correspond to 2016 but to 2012, we have used NPT anyway in lack of its 2016 values, we can think its values correspond to 2016, for the only purpose of showing the procedure of obtaining sustainable progress indicators, however, for this reasons and the ones mentioned in the introduction, the results of the analysis can't be taken as meaningful or valid to draw conclusions concerning the metropolitan regions. The rest of the variables given observations correspond to 2016. The data set used consists of an observation of the vector of the above variables for each of a set of 95 Europe metropolitan regions, concretely, let

$$X^t = (\mathrm{PIC}, \mathrm{NPT}, \mathrm{TAR}, \mathrm{CEP}, \mathrm{GEH}, \mathrm{GUR}, \mathrm{DPB}, \mathrm{PAI}, \mathrm{PAS}, \mathrm{DPR}). \tag{32}$$

be the vector of initial variables. We used a data matrix $\widetilde{X} \in M_{95\times 10}(\mathbb{R})$ where each row is an observation $x_i^t = (x_{i1}, \ldots, x_{i10})$ of X^t corresponding to one of the 95 regions and on the year 2016 (except the observation of NPT), and we had one observation for each region.

4.2 The model

The model was fitted for the standardized variables for the reasons explained in section 3.8 and hence using the correlation matrix $\widehat{R}$ of the initial vector X, computed using $\widetilde{X}$. We adjusted a factor model with 4 factors to the (standardized) data, using confirmatory factorial analysis. We fixed some parameters of the model as usual, concretely, we fixed some of the factor loadings to zero, and we fixed to unit the common factors variances, to allow interpretations. This restrictions were in concordance with the initial exploratory factorial analysis, that is, we only fixed to zero values that were small in the exploratory analysis, to obtain a model with a simple structure without contradicting the observed data. The fixed structure in Q, Θ, and Ψ is given by:

$$Q = \begin{pmatrix} q_{11} & 0 & 0 & q_{14} \\ q_{21} & 0 & 0 & 0 \\ q_{31} & q_{32} & 0 & 0 \\ 0 & q_{42} & 0 & 0 \\ 0 & q_{52} & 0 & 0 \\ 0 & 0 & q_{63} & 0 \\ 0 & 0 & q_{73} & 0 \\ 0 & 0 & 0 & q_{84} \\ 0 & 0 & 0 & q_{94} \\ 0 & 0 & 0 & q_{104} \end{pmatrix} \quad \Theta = \begin{pmatrix} 1 & \theta_{12} & \theta_{13} & \theta_{14} \\ \theta_{21} & 1 & \theta_{23} & \theta_{24} \\ \theta_{31} & \theta_{32} & 1 & \theta_{34} \\ \theta_{41} & \theta_{42} & \theta_{43} & 1 \end{pmatrix} \tag{33}$$

and Ψ is diagonal with no fixed values. The model was fitted under this restrictions using the function *cfa()* in *lavaan* R-package (Rosseel, 2012)(*14*), the method of the unweighted least squares was used (recall section 3.6). We obtained the estimates given in Figure 1.

	Factor1	Factor2	Factor3	Factor4
PIC	1.03	0.00	0.00	-0.44
NPT	0.69	0.00	0.00	0.00
TAR	-0.42	-0.37	0.00	0.00
CEP	0.00	0.82	0.00	0.00
GEH	0.00	0.90	0.00	0.00
GUR	0.00	0.00	0.85	0.00
DPB	0.00	0.00	1.09	0.00
PAI	0.00	0.00	0.00	0.87
PAS	0.00	0.00	0.00	-1.04
DPR	0.00	0.00	0.00	0.83

(a) Estimated loadings matrix

	Factor1	Factor2	Factor3	Factor4
Factor1	1.00	0.27	0.05	0.08
Factor2	0.27	1.00	-0.07	0.16
Factor3	0.05	-0.07	1.00	-0.24
Factor4	0.08	0.16	-0.24	1.00

(b) Estimated factors correlation matrix

PIC	NPT	TAR	CEP	GEH	GUR	DPB	PAI	PAS	DPR
-0.18	0.53	0.61	0.32	0.19	0.28	-0.18	0.24	-0.08	0.31

(c) Estimated specific variances

Figure 1: Estimated parameters of the model

We observe that some specific variances are negative, this is not a serious problem since the value of zero was found in their confidence interval. The estimated correlation matrix $\widehat{Q}\widehat{\Theta}\widehat{Q}^t + \widehat{\Psi}$ using the estimates $\widehat{Q}$, $\widehat{\Theta}$ and $\widehat{\Psi}$ in figure 1 reproduced properly the sample correlation matrix $\widehat{R}$, thus, the model fits adequately the observed data, and we can use its properties.

4.2.1 Interpretation of the factors

Once the model is fitted, the next step is to try to label the factors. To help us with the interpretation, in addition to the loadings, we will use the estimated matrix of correlations between the initial variables and the factors $\widehat{R}_{xf} = \widehat{Q}\widehat{\Theta}$ given in table 1.

Denote by f_1, f_2, f_3 and f_4 the four factors of the model, starting with f_1, we observe that it explains a large part of the variability of PIC (gross domestic product per capita), since $q_{11} = 1.03$. The

	Factor1	Factor2	Factor3	Factor4
PIC	0.99	0.20	0.16	-0.36
NPT	0.69	0.18	0.04	0.05
TAR	-0.51	-0.48	0.00	-0.09
CEP	0.22	0.82	-0.06	0.13
GEH	0.24	0.90	-0.06	0.15
GUR	0.05	-0.06	0.85	-0.21
DPB	0.06	-0.07	1.09	-0.27
PAI	0.07	0.14	-0.21	0.87
PAS	-0.08	-0.17	0.25	-1.04
DPR	0.07	0.14	-0.20	0.83

Table 1: Estimated correlations between initial variables and factors

only remaining common factor explaining PIC is f_4, in a smaller scale. The specific variance for PIC, that is, the part of the variability of PIC not explained by the common factors, is small, we also observe that $q_{11} = 1.03$ is positive, as the correlation between f_1 and PIC, and hence, large values of f_1 will correspond to large values on PIC, similarly, f_1 explains NPT (number of registered patents per capita), except for the specific variance, and it is positively correlated with NPT. Finally, the factor f_1 is negatively loaded in TAR (unemployment rate), and thus, large values of f_1 will imply a decrease in TAR. Taking into account this considerations, we will label the factor f_1 as *socioeconomic development*. This label must be taken as an example, with the only purpose of showing the statistical procedure to derive sustainable progress indicators, since although it is done taking into account the mathematical relationships between f_1 and the initial variables, the author have no knowledge about economics or sociology, and the label was not contrasted with the experts of this matters in IERMB. The same apply to the rest of the labels.

The interpretation of the factor f_2 is easier, the variables CEP (primary energy consumption per capita) and GEH (greenhouse gases per capita) are mainly explained with this factor, with no other factors loading in this variables, in other words, no other factors explain directly this variables, and f_2 is positively correlated with CEP and GEH and therefore the large values in this variables correspond to large values in f_2, thus, we will label this factor as *environmental impact*.

Similarly, factor f_3 explains GUR (urbanization rate) and DPB (population density) with positive loadings in each of this two variables, and is labelled as *urban complexity*, with large values on f_3 implying large values on GUR and DPB.

Finally, for f_4, we observe that is loaded positively on PAI (weight of industrial activity) but negatively on PAS (weight of services activity) and positively loaded on DPR (productive diversity) but negatively on PIC (gross domestic product per capita), this alternated relations of f_4 with the initial variables makes it a factor difficult to interpret, maybe an interpretation would been possible with the feedback of IERMB, but we can't see a label of example clearly, and in consequence, we won't use this factor on the further analysis. Not using f_4, the indicators that we may find will not take into account the variables PAI, PAS and DPR, because this variables are only explained by f_4, except of small indirect explanation by the other factors trough the correlations between them and f_4.

For convenience in the building process of the sustainable progress indicators, we would prefer the opposite of the factor f_2, that is, a factor whose large values correspond to low values in CEP (primary energy consumption) and GEH (greenhouse gases), since we want that large values on the factors coincide to good levels in the aspects involving sustainable progress. This is not a problem, because if (f_1, f_2, f_3, f_4)

give a solution to the model, the factors $(f_1, -f_2, f_3, f_4)$ also give a solution to it, concretely we observe:

Observation 4.2.1. Let (Q, f, u) be a solution to the factor model for a random vector X, with m factors, and let $I_m^i \in M_m(\mathbb{R})$, $I_m^i = diag(1, \dots, 1, -1, 1, \dots, 1)$ with the value -1 in the ith position, that is, the m square matrix with ones on the diagonal, except the value -1 in the ith position. We observe that I_m^i is an orthogonal matrix, and so, it holds (analogously to proposition 2.1.6) that $(QI_m^i, I_m^i f, u)$ is a solution to the factor model for X with m factors, concretely $I_m^i f = (f_1, \dots, -f_i, \dots, f_m)^t$ as desired, and QI_m^i is the matrix Q but with the loadings in the second column with the signs changed. Also, supposing $Cov(f) = \Theta$, we have $Cov(I_m^i f) = I_m^i \Theta I_m^i$, that is, Θ, but with the ith row and column with the signs changed, the specific variances do not change.

Hence, we can consider the factors $f_1, -f_2, f_3$ and f_4, since they explain as well the initial variables. Denoting $\widetilde{f_2} := -f_2$, now the factor $\widehat{f_2}$ is interpreted with the exact same loadings as f_2 but with changed sign in them, and thus, $\widetilde{f_2}$ explains CEP and GEH, but with the opposite relationship as f_2 with this variables, that is, large values of $\widetilde{f_2}$ will coincide with low values in CEP and GEH, as desired, and we will label $\widetilde{f_2}$ as *environmental protection.* The other factors remain the same.

Thus, we consider the factors f_1(*socioeconomic development*), $\widetilde{f_2}$ (*environmental protection*) and f_3 (*urban complexity*). Supposing that their labels were valid, this factors could measure their respective labelled aspects for the metropolitan regions of study, trough the factor scores, with large values on the factor scores meaning a high level in their labels. We will use this factors to build indicators.

4.3 Sustainable progress indicators

The next step to obtain a sustainable progress indicator is deriving simple indices from the factors f_1, $\widetilde{f_2}$ and f_3, the indices will measure the same aspects as the factors, but they will have more adequate properties, this simple indices are integrated into compound indicators, which take into account all the three aspects measured by the factors and, therefore, may be able to measure sustainable progress (in the case that factors trully represent its labels). We will denote $\widetilde{f_2}$ by f_2, abusing of notation, and reminding that f_2 will now refer to the factor *environmental protection*

We will use the term "probability-based index" to refer to any statistic that is a function of several variables and satisfies: *i*) it takes values on a bounded interval; *ii*) it can be used to rank a set of observations; *iii*) it depends on a parametric family of distributions adjusted to the data sample (Marull et al., 2019) (*11*). We will first build three probability-based indices from the empirical distributions of the factors f_1, f_2 and f_3.

Let $\widetilde{X_z}$ be the standardized data matrix summarizing the observations of the metropolitan regions, and consider the factor scores of this observations $F = \widetilde{X_z}\widehat{R}^{-1}\widehat{Q}\widehat{\Theta}$ for the factors f_1, f_2, f_3. Denote F_1, F_2 and F_3 the first, second and third column vector of F respectively, that is, the factor scores of the factors f_1, f_2 and f_3. The simple indices are obtained using F_1, F_2 and F_3. The first step is applying a transformation to the scores, concretely, we apply a Box-Cox monotone power type transformation (Box and Cox, 1964)(*2*) to $F1$, $F2$, and $F3$ independently. Let $F_{ij} = (F_{1j}, \dots, F_{95j})^t$ be the scores of the factor f_j, for $j \in \{1, 2, 3\}$, for all the 95 metropolitan regions. Let F_{ij}^{λ} be the transformed scores of the factor f_j, the Box-Cox transformation is given by

$$F_{ij}^{\lambda} = \frac{(F_{ij} + m)^{\lambda} - 1}{\lambda}$$

for $\lambda \neq 0$, and for some m such that $F_{ij} + m \geq 0$, $\forall i = 1, \dots, 95$. This transformation is monotone increasing and therefore it does not change the classification that factor f_j does of the regions, that is, if $F_{ij} \leq F_{kj}$, then $F_{ij}^{\lambda} \leq F_{kj}^{\lambda}$, moreover, the transformed factors are more symmetric, and we use this property to adjust the transformed factor scores to a Laplace distribution. The Laplace's density and cumulative distribution functions are given by:

$$f(x;m,b)=\frac{1}{2b}exp\Big(-\frac{|x-m|}{b}\Big) \quad \text{and} \quad \Phi(x;m,b)=\frac{1}{2}+\frac{1}{2}\mathrm{sgn}(x-m)-\exp\Big(-\frac{|x-m|}{b}\Big) \qquad (34)$$

respectively, where sgn denotes the function sign. This is the parametric family of distributions that the three indices will depend on. We observe that the Laplace's density is symmetric, and therefore it will likely fit to the transformed scores empirical distribution, choosing the adequate parameters m and b. m is called location parameter and it indicates the axe of symmetry of the density function, its maximum likelihood estimate is the empirical median $\widehat{m}$ (Norton, 1984) (*12*), whereas a maximum likelihood estimator for b is $\widehat{b}=\frac{1}{n}\sum_{i=1}^{n}|x_i-\widehat{m}|$. Hence, denoting $tF_j; j\in\{1,2,3\}$ the vectors of transformed scores, we adjust the transformed scores tF_j to the Lapalce distribution with the parameters of maximum likelihood $\widehat{m}_j$ and $\widehat{b}_j$ computed using tF_j. The procedure is visualized in the next two graphics, for F_3:

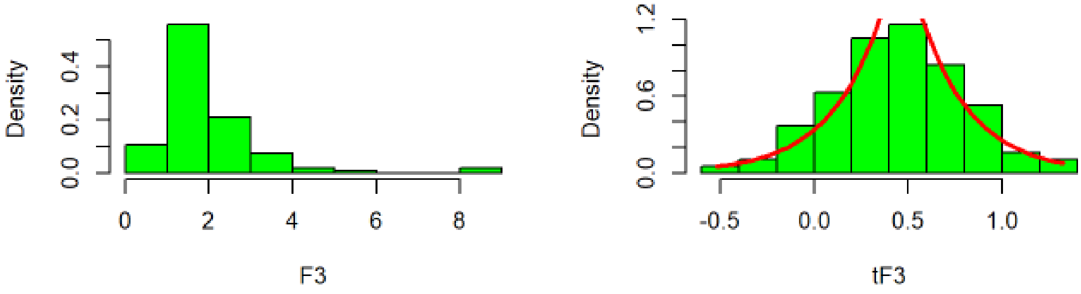

Figure 2: Factor scores adjusted to Laplace distribution

In the two histograms, left for F_3 and right for transformed scores, we can see the symmetry achieved with the Box-Cox transformation, the red lines represent the density function of Laplace with parameters $\widehat{m}_3$ and $\widehat{b}_3$, the fit for F_3 is good, although it may not be the case in general. Let $\Phi_j(x)$ be the cumulative distribution function of parameters $\widehat{m}_j$ and $\widehat{b}_j$, for $j\in\{1,2,3\}$, that is, the cumulative distribution function adjusted to each transformed scores, in the figure 3 is shown the adjust of $\Phi_j(x)$ to the empirical cumulative distribution function of tF_j.

The probability based index for F_j, is given by $I_j := \Phi_j(x)$, thus, it depends on the Laplace distribution adjusted to the the scores tF_j, and it takes values in $[0,1]$ and ranks the observations of the metropolitan regions in the same order as given by tF_j, and hence in the same order as F_j, since $\Phi_j(x)$ is a cumulative distribution function. Therefore $I_j := \Phi_j(x)$ will classify and measure the observations of the regions as the scores F_j, but it have adequate properties, as we want. The value of I_j for a metropolitan region i is given by

$$I_{ij} := \Phi_j(tF_{ij})$$

and the larger the value I_{ij} is, the better positioned will be the region i in terms of the aspect explained by the factor f_j. In fact, I_{ij} can be interpreted as the probability to obtain a value less or equal than the observed tF_{ij} within all the scores in tF_j, since tF_{ij} is supposed to be an observation of a Laplace variable with cumulative distribution function Φ_j.

Once the simple indices I_1(*economic development*), I_2(*environmental protection*) and I_3(*urban complexity*) are obtained, we can search for a compound index. We will consider as a compound index any linear combination of the three simple indices of the form:

$$S = w_1I_1 + w_2I_2 + w_3I_3;\ w_1+w_2+w_3=1 \qquad (35)$$

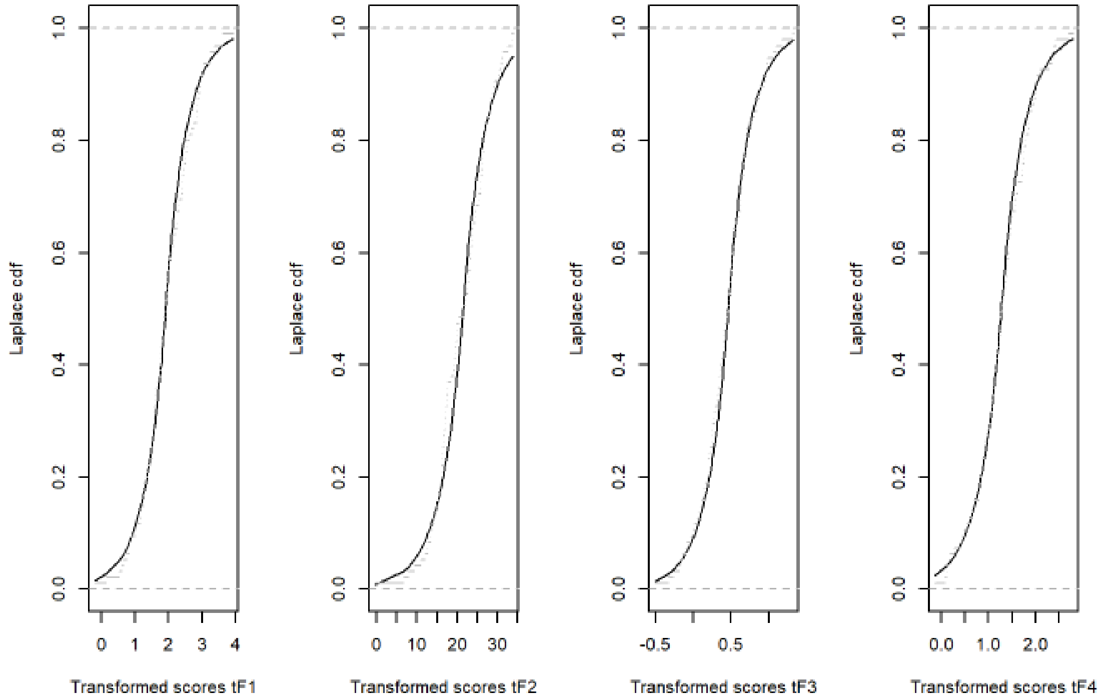

Figure 3: Black: adjusted distribution vs Grey: empirical distribution of transformed scores

This way, S will also take values in $[0, 1]$ and it will take into account all the three aspects explained by the simple indices. If we had been able to interpret f_4, we also included a simple index for it in the above expression, and the compound indicators would take into account more features, maybe related with the economic activity of the metropolitan regions, due the variables explained by f_4.

We propose 3 compound indicators:

$$S_1 = \frac{1}{3}I_1 + \frac{1}{3}I_2 + \frac{1}{3}I_3 \tag{36}$$

$$S_2 = \frac{2}{4}I_1 + \frac{1}{4}I_2 + \frac{1}{4}I_3 \tag{37}$$

$$S_3 = \frac{1}{4}I_1 + \frac{2}{4}I_2 + \frac{1}{4}I_3 \tag{38}$$

S_1 is an indicator that rewards balanced values in the economic, social, ecological and urban aspects, hence, we will denote it as *inclusive development.* S_2 rewards high values on I_1(*socioeconomic development*), we will label it *socioeconomic sustainability.* Finally, S_3 rewards high values on I_2(*environmental protection*), and we will label it as an *environmental sustainability* indicator. Figure 4 shows the values taken by I_1, I_2, I_3, S_1, S_2 and S_3 for all the 95 metropolitan regions, in the year 2016.

4.4 Discussion

The values of the indicators in figure 4 measure the level of their respective labels on each metropolitan region in the year 2016, supposing the labels were accurate. values close to 1 indicate a good position in the aspect evaluated, whereas values close to 0 indicate a poor performance. The effects of the weightings are clear, for example, Madrid and Barcelona having low values on I_1(*socioeconomic development)* and large values on I_2(*environmental protection)* are clearly rewarded by the index S_3(*environmental sustainability*) and penalized by the index S_2(*socioeconomic sustainability*), whereas London is rewarded by

S_2(*socioeconomic sustainability*). An extensive discussion of the values in figure 4 could we performed. Disposing of data of more years, the evolution in the values of the indicators could be observed, in that case, it is possible to evaluate the progress or involution of each aspect measured by the indicators in the metropolitan regions over the years. Supposing the labels of the indicators were really valid, the indicator S_1(*inclusive development*) is a balanced indicator over all socioeconomic, ecological and urban aspects, taking into account the dimensions characterizing sustainable progress, thus, it might be a candidate to sustainable progress indicator if evaluated over different years. Its important to remind that, as we said in the introduction, the values in 4 can not be used to draw conclusions about the aspects with which we, with the purpose of showing this procedure, have labelled the indicators as an example, since this labels have not been discussed with experts on the matter.

5 Conclusion

Factorial analysis has proved to be one useful tool to analyse quantitatively, aspects with qualitative dimensions, reducing the inherent subjectivity involving qualitative valuations, to an interpretation of the effects of a factor over known and observed variables, where the consensus could be greater. Recall the case of the ”overall level of intelligence”, it could have a greater consensus understanding a factor as the ”overall level of intelligence” if their large values imply large values on a series of exams qualifications, than tell, based directly on the exams qualifications and with no other tool, the ”overall level of intelligence” of a student.

6 Acknowledgments

I would like to thank all the people who have contributed to make this document possible, starting with my tutoress for this work, Mercè Farré, who introduced me to multivariate statistical analysis, trough its complete lessons, taking into account the applications but also the mathematics of every theory. She proposed me to do this work, and proofread it with precious annotations, I wish to thank her also hers help in the data analysis. A sincere thank you to all the people in IERMB that worked in preparing the analyzed data set and thinking of it, they gave me the opportunity to apply the procedure explained in this document with actual data, in particular, I want to thank Joan Marull for dedicating his time in explaining me the study in which the analysis is based on, when this work was proposed.

Thanks to my family, for caring for me. Thanks to my friends, for making me feel accompanied.

References

(*1*) Banachiewicz, T. (1937). Zur Berechnung der Determinanten, wie auch der Inversen und zur darauf basierten Auflosung der Systeme linearer Gleichungen. *Acta Astronom. Ser. C 3*, 41–67.

(*2*) Box, G. E., and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological) 26*, 211–243.

(*3*) Cedó, F., and Reventós, A., *Geometria plana i àlgebra lineal*; Univ. Autònoma de Barcelona: 2004, pp 448–451.

(*4*) Härdle, W. K., and Simar, L., *Applied Multivariate Statistical Analysis*; Springer Berlin Heidelberg: 2012.

(*5*) Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika 32*, 443–482.

(*6*) Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika 34*, 183–202.

(*7*) Joreskog, K. G., and Goldberger, A. S. (1971). Factor analysis by generalized least squares. *ETS Research Bulletin Series 1971*, i–32.

(*8*) Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika 23*, 187–200.

(*9*) Lin, M., and Sra, S. (2014). Completely strong superadditivity of generalized matrix functions. *arXiv preprint arXiv:1410.1958.*

(*10*) Mardia, K., Kent, J., and Bibby, J., *Multivariate Analysis (Probability and Mathematical Statistics)*; London: Academic Press: 1979.

(*11*) Marull, J., Farré, M., Boix, R., Palacio, A., and Ruiz-Forés, N. (2019). Modelling urban networks sustainable progress. *Land Use Policy 85*, 73–91.

(*12*) Norton, R. M. (1984). The double exponential distribution: Using calculus to find a maximum likelihood estimator. *The American Statistician 38*, 135–136.

(*13*) Peña, D. Análisis de Datos Multivariantes. Madrid: McGraw Hills, 2002.

(*14*) Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software 48*, 1–36.

(*15*) Schur, J. (1917). Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind. *Journal für die reine und angewandte Mathematik 1917*, 205–232.

(*16*) Thomson, G. H. (1935). The definition and measurement of" g"(general intelligence). *Journal of Educational Psychology 26*, 241.

	I1	I2	I3	S1	S2	S3
London	0.97	0.28	0.95	0.73	0.79	0.62
Paris	0.91	0.64	0.92	0.82	0.84	0.78
Madrid	0.15	0.91	0.74	0.60	0.49	0.68
Barcelona	0.09	0.95	0.71	0.58	0.46	0.68
Berlin	0.28	0.22	0.16	0.22	0.24	0.22
Ruhrgebiet	0.33	0.13	0.91	0.46	0.43	0.38
Roma	0.54	0.83	0.57	0.65	0.62	0.69
Milano	0.78	0.94	0.89	0.87	0.85	0.89
Manchester	0.29	0.65	0.78	0.57	0.50	0.59
Athina	0.06	0.87	0.92	0.62	0.48	0.68
Hamburg	0.77	0.17	0.60	0.51	0.58	0.43
Amsterdam	0.84	0.25	0.30	0.46	0.56	0.41
Napoli	0.40	0.79	0.98	0.72	0.64	0.74
Marseille	0.23	0.20	0.60	0.34	0.32	0.31
Budapest	0.02	0.84	0.25	0.37	0.28	0.49
Warszawa	0.17	0.34	0.05	0.19	0.18	0.23
Munchen	0.94	0.23	0.67	0.61	0.70	0.52
Lisboa	0.05	0.95	0.53	0.51	0.40	0.62
Wien	0.79	0.21	0.70	0.57	0.62	0.48
Stuttgart	0.92	0.19	0.89	0.67	0.73	0.55
Katowice	0.95	0.04	0.80	0.60	0.68	0.46
Frankfurt	0.87	0.16	0.81	0.61	0.68	0.50
LDV	0.07	0.50	0.50	0.36	0.29	0.39
Praha	0.26	0.19	0.33	0.26	0.26	0.24
Valencia	0.14	0.86	0.19	0.40	0.33	0.51
Bruxelles	0.78	0.66	0.71	0.72	0.73	0.70
WMUA	0.30	0.55	0.98	0.61	0.53	0.60
Bucuresti	0.70	0.21	0.94	0.62	0.64	0.51
Torino	0.21	0.91	0.09	0.40	0.36	0.53
Stockholm	0.97	0.40	0.86	0.74	0.80	0.66
Dublin	0.98	0.14	0.09	0.40	0.55	0.34
Kbenhavn	0.94	0.35	0.62	0.64	0.71	0.56
Köln	0.78	0.12	0.92	0.61	0.65	0.48
Sevilla	0.31	0.82	0.29	0.47	0.43	0.56
AlElche	0.17	0.61	0.20	0.33	0.29	0.40
Glasgow	0.30	0.16	0.25	0.24	0.25	0.22
Lyon	0.74	0.59	0.76	0.70	0.71	0.67
Rotterdam	0.69	0.03	0.76	0.49	0.54	0.38
Liverpool	0.11	0.50	0.87	0.49	0.40	0.50
Porto	0.17	0.95	0.42	0.51	0.43	0.62
Leeds	0.54	0.36	0.23	0.38	0.42	0.37
Sofia	0.12	0.38	0.06	0.19	0.17	0.24
Göteborg	0.89	0.76	0.39	0.68	0.73	0.70
MaMarbella	0.20	0.79	0.19	0.39	0.34	0.49
Helsinki	0.92	0.32	0.52	0.59	0.67	0.52
Bordeaux	0.82	0.19	0.76	0.59	0.65	0.49
Düsseldorf	0.89	0.10	0.93	0.64	0.70	0.50
MuCartagena	0.91	0.18	0.77	0.62	0.69	0.51
Kraków	0.89	0.18	0.79	0.62	0.69	0.51
Leicester	0.26	0.86	0.61	0.58	0.50	0.65
Nantes	0.50	0.80	0.41	0.57	0.55	0.63
Toulouse	0.43	0.37	0.46	0.42	0.42	0.41
Dresden	0.53	0.33	0.40	0.42	0.45	0.40
Nürnberg	0.66	0.16	0.52	0.45	0.50	0.38
Malmö	0.90	0.67	0.13	0.57	0.65	0.59
Gdansk	0.45	0.11	0.23	0.26	0.31	0.22
Hannover	0.48	0.18	0.39	0.35	0.38	0.31
Utrecht	0.86	0.78	0.77	0.80	0.82	0.80
Palermo	0.39	0.88	0.26	0.51	0.48	0.60
Brescia	0.81	0.68	0.20	0.56	0.63	0.59
Bari	0.71	0.73	0.72	0.72	0.72	0.72
Bremen	0.63	0.24	0.30	0.39	0.45	0.35
Rouen	0.39	0.23	0.74	0.45	0.44	0.40
Grenoble	0.36	0.23	0.19	0.26	0.28	0.25
Cádiz	0.08	0.81	0.14	0.34	0.28	0.46
Zagreb	0.19	0.75	0.05	0.33	0.30	0.44
Ostrava	0.65	0.07	0.73	0.48	0.52	0.38
Brno	0.43	0.20	0.56	0.40	0.40	0.35
MannheimL	0.89	0.05	0.88	0.61	0.68	0.47
Poznan	0.51	0.10	0.12	0.24	0.31	0.21
NUponTyne	0.06	0.62	0.01	0.23	0.19	0.33
Bilbao	0.25	0.90	0.59	0.58	0.50	0.66
Montpellier	0.27	0.11	0.64	0.34	0.32	0.28
Bristol	0.80	0.66	0.85	0.77	0.78	0.74
ACoruña	0.54	0.78	0.32	0.55	0.55	0.60
Strasbourg	0.25	0.71	0.36	0.44	0.39	0.51
StokeOnTrent	0.15	0.59	0.34	0.36	0.31	0.42
Catania	0.57	0.78	0.41	0.59	0.58	0.64
Thessaloniki	0.03	0.91	0.03	0.32	0.25	0.47
Bergamo	0.82	0.66	0.23	0.57	0.63	0.59
Nice	0.26	0.37	0.32	0.32	0.30	0.33
Lódz	0.08	0.35	0.16	0.20	0.17	0.24
sGravenhage	0.82	0.01	0.96	0.60	0.65	0.45
Rennes	0.71	0.80	0.19	0.57	0.60	0.62
OviedoGijon	0.28	0.80	0.16	0.41	0.38	0.51
Antwerpen	0.53	0.23	0.47	0.41	0.44	0.36
Leipzig	0.40	0.09	0.54	0.34	0.36	0.28
Firenze	0.39	0.92	0.11	0.47	0.45	0.58
Riga	0.17	0.71	0.08	0.32	0.28	0.42
Bologna	0.58	0.88	0.05	0.50	0.52	0.60
BSW	0.91	0.14	0.60	0.55	0.64	0.45
Zaragoza	0.44	0.64	0.23	0.44	0.44	0.49
Vigo	0.60	0.80	0.56	0.65	0.64	0.69
Padova	0.38	0.84	0.03	0.42	0.41	0.52
Verona	0.75	0.56	0.07	0.46	0.53	0.49

Figure 4: Values of the simple indicators I_1(*socioeconomic development*), I_2(*environmental protection*) and I_3(*urban complexity*), and the compound indicators S_1(*inclusive development*), S_2(*socioeconomic sustainability*), and S_3(*environmental sustainability*). LDV: LilleDunkerqueValenciennes, WMUA: WestMidlansUrbanArea, AlElche: AlicanteElche, MaMarbella: MálagaMarbella, MuCartagena: MurciaCartagena, MannheimL: MannheimLudwigshafen, NUponTyne: NewcastleUponTyne, BSW: BraunschweigSalzgitter-Wolfsburg.